ICMPC15
ESCOM10
2018

Proceedings

Graz La Plata Montréal Sydney

ISBN: 978-3-200-05771-5

How to cite this volume
Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Proceedings of ICMPC15/ESCOM10: Introduction

Richard Parncutt[1] and Sabrina Sattmann[2]

Centre for Systematic Musicology, University of Graz, Austria

[1]parncutt@uni-graz.at, [2]sabrina.sattmann@uni-graz.at

The 15th International Conference on Music Perception and Cognition (ICMPC15) and the 10th triennial conference of the European Society for the Cognitive Sciences of Music (ESCOM10) were held simultaneously from 23 to 28 July 2018 in four different global locations: Graz, Austria; La Plata, Argentina; Montreal, Canada; and Sydney, Australia. For the first time, all talks were live streamed. Moreover, each talk was seen during the conference at another hub, either as a live stream followed by audio discussion or as a video followed by written discussion. All talks were available to registered participants as videos during and after the conference.

This strategy increased the number of participants by about one half relative to previous comparable conferences. It also opened the conference to many colleagues who would otherwise have been unable to participate for financial reasons—in particular, those from South America and Australia, but also students at all locations. This in turn increased the conference's cultural diversity. By reducing the distance that the average participant traveled, we were able to reduce pro capita carbon emissions by roughly 70% relative to a comparable scenario in which all participants traveled to Graz.

Of the 685 submissions that we received for talks, posters, workshops, or demonstrations, 635 were accepted following a thorough peer review procedure. Authors of every abstract received three anonymous reviews. On the basis of mean reviewer grades, talk/poster submissions were assigned to one of four categories (long talks, short talks, posters, rejects), and workshop/demonstration submissions were either accepted or rejected. At each hub, there were generally more posters than short talks, and more short talks than long talks.

All active participants (all presenters of long talks, short talks, and posters) were invited to contribute to this electronic proceedings volume, and 105 did so. Unlike their original abstract submissions, contributions to this volume were not reviewed, nor were submissions edited.

This volume contains only proceedings contributions, in alphabetical order of first author. For further information about the conference, see icmpc.org/icmpc15. For introductions from society and hub organizers, see the electronic abstract book.

We would like to take this opportunity to thank all those who trusted us with their research reports and carefully followed our guidelines, without which the conference and this proceedings book would not have been possible. We especially wish to thank the members of the ICMPC executive committee, the ESCOM executive council, and the organizers, organizing committees, technicians and hospitality team members at all four hubs.

Richard Parncutt and Sabrina Sattmann

Centre for Systematic Musicology
University of Graz, Austria
September 2018

Table of Contents

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Structure of Chord Progressions Influences Listeners' Enjoyment and Absorptive States in EDM

Kat Agres[1] and Dorien Herremans[1,2]

[1] *Social and Cognitive Computing Department, Institute of High Performance Computing, A*STAR, Singapore*

[2] *Information Systems Technology and Design, Singapore University of Technology and Design, Singapore*

[1]kat_agres@ihpc.a-star.edu.sg, [2]dorien_herremans@sutd.edu.sg

Abstract

The field of music cognition has given comparatively little consideration to the topic of altered listening states, such as audience flow, trancing, and absorptive states. Some research has investigated the relationship between musical features (such as repetitiveness or information-theoretic characteristics) and enjoyment of the music, but the impact of musical structure on absorptive states has yet to be thoroughly addressed. The present study sought to fill this gap in the literature by examining harmonic structure, enjoyment, and absorptive listening states in Uplifting Trance (UT), a sub-genre of Electronic Dance Music (EDM). Rather than examine the obvious (and previously considered) connection between driving rhythmic cues and absorptive listening states, we aimed to investigate the effect of harmonic repetition on listening states. Based on previous work (Agres, et al, 2017), we generated a number of UT excerpts with varying degrees of harmonic repetition. These excerpts were used in an online listening task with 19 experienced trance/EDM listeners. We discovered that chord progressions significantly influenced both listeners' enjoyment of the music as well as their reported absorptive listening states. Overall, the results from this study indicate that moderately complex chord sequences elicit greater enjoyment than very repetitive chord sequences or sequences that violate listeners' expectations. A significant correlation between highly enjoyed sequences and absorptive listening states was also observed.

Introduction

From tribal cultures to western electronic dance music (EDM) contexts, trance music helps listeners achieve altered listening states (Kartomi, 1973; Kakouri, 1965). The precise musical features and mechanisms responsible elucidating these altered states are still a topic of investigation by scientists around the world (Becker-Blease, 2004; Becker, 2012; Fachner, 2011; Neher, 1962; Trost, Frühholz, Schön, Labbé, Pichon, Grandjean & Vuilleumier, 2014). The relationship between harmonic structure and enjoyment of EDM has previously been investigated by the authors (Agres, Herremans, Bigo & Conklin, 2017), however, the influence of chord progressions on *absorptive listening (AL) states* has remained unexplored.

Different terms have been used in the literature to refer to the wide range of experiences that may be elicited from music, such as 'trancing' and 'altered states of consciousness' (see for example Ludwig, 1966; Silverman, 1968; and Becker-Blease, 2004). Different connotations are attached to each of these alternatives. In this paper, we focus on 'absorptive listening states', which refer to the experience of being deeply engulfed in music. Although there is intuitively a connection between enjoyment and the sense of being absorbed in the music, this topic has not been formally addressed. The current work makes use of an empirical listening task with EDM experts to fill this gap in the literature.

Existing research has examined the influence that rhythmic and percussive elements have, through entrainment and auditory driving, on heightened enjoyment (Becker-Blease, 2004; Becker, 2012; Fachner, 2011; Neher, 1962; Trost et al., 2014). Because ample research suggests that these temporal cues do have an impact on absorptive listening states, the authors have instead chosen to examine a feature of trance music less studied in this context: harmonic structure.

In previous work, a listening experiment was conducted to examine the relationship between harmonic structure and enjoyment of EDM trance music (Agres et al., 2016; Agres et al., 2017). In this study, expert participants listened to excerpts of uplifting trance (UT)[1] music containing varying degrees of harmonic complexity (details in Method section below). The experiment by Agres et al. (2017) provided evidence for a significant connection between enjoyment and harmonic structure of UT excerpts, as defined in terms of repetitiveness, complexity, and tension. The degree of harmonic complexity of the UT stimuli was carefully controlled by imposing certain *semiotic structures*, or patterns of chord progressions, on the 16-bar-long stimuli. Fourteen different semiotic structures (which varied from extremely repetitive to moderately complex) were examined, with one such example being '`AABB-AACC`'. In this semiotic structure, the first chord is repeated for four bars (note that each element in the pattern represents two bars of music), followed by a different chord for four bars, which is then followed by the first chord for four bars, and so on. Also note that each element represents a unique chord, not a chord name. For more information regarding the stimuli and how they were generated, the reader is referred to Agres, et al. (2017).

The results of the study indicated that trance excerpts with moderately repetitive chord sequences are most highly enjoyed, with participants demonstrating a preference for moderate harmonic complexity and tonal tension (as defined by Herremans & Chew, 2017). Further, sequences whose structure violated the expected form (such as `AABB-AAAA`, in which the third '`A`' chord in the second half of the sequence is unexpected), created a lack of enjoyment in listeners. These findings suggest that the enjoyment of uplifting trance music may be related to the predictability of its underlying harmonic

[1] Uplifting trance is a sub-genre of EDM typically in the range of 135-150 beats per minute, with a song structure often comprised of intro, breakdown, build-up, release, anthem and outro sections (Madrid, 2008).

structure: extremely repetitive chords as well as unexpected chords are not always enjoyed by listeners, whereas chord progressions that are moderately expected are better enjoyed. The pattern of results can be described by a Wundt curve (see Berlyne, 1970), and this inverted-U relationship between complexity and enjoyment (or liking) has also been found in other musical genres, such as jazz (Gordon and Gridley, 2013) and bluegrass music (Orr and Ohlsson, 2001). Arguably, the one of the highest forms of enjoyment of UT music is reaching an absorptive state, however Agres et al. (2017) did not directly test the connection between chord progressions and AL states.

In both modern EDM settings and in numerous tribal cultures, trance music has been associated with reaching absorptive states of consciousness, yet again, no research has investigated whether harmonic structure contributes to this experience. Therefore, this study aimed to investigate the influence of harmonic structure on reaching AL states. We also sought to replicate and further validate the results of Agres, et al. (2017), which found that harmonic structure has a significant influence on enjoyment. In the present study, conducted solely with expert listeners, we use longer trance excerpts that are more ecologically valid, and provide listeners with a greater opportunity to "lose themselves in the music." Finally, we sought to explore the relationship between enjoyment ratings and reported absorptive listening states.

Behavioural Experiment

In an online listening experiment, we investigated the nature of the influence of harmonic structure on both enjoyment and self-reported ability to reach absorptive listening states. Several hypotheses drove this project. First, we hypothesised that harmonic structure influences the (self-reported) ability of listeners to reach AL states. Secondly, we predicted that harmonic structure (i.e., chord sequences) in UT music influences listeners' enjoyment ratings in longer, more ecological musical stimuli (we aimed to replicate and extend the results from Agres et al. (2017), which used brief UT excerpts). Our third hypothesis was that enjoyment and absorption are linked responses; that is, that listeners' enjoyment of the music would be directly related to their self-reported AL state.

Method

Participants. The study was advertised through email and on trance production forums, which resulted in a total of 25 responses. Because we wanted a population of expert listeners, only participants with a minimum of one year of experience DJing, composing, or producing trance music were included in the study, which resulted in a total of 19 participants (mean age = 33.4 yrs, std = 10.5 yrs; 19 male). These expert participants had on average 9.6 years (std = 7.7 yrs) experience DJing, composing, or producing trance music. To encourage and reward participation in the study, a £25 Amazon voucher was given to one randomly selected participant.

Stimuli. For this experiment, we created extended versions of a subset of the stimuli from Agres, et al. (2017). The original chord progressions from Agres, et al. (2017) were generated using a statistical model trained on a corpus of chord progressions from the anthem section of 100 uplifting trance songs[2]. For this research, we selected the three most enjoyed and three least enjoyed stimuli from the previous experiment (see Table 1), hereafter referred to as *Liked* and *Disliked* stimuli. The original sequences, taken from the first study, were expanded from 30 seconds to 120 seconds for the present experiment by adding a breakdown and a build-up section to each excerpt. The stimuli in Agres et al. (2017) were composed to span a degree of repetitiveness and complexity. To be consistent with the template song on which the original stimuli were based[3], the breakdown section consisted of the same chords as the anthem, and the build-up section was comprised of only one chord (the first chord of the anthem). The resulting stimuli are available online[4].

Table 1. The three most Liked and Disliked chord progressions from Agres et al. (2017), which were expanded for the current study. Note that the stimuli are listed in descending order of enjoyment, with the most Liked and most Disliked progressions in bold font.

Liked	Disliked
DBmGA-BmGABm	DDF#mF#m-F#mF#mDD
BmGDA-BmGDA	EmEmDD-EmEmEmEm
BmBmGG-BmBmGG	**GGCC-CCGG**

Procedure. An online listening study was conducted using Qualtrics[5]. On every trial, listeners heard a 2-min UT excerpt (including breakdown, build-up, and anthem sections). After listening to each stimulus, participants first rated their enjoyment of the excerpt on a Likert scale from 1-7 (where 1 represented 'Not at all' and 7 represented 'Very much'). Then participants rated the state of absorption they achieved while listening to the excerpt, specifically, the extent to which they felt able to lose themselves in the music, e.g., "achieve a trance-like state of mind, become totally absorbed, or reach a state of flow." This judgement was again made using a Likert scale from 1 to 7, with 1 representing 'Not at all' and 7 representing 'Very much'.

At the end of the listening study, the participants were asked to complete a brief questionnaire about their trance music listening habits and expertise. In this questionnaire, the participants with experience *composing* trance pieces were asked to consider how their choice of chord progressions might affect the enjoyment or absorptive listening states of their audience, and the extent to which they take this into consideration when creating/producing a new piece of trance music. They were also asked how many years of experience they had as a DJ, composer, or music producer; how many hours per week they listen to UT music; and whether they have experienced trance-like absorptive flow states when listening to music.

[2] The corpus of 100 UT songs is available online at http://katagres.com/trance_experiment

[3] Uplifting Trance Logic Pro X Template by Insight, DAW Templates, Germany

[4] The stimuli are available online at http://katagres.com/trance_experiment

[5] http://qualtrics.com

Results and Discussion

We tested all three hypotheses outlined in the Behavioural Experiment section. First, we tested the influence of chord structure on the ability of the listeners to reach AL states. Second, we examined the influence of chord structure in longer stimuli on listeners' enjoyment, and tested whether the results correspond to those in Agres et al (2017). Finally, the correlation between enjoyment and AL states was explored.

The results from the experiment confirm that there are meaningful relationships between harmonic structure and AL states, as well as subjective enjoyment and AL states. A mixed effects analysis, with AL states as the dependent variable and chord sequence as the independent variable, indicated a significant influence of chord sequence on AL states ($F = 2.73$, $p < .05$), providing evidence that harmonic structure does influence absorptive listening states. The chord progression corresponding to the highest average AL state rating was `BmBmGG-BmBmGG`, while that corresponding to the lowest average AL state rating was `GGCC-CCGG`. The semiotic structure of the latter (`AABB-BBAA`) violates expectations of musical form, as a change would be expected at the beginning of the second half of the sequence; this is not the case in the highest rated stimuli (`AABB-AABB`).

A second mixed effects analysis, with enjoyment ratings as the dependent variable and chord sequence as the independent variable, yielded a significant effect of chord progression ($F = 3.48$, $p < .01$), supporting the hypothesis that harmonic structure influences the enjoyment of UT excerpts. The three chord progressions that were most enjoyed in this listening study correspond to the most enjoyed (moderately complex) sequences from Agres et al (2017), although it should be noted that their ranked order of enjoyment differs.[6] A mixed effects analysis was performed to test whether the most liked sequences from Agres et al (2017) correspond to the most enjoyed stimuli in the current study. In this analysis, enjoyment ratings (from the current study) were the dependent variable, liked/disliked stimuli (labels from Agres et al., 2017) were the independent variable, and participant was included as a random effect. This analysis provided significant evidence that the most liked stimuli from the previous work do indeed correspond to the most enjoyed in the current study ($R2 = 0.59$, $F(1, 1) = 7.89$, $p < 0.01$), and confirms more generally the significant influence of harmonic structure on enjoyment in these longer, more ecologically valid stimuli.

Finally, the relationship between enjoyment and AL states was investigated. First, we examined which stimuli elicited the highest and lowest enjoyment and AL state ratings. As mentioned above, the sequence yielding the highest enjoyment rating of this study was `BmBmGG-BmBmGG` (avg rating = 4.68, std error = 0.25). Similarly, the stimulus with the highest AL state rating was also `BmBmGG-BmBmGG` (avg rating = 4.32, std error = 0.31). The stimulus with the lowest enjoyment rating was `GGCC-CCGG` (avg rating = 3.53, std error = 0.42), and again, the stimulus with the lowest AL state rating was `GGCC-CCGG` (avg rating = 3.00, std error = 0.37). A linear correlation confirmed that these two variables, enjoyment and AL state, are significantly related ($\rho = 0.60$, $p < 0.001$). This relationship is depicted in Figure 1.

The results above indicate that chord progressions influence both AL states and enjoyment. In addition, the results confirm that enjoyment and AL states are correlated. In order to verify that the influence of chord progressions on absorptive states was not simply due to the influence of enjoyment, we tested whether the effect of chord structure on AL state ratings was mediated by enjoyment. A mediation analysis indicated that enjoyment did not have a significant mediating effect on AL states (ACME: $p=0.56$). That is, the influence of chord structure on AL states was not caused by the relationship between chord structure and enjoyment, confirming the significant effect of harmonic structure on absorptive listening states.

Results of the post-study questionnaire were also examined. Out of the 19 participants, 11 (or 58% of this expert group) indicated that they do carefully select chord progressions (when producing or composing trance tracks) to create a desired effect on the audience's enjoyment or absorption of the music. One of the participants, who was a producer with significant experience, reported that the chord progression "takes the listener on a journey", and that the chords need to be carefully chosen so that the music sounds engaging but not boring. Another respondent claimed that chord progressions are important for setting up expectations in the listener, and that delaying the resolution of the build-up creates a powerful harmonic tension in listeners, which is one of the most pleasurable aspects of trance listening. Although anecdotal, these responses indicate that composers of this genre do attempt to select chord progressions with the specific intention of guiding listeners' affective responses and states of consciousness. These subjective reports help support and confirm the important role of harmonic structure in trance music listening.

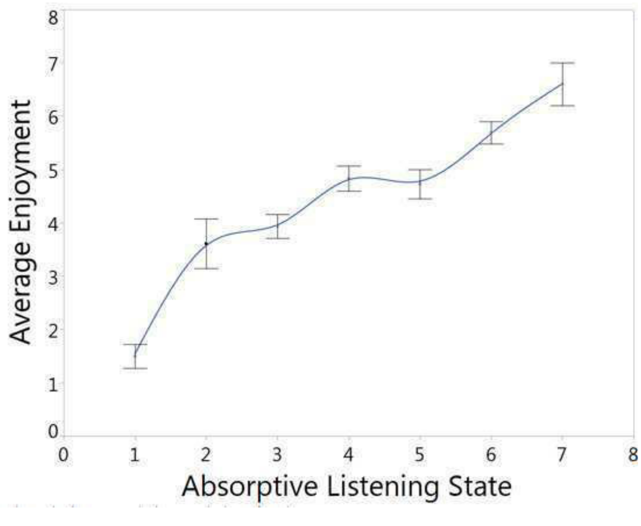


Figure 1. Average enjoyment ratings correlated with absorptive listening state ratings for all of the stimuli in the study. The error bars show the standard error from the mean.

[6] Here, the progression with the highest enjoyment, `BmBmGG-BmBmGG`, is followed by `BmGDA-BmGDA`, and then by `DBmGA-BmGABm`, in order of average enjoyment rating.

Conclusion

We investigated three hypotheses in this study. First, we examined whether harmonic structure in UT music influences listeners' (self-reported) ability to reach absorptive listening states. Secondly, we investigated whether harmonic structure has an influence on listeners' enjoyment, further validating the results from Agres et al. (2017) by using longer, more ecologically valid musical excerpts. Finally, we hypothesised that the enjoyment of UT music and ability of listeners to reach absorptive states are associated.

The results of the experiment confirm each of the above hypotheses. More specifically, our findings indicate that the structure of chord progressions influences both listeners' enjoyment of the music and their sense of being absorbed in the music. UT excerpts that are most enjoyed are those which tend to elicit absorptive listening states in listeners.

Although our results also support a meaningful connection between harmonic structure and absorptive listening states, further research is warranted on this topic. In our experiment, four of the participants reported not having previously experienced absorptive listening states while listening to trance, which we assume may have influenced their AL state ratings to some extent. Furthermore, Haerlin (1998) has suggested that many listeners need around 13–15 minutes in order to fully reach AL states through auditory driving (in their case, drumming). Although subjective reports differ, this important point of giving the participants the necessary time to induce an absorptive listening state should be carefully considered in the future. Nevertheless, of the studies that explore which specific musical characteristics have an impact on trancing and absorptive listening states, the majority focus on rhythmic cues and auditory driving effects (Becker-Blease, 2004; Becker, 2012; Fachner, 2011; Neher, 1962; Trost et al., 2014). This is the first work, to the authors' knowledge, to confirm the meaningful connection between harmonic structure in UT music and absorptive listening states.

Acknowledgements. The authors would like to thank Associate Professor Louis Bigo from the Université Lille 3 for his help in creating the stimuli for this study. This research was partially supported by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859, as well as the SRG ISTD 2017 129 grant.

References

Agres, K., Herremans, D., Bigo, L., & Conklin, D. (2017). Harmonic structure predicts the enjoyment of uplifting trance music. *Frontiers in psychology*, 7, 1999.

Agres, K., Bigo, L., Herremans, D., & Conklin, D. (2016). The effect of repetitive structure on enjoyment in uplifting trance music. In *Proceedings of the 14th International Conference for Music Perception and Cognition (ICMPC)*. San Francisco, CA. pp. 280-282.

Becker, J. (2012) Rhythmic entrainment and evolution. In J. Berger & G. Turow (Eds.), *Music, science, and the rhythmic brain: Cultural and clinical implications (1)*. New York: Routledge., pp. 49-72.

Becker-Blease, K. A. (2004) Dissociative states through new age and electronic trance music. *Journal of Trauma and Dissociation*, 5(2), pp. 89-100.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics*, 8(5), 279-286.

Fachner, J. (2011). Time is the key–music and altered states of consciousness. In Cardeña E. and Winkelman M. (Eds.), *Altering consciousness: A multidisciplinary perspective, (1*) Connecticut, US: ABC-CLIO/Greenwood Publishing Group, pp. 355–376.

Gordon, J., & Gridley, M. C. (2013) Musical preferences as a function of stimulus complexity of piano jazz. *Creativity Research Journal*, 25(1), pp. 143-146.

Haerlin, P. (1998) Bewußtseinsverändernde Klanginstrumente in der Psychotherapie. *Psychotherapeut*, 43(4), pp. 238-242.

Herremans, D., & Chew, E. (2016). Tension ribbons: Quantifying and visualising tonal tension. Proceedings of the Second International Conference on Technologies for Music Notation and Representation (TENOR), Cambridge, UK.

Kakouri, K. (1965). Dionysiaka; Aspects of the Popular Thracian Religion of Today. *Trans. H. Colaclides*. Athens: G. C. Eleftheroudakis.

Kartomi, M. J. (1973) *Music and trance in central Java*. Ethnomusicology, 17(2), pp. 163-208.

Ludwig, A. (1966) Altered States of Consciousness. *Archives of General Psychiatry,* 15(3), pp. 225-234.

Madrid, A. L. (2008) *Nor-tec rifa!: Electronic dance music from Tijuana to the world*. New York: Oxford University Press.

Neher, A. (1962) A physiological explanation of unusual behavior in ceremonies involving drums. *Human Biology*, 34(2), pp. 151-160.

Orr, M. G., & Ohlsson, S. (2001) The relationship between musical complexity and liking in jazz and bluegrass. *Psychology of Music*, 29(2), pp. 108-127.

Silverman, J. (1968) A paradigm for the *study of altered states of consciousness*. *The* British Journal of Psychiatry, 114(515), pp. 1201-1218.

Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., & Vuilleumier, P. (2014) Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage*, 103, pp. 55-6

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Brawlers, Bawlers, and Bastards: Vocal Timbre as a Marker of Recurring Archetypal Characters in the Music of Tom Waits

Joshua Albrecht[1]

[1] *Music Department, The University of Mary Hardin-Baylor, United States*

[1]jalbrecht@umhb.edu

Abstract

A wide range of very distinctive timbres and a high degree of theatricality characterizes Tom Waits' music. Several scholars have identified a tendency for Waits to portray fictional characters and personas in his songs, rather than the more typical approach of presenting his songs as semi-autobiographical. This study explores the hypothesis that Waits' songs can be categorized by vocal timbre into a relatively small number of subsets, and that these groups of songs represent recurring archetypal characters in Waits' output. To test this hypothesis, participants grouped 40 five-second excerpts randomly sampled from 146 of his latter-career output into as many as eight groups, and then progressively linked the most similar groups. They also provided descriptions for each of the groups. These similarity measures were used to estimate the distances between each of the songs in Waits' late career. Cluster analysis suggests the presence of 7 distinct clusters, and these clusters are supported by different timbral descriptors. Further content analysis is performed on the lyrics of these songs by clusters, but no significant ANOVAs suggest that if there are distinct characters in Waits' songs, they cannot be detected only by sentiment analysis of the lyrics.

Introduction

Tom Waits' music is typically off-kilter and defies easy categorization. His music can be characterized by a number of distinctive narrative and musical techniques that are unusual in popular music. Many of his songs incorporate irregular hypermeter (Thomas, 2016), nontraditional noise sounds and the blurring of formal boundaries (Jones, 2007), and the use of unreliable narration (Hoskyns, 2009).

Probably the most immediately recognizable aspect of Tom Waits' music is his distinctly rough vocal timbre(s). As Solis (2007, 41) writes, "it is common, to the point of meaninglessness, to say that Waits is possessed of a deep, gravelly, 'whisky' voice." Instead, Solis argues that Waits uses many different voices, each distinct and recognizable. Moreover, many songs share similar vocal timbres, suggesting links or shared meaning between them.

Another important feature of Waits' music is its deep theatricality. From early in his career, there have been strong interactions between Waits' music, his life, and theatrical elements (Hoskyns, 2009). This is most clearly reflected in four albums of music composed for the stage: *Franks Wild Years, The Black Rider, Alice,* and *Bone Machine*. Yet the portrayal of characters permeates his entire output. Unlike much of rock music since the 1960s, which claims some form of implicit autobiography as the voice of the singer/songwriter (Starr & Waterman, 2003), Waits' songs are (often overtly) inhabited by fictional personas (see BaileyShea, 2014) who speak in distinct voices, both figuratively (lyrical meaning) and literally (unique vocal timbres).

The interrelationship between theatricality, fictional personas, and timbral distinctions results in a high degree of intertextuality. Even on albums that are not explicitly song cycles, there are frequent textual and musical cross-references and quotations. For example, the word 'moon' appears in 72 of his 255 studio songs, and 18 of his 19 studio albums. The high degree of intertextuality has ramifications for interpreting the meaning of the songs.

Hypothesis

In this paper, I explore the hypothesis that there are a relatively small number of character archetypes in Waits' music, whose songs are marked by a high degree of intertextuality and shared vocal timbre. These characters are portrayed as fictional persona types throughout his output. I examine this hypothesis using an exploratory technique that combines perceptual grouping of song excerpts by vocal timbre with textual analysis.

Method

It might seem that the easiest method of testing the hypothesis that there is a small number of timbral groupings in Waits' songs would be to use automatic feature extraction to examine timbral features of his music, but there are a number of methodological drawbacks to this approach. In the first case, the cognitive perception of timbre is a complicated, multi-faceted set of auditory attributes that do not easily map onto acoustic properties (McAdams & Giordano, 2008), so that even similarities within acoustic parameters between songs may not mean that these songs would be perceived as timbrally similar. Secondly, at the time of this project, I did not have access to the original mastered recordings, so the tracks examined consisted of Waits' voice along with instrumentation. Not being able to isolate the voice, any similarities obtained by automatic feature extraction alone may be influenced by similarities in instrumentation or style.

Ultimately, to understand perceptions of timbre, there is no substitute for human ratings. Therefore, this study employed an empirical approach to study perceptions of timbral similarity using human participants and a corpus study of lyrics to examine intertextuality.

Sample

Tom Waits' entire output is expansive: 255 vocal tracks over 19 studio albums, dozens of instrumental tracks, live albums of concerts, and soundtracks to movies, among other recordings. This study's focus is the way Waits represents personas in his music, so it was important to limit the sample to that music over which he had the most artistic control, something not always possible when writing soundtracks for another director's vision. Other constraints beyond artistic vision may influence the way Waits uses vocal timbre in live music, such as his health at the time of recording, the rigor of his touring schedule, the acoustics or recording equipment of the venue, or the physical proximity of his audience. For the purposes of this study, therefore, only Waits' studio albums were selected, and instrumental tracks were not used.

However, 255 tracks were still deemed too many to be practical for this study, and so only those tracks appearing on *Swordfishtrombones* and later were used, but not counting his compilation album *Orphans*. This album was chosen as the starting point of this study because it marks a sizeable shift in Waits' career and style. It is the first album after Waits was married to Kathleen Brennan, who had a strong influence on his artistic vision and goals. This album is also the beginning of a marked shift toward very distinct vocal timbres, a feature that could be seen in nascent form earlier but was not as pronounced until *Swordfishtrombones.* After filtering Waits' output in this way, 146 total vocal tracks remained.

For the purposes of timbral comparisons, it was only important to sample brief excerpts from each of the remaining 146 tracks. Prior research has shown that listeners can extract a remarkable degree of information out of very short excerpts (Gjerdingen & Perot, 2008). Plazak and Huron (2011), for example, found that instrumentation could be discerned in as little as 100 ms and the gender of the voice could be discerned in as little as 800 ms. In addition, to aid in the comparison of excerpts, it was important to keep the excerpts short. It was therefore deemed that five-second excerpts were long enough. Consequently, the first five-second excerpt of each song was sampled in which Waits' voice could be heard for at least four seconds of it.

Lyrics corpus

To examine inter-textual links between songs, the lyrics for each song were downloaded from tomwaitslibrary.info, an extensive fan-made repository of Tom Waits information. This source was chosen because the liner notes included with the albums sometimes have misprints or leave out the texts of some songs entirely, most often with his spoken word songs. The fan page appears to be painstaking transcriptions of the words actually used in the recordings.

The texts for each song were analyzed using two lexicons. The first lexicon was the NRC Word-Emotion Lexicon (NRC). This lexicon consists of 14,182 words categorized into positive and negative sentiments and the emotions anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The R package 'tidytext' was used to analyze Waits' song lyrics using NRC.

The second content analysis program used was the Linguistic Inquiry and Word Count program (LIWC2015), consisting of over 90 variables that measure properties of texts (Pennebaker *et al*., 2015). Specifically, LIWC2015 measures the degree to which texts suggest logical argument, the relative social status of the text, levels of authenticity in the text, the overall positive or negative emotional tone of the text, the degree of anxiety, anger, sadness, and positive emotion of the text, social words, affiliation, achievement, power, reward focus, and risk prevention focus, among dozens of other dimensions.

Participants

A total of 71 undergraduate students at the University of Mary Hardin-Baylor were recruited for the task involving sorting songs by vocal timbre. 44 of the participants were female, and 27 were male, with a mean age of 21.0 (sd = 6.9). 54 participants reported Rock music as the musical genre they mainly listen to, with 12 reporting classical music, and 5 reporting jazz as their preferred genre. The mean Goldsmith's Musical Sophistication Index (Müllensiefen, *et al.,* 2014) for general musical sophistication was 96.9 out of 126 total (sd = 12.0). Average scores of subsets of the Gold-MSI measure included 47.4 out of 63 for active engagement (sd = 6.8), 51.1 out of 63 for perceptual abilities (sd = 6.3), 35.4 out of 49 for musical training (sd = 7.3), 35.2 out of 42 for emotions (sd = 4.6), and 36.2 of 49 for singing abilities (sd = 6.2).

Procedure

The primary purpose of the human participants phase of this study was to get perceptual data about how similar and dissimilar each excerpt was to all the others. Ideally, each participant would group together all 146 of the excerpts by timbral similarity. Unfortunately, the sheer number of comparisons needed between the entire dataset was prohibitive. Instead, participants were given a randomly selected subset of 40 of the 146 excerpts.

In a study of sorting methodologies, Giordano, *et al.* (2011) found that there were advantages to both free sorting, in which participants sorted into however many groups they deemed most relevant, and hierarchical sorting, in which the most similar items were grouped successively until there were only two groups left. However, free sorting suffered from unreliability and hierarchical sorting suffered from its overly time-consuming nature. Instead, they found that a truncated hierarchical sort, with *n* / 5 possible groups was enough to provide sufficient resolution without taking too much time.

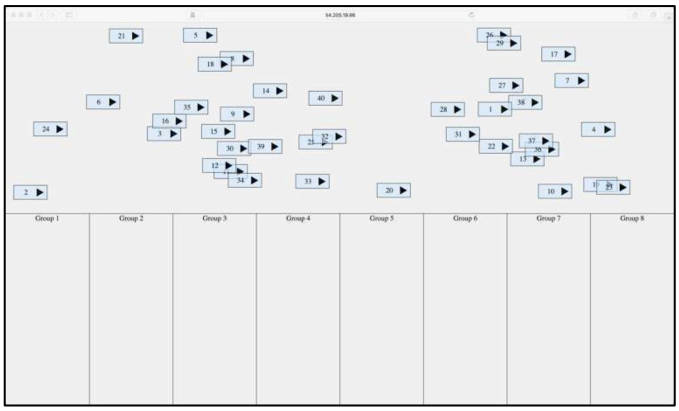


Figure 1. The sample interface. 40 randomly selected excerpts were scattered across the top of the screen, and participants dragged them into as many as eight categories on the bottom of the screen.

Therefore, for this study I used a truncated hierarchical sort, which happened in three phases. Because there were 40 excerpts, participants could sort them into as many as eight categories (or fewer if deemed appropriate) in the first phase. At the beginning, the 40 excerpts were scattered randomly across the top of the screen (see Figure 1 for a sample screenshot). Before sorting could start, participants listened to all 40 excerpts in numerical order as they lit up to indicate which excerpts were playing. Once heard, participants were free to move the excerpts around and into whatever categories seemed most appropriate, and they could click on any excerpt as many times as they wished to hear it again.

For this study, participants were explicitly told to ignore the instrumentation of the excerpts, the texture, the genre, the meaning of the words, and other factors, and to *only* group excerpts by vocal timbre. They were told to group excerpts that were most similar and to ensure that excerpts that were the most different from one another were sorted into different groups.

After the initial sort, participants were presented with each of their assembled groups in the second phase, listening to all excerpts in the group in random order. They were then asked to provide a description of the vocal timbre used in that group. In the third phase, each group appeared with the description they provided in the second phase. Participants again listened to groups by clicking on them, and joined the two that were most similar. Once joined, they again listened to groups and selected the two that were most similar. Joining continued until there were only two groups remaining.

Results

Participant grouping responses

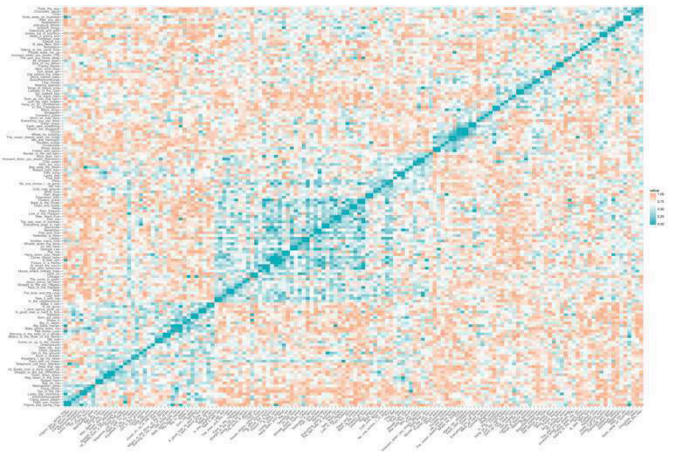


Figure 2. Dissimilarity matrix for the participant grouping responses for the entire dataset. Maximum dissimilarity is in red and minimum dissimilarity is in blue.

Participant data were used to estimate the timbral similarity between each pair of excerpts. The similarity measures were calculated by dividing the number of times two excerpts were grouped together at any stage of the truncated hierarchical grouping by the number of times they *could have been* grouped together. So, if two excerpts were always grouped together every time they could have been, they would have a similarity of 1, and if they were never grouped together every time they could have been, they have a similarity of 0. For example, "I'll Shoot the Moon" and "Another Man's Vine" appeared in the same study four times, in which there were 26 total grouping stages. Of those, they were grouped together only 9 times, meaning that the similarity between the two was 9 / 26 = .346, or 34.6%. Every combination of excerpts appeared together in the same experiment at least once. To determine how many distinct groups naturally occurred in the data, cluster analysis was employed. Cluster analysis is driven by *dis*similarity data, so the similarity scores were subtracted from 1. Therefore, if two excerpts were never grouped together, their dissimilarity rating was 1. The dissimilarity matrix for all 146 excerpts appears in Figure 2. Maximum dissimilarity is shown in red and minimum dissimilarity is shown in blue.

Determining the optimal number of clusters in a dataset is not a trivial task. A number of metrics have recently been proposed to try to minimize the role of intuition, though most statisticians acknowledge that researcher intuition remains an important part of interpreting grouping. The gap statistic is a test of how many clusters are most appropriate in a given dataset by comparing the total within-cluster variation for different numbers of clusters against the expected values under null reference distribution of the data (Tibshirani *et al.,* 2001). Using a gap statistic, the optimal number of clusters appears to be 7 clusters (Figure 3). A hierarchical clustering dendrogram can also be a useful way to visualize how many distinct clusters seem appropriate for the data. A dendrogram showing Ward's method appears in Figure 4, with the 7-cluster solution shown. This solution appears robust, with only a 3-cluster solution appearing to have more distance between cluster heights. As a final check on the 7-cluster solution, silhouette scores are plotted in Figure 5. As can be seen in the Figure, the average silhouette scores are not high, but there are only 19 excerpts with negative silhouette widths, suggesting the presence of weak clustering.

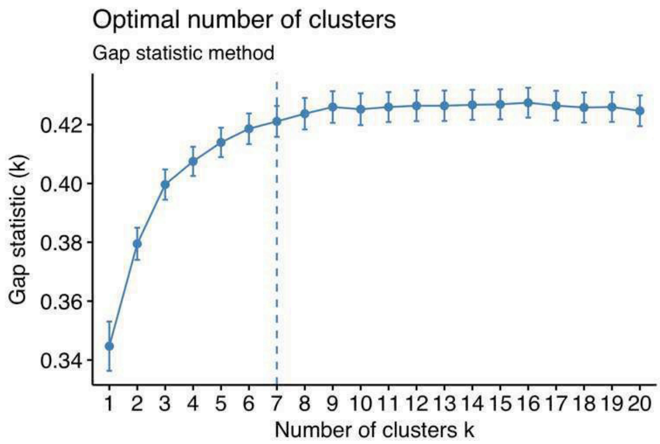


Figure 3. Optimal number of clusters according to the gap statistic. The best number of clusters appears to be between 7.

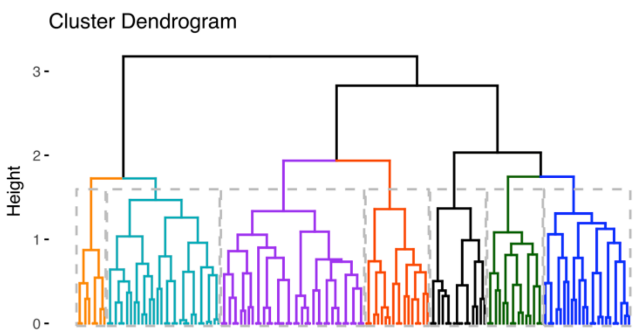


Figure 4. A 7-cluster dendrogram using Ward's method.

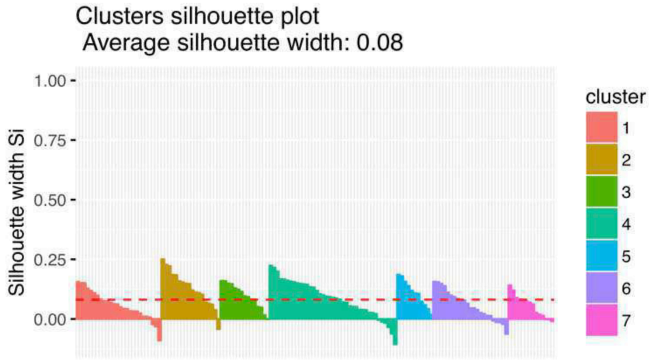


Figure 5. A 7-group hierarchical circle cluster using Ward's method

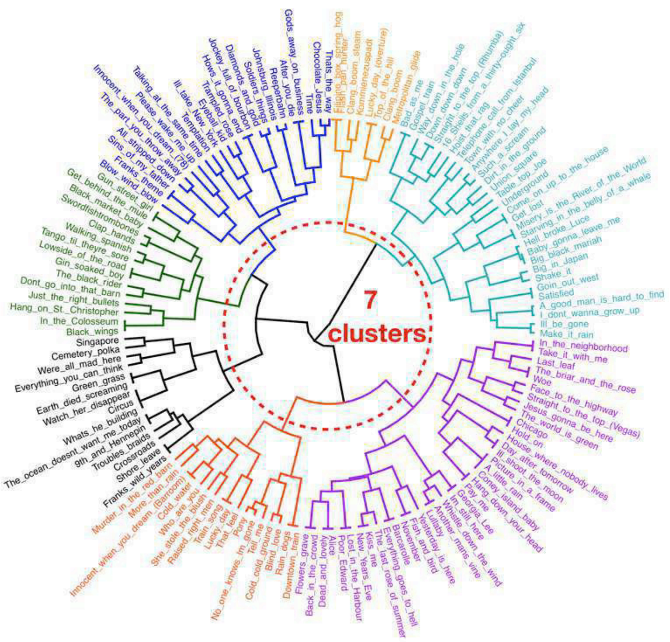


Figure 6. Clustering results for 7 clusters.

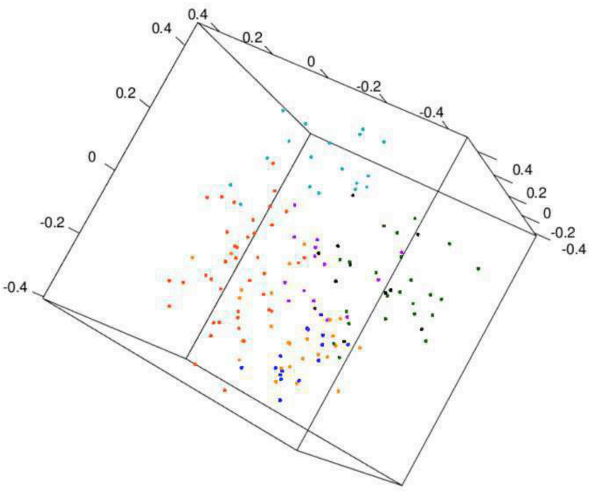


Figure 7. 3-dimensional multi-dimensional scaling for k=7 k-means clustered data.

The excerpt names for the hierarchical clustering using Ward's method are shown in Figure 6. Here, the dendrogram is represented circularly for reasons of space. A k=7 k-means cluster analysis represented using multi-dimensional scaling in three dimensions is shown in Figure 7. Again, clustering is apparent, but weak.

Participant timbre descriptions

Once initial groups were established, each participant provided a description of the timbral quality of the voice for each group as free responses. These free responses were cleaned by removing all words that were not descriptors. Generally, adjectives were kept and other words removed. Modifiers of the adjectives, like "really," "very," or "kind of" were discarded, and compound descriptors were hyphenated, such as in "not-deep." When appropriate, nouns were turned into adjectives, so "breathiness" was made into "breathy." When there were two adjectives in different forms, the simpler form was retained, so "stranger" was changed to "strange." References to specific proper nouns were retained and hyphenated, as in "Muppets" or "Frank-Sinatra." Finally, spelling mistakes were corrected and all punctuation was removed.

After cleaning the timbral descriptors, there remained 6,124 total terms and 407 unique terms describing Wait's vocal timbres. The twelve most common terms overall are shown in Table 1. The table reads like an album review for one of Waits' albums. Of note is that raspy is far and away the most common descriptor for Waits' music, appearing a remarkable 9.5% of the total number of terms used.

Table 1. The most common timbral description terms overall

Most common terms (146 excerpts)	Number of appearances
raspy	582
deep	181
smooth	168
rough	160
breathy	158
growly	128
soft	122
nasal	100
gravelly	95
scratchy	94
low	93
harsh	91

Every term that was connected with any of the excerpts in each of the k=7 k-means clusters was summed for each cluster. Rank-ordered lists for timbral descriptors by cluster are shown in Table 2. The number of excerpts in each cluster is shown parenthetically in the table heading. The number of appearances for each term is shown parenthetically beside the term. Unsurprisingly, 'raspy' is the most common timbral descriptor for every cluster, but the rank-ordering of the next most common terms reveals interesting differences between patterns of perceptions of vocal timbre for each cluster. For example, after raspy, C1 is described as deep, smooth, and breathy, whereas C6 is rough, growly, screaming, and harsh, and C4 is speech-like, deep, and low. The rank-ordered differences are consistent with distinct timbral characteristics for each cluster, and suggest consistent timbral portrayals of different vocal characters.

Table 2. The most common timbral description terms by cluster. Number of excerpts in each cluster is shown in parentheses in the header, and number of times each term appears in the cluster is shown in parentheses after the term.

C1 (33)	C2 (22)	C3 (15)	C4 (21)	C5 (16)	C6 (16)	C7 (23)
raspy (150)	raspy (83)	raspy (48)	raspy (93)	raspy (58)	raspy (52)	raspy (97)
deep (76)	nasal (45)	smooth (35)	speech-like (42)	rough (18)	rough (31)	growly (38)
smooth (74)	rough (29)	breathy (31)	deep (33)	deep (15)	growly (25)	rough (38)
breathy (63)	harsh (19)	soft (21)	low (29)	smooth (15)	scream-ing (21)	deep (27)
soft (45)	speech-like (19)	relaxed (16)	soft (27)	breathy (14)	harsh (15)	yelling (22)
low (38)	jazzy (18)	light (15)	breathy (24)	growly (14)	nasal (14)	gravelly (20)
scratchy (30)	light/ growly (13)	airy (11)	smooth (24)	heavy (13)	scratchy (13)	husky (17)

Textual analysis by cluster

The texts for each of Waits' songs were analyzed using NRC and LIWC2015, and the results of the analysis were grouped by the k-means clusters found in the participants' grouping data. Using NRC and LIWC2015 has a number of benefits. However, there are also drawbacks in using automated content analysis libraries on song lyrics. In the first case, song lyrics are often used poetic, and there is often play in the meaning of words or phrases in which less common meanings are employed or in which words relate to one another in complicated ways. Because word lexicons were designed to measure naturalistic language, not poetry, and they simply compare individual words against a library of terms, they cannot really take into consideration the context of the song. More seriously, lexicons are designed to analyze large texts of thousands of words, but the lyrics for these songs are much shorter (M = 232.4 words, sd = 111.0). As the authors of LIWC2015 contend, "the more words you analyze, the more trustworthy are the results," and short texts "should be looked at with a certain degree of skepticism."

A one-way ANOVA was performed to compare the effects for each of several parameters on cluster identity. Parameters examined for NRC included positive and negative sentiment, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. Parameters examined for LIWC2015 included analytical thinking, authenticity, emotional tone, clout, positive emotion, negative emotion, anger, sadness, anxiety, social words, cognitive processes, sexuality, core drives and needs, affiliation, achievement, power, reward focus, risk prevention focus, money, religion, death, and swear words, among many others. Of these, only core drives and needs ($F(1, 143) = 7.1$, $p = .008$) and power ($F(1,143) = 5.2$, $p = .02$) showed significant effects. Cluster means for core drives and needs are shown in Figure 8. After correcting for multiple tests using a Bonferroni correction, none of the parameters revealed significant differences at $p < .002$. Contrary to the results regarding perceptions of timbral features, these results are not consistent with the notion that Waits codes for character archetypes in the lyrics of the songs in timbral categories. It is likely that the crude approach of content analysis is unable to detect differences in song texts by cluster. Closer analysis of lyrics is likely needed to investigate textual links or differences between clusters.

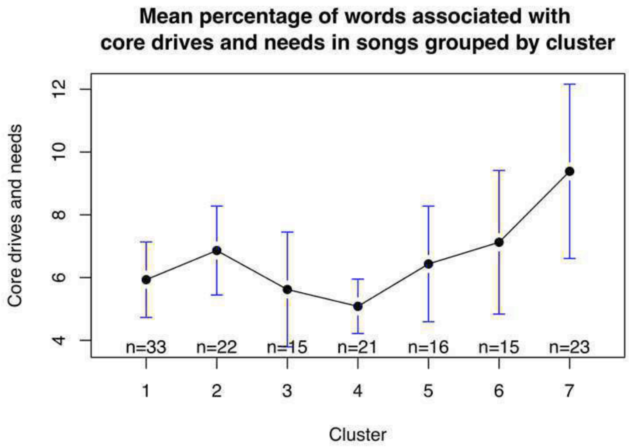


Figure 8. Mean percentage of words associated with core drives and needs in songs grouped by cluster

Conclusion

Participant grouping data for the corpus of studio songs comprising the latter half of Waits' career is consistent with the existence of a small number of groups based on the vocal timbre he employs in his songs. It appears that seven groups is an appropriate number of song subsets (see Figure 3), but there are also clear supersets; two or three clusters would be appropriate as well. Additionally, although seven clusters emerge, more refined subsets could be possible with further data collection. All songs were compared with all others, but a few pairs of the 146 songs only appeared in one study simultaneously, and several more pairs were only compared by two participants, limiting the power of detecting distances. A close examination of Figure 2 reveals the presence of dozens of small clusters of 3-5 songs that are highly similar, and with more data collection more distinct sub-clusters could possibly emerge within the dataset.

Although perceptual timbral clustering was consistent with distinct groups within Waits' output, the automated inter-textual analysis performed through sentiment analysis in LIWC2015 and the NRC emotion lexicon produced decidedly negative results. None of the categories examined produced significant differences between derived clusters, except for the borderline cases of "power" and "core drives and needs" (see Figure 8). If there are indeed distinct character archetypes present in Waits' songs that are linked to timbral differences, either these are not reflected in the texts of the lyrics, or they are not detectable using the crude tool of automated sentiment analysis. Given the delicate balance of expression and wordplay reflected in many songs, however, it is likely that a much better strategy would be close analyses of individual songs, rather than an automated approach. This is my plan for next steps in analyzing the data.

Nevertheless, the participant descriptions of vocal timbres by song groups are also consistent with distinctions between groups. Although specific character archetypes like "hobo," "circus act," or "used car salesman" cannot be gleaned from differences between descriptors, more general characteristics of clusters can be seen in Table 2. Given the large number of songs in the groups that do emerge (15-33), it may be the case that these reflect more generalized character types rather than specific character archetypes. The more specific character archetypes I originally hypothesized may still be there, but they may exist in smaller subsets of fewer songs each. For any listener familiar with Waits' output, it is undeniable that his skill-set includes many more than seven distinct vocal timbres.

Generally, however, these results are consistent with the existence of perceptual sub-clusters within Waits' songs based on vocal timbre. These results suggest new corners to examine, but are promising for the original hypothesis, even if the question of specific recurring personas or character archetypes must be left open for further investigation.

Acknowledgements. I would like to thank Danny Kirchmeier for his help in developing the software and online platform for the study's interface, and to the University of Mary Hardin-Baylor Faculty Development Grant for travel funding to ICMPC.

References

BaileyShea, M. L. (2014). From me to you: Dynamic discourse in popular music. *Music Theory Online, 20.4*. Retrieved from http://www.mtosmt.org/issues/mto.14.20.4/mto.14.20.4.baileyshea.html

Brackett, D. (1995). *Interpreting Popular Music.* Cambridge: Cambridge University Press.

Giordano, B.L., Guastavino, C., Murphy, E., Ogg, Mattson, Smith, B.K., & McAdams, S. (2011). Comparison of methods for collecting and modeling dissimilarity data: Applications to complex sound stimuli. *Multivariate Behavioral Research, 46*, 779-811.

Gjerdingen, R., & Perrot, D. (2008). Scanning the dial: The rapid recognition of musical genres. *Journal of New Music Research*, *37* (2), 93-100.

Heidemann, K. (2016). A system for describing vocal timbre in popular song. *Music Theory Online, 22.1*. Retrieved from http://www.mtosmt.org/issues/mto.16.22.1/mto.16.22.1.heidemann.html

Holmes, P. A. An exploration of musical communication through expressive use of timbre: The performer's perspective. *Psychology of Music, 40* (3), 301-323.

Hoskyns, B. (2009). *Lowside of the road: A life of Tom Waits*. New York: Broadway Books.

Jones, A. (2007). Musical Apocalypse: Tom Waits' *Bone Machine*. *FORUM: University of Edinburgh Postgraduate Journal of Culture and the Arts 5*.

McAdams, S., & Giordano, B., (2016). The perception of musical timbre. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford Handbook of Music Psychology* (pp. 113-124). Oxford: Oxford University Press.

Montandon, M. (Ed.) (2005). *Innocent when you dream: The Tom Waits Reader*. New York: Carroll & Graf.

Müllensiefen, D., Gingras, B., Steward, L., & Musil, J. (2014). The musicality of non-musicians: An index for measuring musical sophistication in the general population. *PLoS ONE, 9* (2), e89642.

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Plazak, J., & Huron, D. (2011). The first three seconds: Listener knowledge gained from brief musical excerpts. *Musicae Scientiae, 15* (1), 29-44.

Starr, L., & Waterman, C. (2003). *American Popular Music: From Minstrelsy to MTV*. Oxford: Oxford University Press.

Solis, G. (2007). "Workin' hard, hardy workin'/ Hey Man, you know me": Tom Waits, sound, and the theatrics of masculinity. *Journal of Popular Music Studies, 19* (1), 26-58.

Thomas, M. E. (2016). Text and temporality: Toward an understanding of rhythmic irregularities in the music of Tom Waits. *Music Theory Online 22.4*. Retrieved from http://mtosmt.org/issues/mto.16.22.4/mto.16.22.4.thomas.html

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B, 63* (2), 411-423.

Tsai C., Wang, L., Wang S., Shau, Y., Hsiao, T., & Auhagen, W. Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception 27* (4), 209-221.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Extracting majorness as a perceptual property of music

Anna Aljanaki,[1] Gerhard Widmer[2]

Institute of computational perception, Johannes Kepler University, Austria

[1]aljanaki@gmail.com, [2]gerhard.widmer@jku.at

Abstract

For the tasks of automatic music emotion recognition, genre recognition, music recommendation it is helpful to be able to extract mode from any section of a musical piece as a perceived amount of major or minor mode (majorness) inside that section, perceived as a whole (one or several melodies and any harmony present). In this paper we take a data-driven approach (modeling directly from data without giving an explicit definition or explicitly programming an algorithm) towards modeling this property. We collect annotations from musicians and show that majorness can be understood by musicians in an intuitive way. We model this property from the data using deep learning.

Introduction

With Western popular tonal music, the term "mode" is often used dichotomously to refer only to major (Ionian) or minor (Aeolian) mode (omitting the rest of the modes, such as harmonic minor, blues scales, etc.). For a certain combination of harmony and melody, labeling an excerpt of music as "major" or "minor" can be subjective, ambiguous, or even impossible (especially when modulations and/or key signature changes are present inside the segment). The tonal hierarchy also needs time establish itself (Parncutt, 1989) and the perception of the tonal centroid and mode may change while this is happening. In music information retrieval context, the category of mode can sometimes be treated probabilistically (e.g., as a probability of an excerpt being in major mode, as predicted by an algorithm), resulting in a continuous property (Saari, 2011) and (Friberg, 2011).

In this paper we will call this property majorness, following (Parncutt, 1989) and MIRToolbox (Lartillot, 2008). There exist music analysis tools that permit to extract it, such as MIRToolbox and QM Vamp Plugin (Noland, 2007). The algorithms implemented in these tools rely on a pitch class profile based estimation of key, and produce a result that is not exactly similar to a perceptual estimation of majorness, as annotated by musicians (Friberg, 2011).

The concept of majorness lacks a clear musicological definition, and therefore it is difficult to design an algorithm to extract it in the same straightforward way as an algorithm for onset detection can be designed. However, given enough training data, it is possible to learn the property directly from the data. In this paper we investigate this approach.

Methods

In this section, we will first describe our data collection approach (a hybrid of pairwise comparison and absolute ranking), and then describe the deep learning method that was used to create a model of majorness.

Data collection

It is very difficult for an annotator to rate a vaguely defined and subjective concept such as majorness on an absolute scale (Madsen, 2013). Comparing two examples given a certain criterion is an easier task. However, pairwise comparisons require factorially (in relation to the number of examples) more ratings, as compared to linear number of absolute ratings, even when using only a part of the full comparison matrix (Madsen, 2013). In order to learn majorness from data, we need to annotate at least several thousand song excerpts, this amount of songs would require millions of pairwise comparisons, which is prohibitively expensive.

We decided to combine the two approaches and first create a scale using pairwise comparisons, and then collect absolute ratings on that scale.

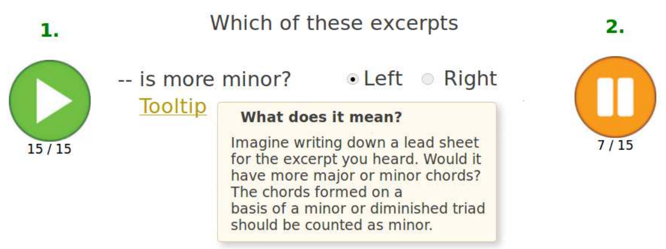


Figure 1. Pairwise comparisons interface with a hovered tooltip (translated from a Russian interface on toloka.yandex.ru).

Pairwise comparisons

On a crowd-sourcing platform we hire 80 musicians (5 per pair) to compare pairwise 100 musical excerpts of 15 seconds on their majorness. We get the music from creative-commons licensed websites and chose 100 songs from different genres and with different valence/arousal values. Figure 1 shows the interface that was shown to the annotators on the Toloka crowd-sourcing platform (toloka.yandex.ru). From the pairwise comparisons, we obtain a ranking of pieces from the most certainly minor ones, through the ambiguous ones, to the most certainly major ones.

Absolute rankings

From the ranking we sample 10 excerpts as examples and collect ratings of perceived majorness for 5000 excerpts, also belonging to various music genres (rock, pop, classical, jazz, blues, etc.). Figure 2 shows the interface that was used to collect these ratings. An annotator compares a piece to an example, and if a piece is more minor than an example, listens to the next example to the right, until the current piece can be placed between two examples (it's more minor than example to the left, and less minor than example to the right). In this way, absolute ratings from 1 to 10 are obtained.

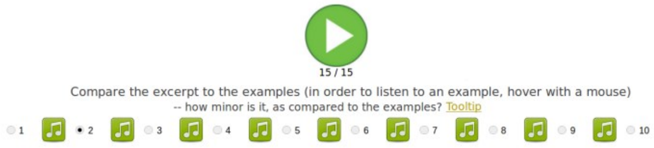


Figure 2. An interface for absolute scale ratings (translated).

Deep learning model

From every musical excerpt in the dataset we extract a mel-spectrogram with 299 mel-filters with a half-overlapping Hanning window of 2048 (44.1k sampling rate). We train a fully convolutional neural network (Inception architecture) with a mean squared error loss (regression task) on the averaged absolute ratings. More details about the model can be found in (Aljanaki, 2018).

Results

Data

The consistency of the annotations without any unreliable rater filtering is 0.69 Cronbach's alpha (0.33 Krippendorff's alpha). These valuse indicate low consistency. To improve the annotations, some of the annotators were removed based on their disagreement with the rest. Figure 3 shows a histogram of the resulting annotations. The data is normally distributed, showing that the annotators avoided the extremes (completely major and completely minor).

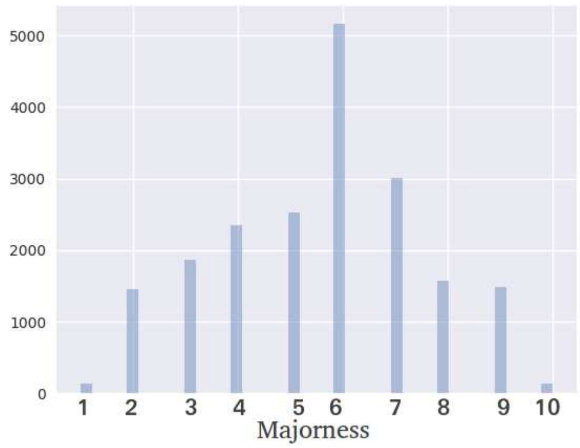


Figure 3. Distribution of the annotations for majorness.

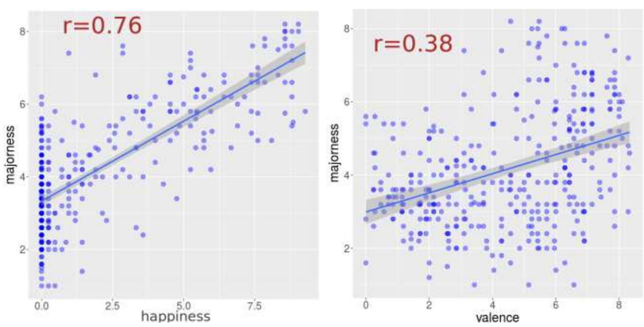


Figure 4. Correlation of majorness with valence and happiness.

It has been shown that majorness is a useful feature for predicting emotion in music (Gabrielsson, 2001). We included the songs from a soundtracks dataset annotated with emotion, both dimensional (valence and arousal) and categorical (five basic emotions, which are used both as categories and as dimensions) (Eerola, 2011). Figure 4 shows the correlation of majorness as annotated in our dataset with valence and happiness (used as a dimension) on the 360 songs from the soundtracks dataset. There is a strong correlation with happiness, which is expected from majorness. The correlation with valence is less strong.

Predicting mode on WTC

An Inception model trained on this data, as explained before, could predict majorness on the test set with a Pearson's correlation of 0.48. In case of a neural network, it is difficult to understand, what exactly the model has learned. In order to better understand it, we used a collection of pieces in different tonalities – the Well-Tempered Clavier (WTC) by J.S. Bach (both books). The WTC contains 96 preludes and fugues, 48 major and 48 minor ones. We used recordings of Glenn Gould's performances and extracted the mel-spectrograms from the first 12 seconds of each prelude or fugue.

With a continuous majorness feature predicted by the model as an independent variable, we train a logistic regression to predict binary major and minor mode. This model has 70% accuracy when 10-fold cross-validated (random baseline is 50%). Clearly, what a model has learned is related to mode, but not exactly mode.

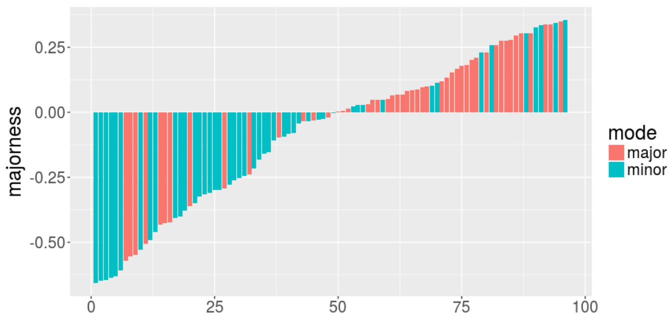


Figure 5. Majorness vs major and minor mode in WTC.

Figure 5 shows how the algorithm performed on minor and major pieces from WTC. The order from left to right is not corresponding to the order of the pieces in WTC, but to the majorness predicted by the neural network. By examining the mistakes (red rows below the x axis, and blue rows above the x axis), we can see that slower and more pensive performances of major pieces were classified as minor (prelude from BWV852, fugue from WV892), and faster, more energetically performed minor pieces were classified as major (prelude from BWV875, prelude from BWV871).

Conclusion

Even without being given a formal definition of majorness, musicians could somewhat intuitively understand this property, and agree on it when annotating music. The property that they annotated turned out to be very close to valence (happiness) dimension of musical emotion. Using this data, we trained a neural network to predict majorness from musical audio. It appears that except for mode itself, a neural network also takes into account other (perhaps, more performative) aspects.

Acknowledgements.
This work was supported by the European Research Council (ERC) under the EUs Horizon 2020 Framework Programme (ERC Grant Agreement number 670035, project "Con Espressione"). This work was also supported by FCS grant.

References

Saari, P., Eerola, T., & Lartillot, O. (2011). Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(6), 1802–1812

Friberg, A., Hedblad, A. (2011) A Comparison of Perceptual Ratings and Computed Audio Features. Proceedings of the SMC 2011, 122–157

Lartillot, O., Toiviainen, P., Eerola, T. (2008) A matlab toolbox for music information retrieval. Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization

Noland, K. and Sandler, M. (2007) Signal Processing Parameters for Tonality Estimation. In Proceedings of Audio Engineering Society 122nd Convention, Vienna

Madsen, J., Jensen, B. S., & Larsen, J. (2013). Predictive Modeling of Expressed Emotions in Music Using Pairwise Comparisons (pp. 253–277). Springer, Berlin, Heidelberg.

Parncutt, R. (1989). Harmony: A psychoacoustical approach. Springer series in Information Sciences, Vol. 19

Eerola, T. & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. Psychology of Music, 39(1), 18-49.

Aljanaki, A. & Soleymani, M. (2018). A data-driven approach to mid-level perceptual musical feature modeling. In Proceedings of the 19th International Conference for Music Information Retrieval

Gabrielsson, A. & Linstrom, E. (2001) The Influence of Musical Structure on Emotional Expression. Music and Emotion:
Theory and Research, (pp. 223-248) Oxford University Press

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Role of Music Performer Gesture in Creating Expressive Sounding Music

Emma Allingham,[1] Geoff Luck[2]

[1]*Department of Music, Art and Culture Studies, University of Jyväskylä, Finland*

[2]*Department of Music, Art and Culture Studies, University of Jyväskylä, Finland*

[1]emmaallingham21@gmail.com, [2]geoff.luck@jyu.fi

Abstract

Music performer gesture is associated with expressive intent, artistic interpretation of the music, and visual communication of emotions. Embodied approaches to music cognition suggest that body movements are important to the musician's process of creating expressive music, but whether different approaches to body movement during performance result in music that sounds more, or less expressive has not, so far, been empirically tested. To explore this, the current study investigated the importance of violinists' approaches to bodily movement in the creation of expressive sounding music. Three violinists performed 8 melodies, each intended to convey one of the emotions *happy, sad, tender,* or *scary*, under the performance conditions *immobile* (as little movement as possible) and *visually expressive* (expressing the music visually). Listeners then rated perceived expressivity and perceived discrete emotions in each performance on a 7-point linear scale. First, the validity of the emotion labels given to the melodies were tested using listener ratings of emotion, and 6 of the 8 melodies were confirmed as unambiguously conveying the intended emotion. Analysis of expressivity ratings revealed that the *immobile* condition significantly reduced ratings of perceived expressivity only in the two *sad* melodies, and significantly increased ratings of perceived expressivity in one *happy* melody. The results suggest that a performer's approach to expressive body movement can affect perceptions of audible expressivity, but that this effect is influenced by various factors such as the emotional content and technical difficulty of the music, and individual performer differences

Introduction

The topic of bodily gesture in music performance is a relatively young but growing area of research in music psychology. Most previous research in this field has focused on examining the visually communicative qualities of gestures, and the relationship between performers' body movements and their expressive intentions. For example, musicians seem to manifest expressive intentions in gesture by moving more when playing expressively and less when playing inexpressively (e.g Davidson, 1993), and by moving more when focusing on expressing an emotion than when focusing of feeling an emotion (Van Zijl & Luck, 2013). In addition, patterns in musicians' gesture show that body movements occur consistently and systematically in relation to structural features of the music and the musicians' interpretation (Demos, Chaffin, and Logan, 2017; MacRitchie, Buck, & Bailey, 2013; Wanderley, 2002; Wanderley, Vines, Middleton, McKay, & Hatch, 2005). It has also been shown that audiences have a visual bias when judging Western classical music performers (Tsay, 2013), and that performer gesture can visually communicate musical emotions, expressive intentions, musical phrasing and musical tension (Broughton & Stevens, 2009; Dahl, & Friberg, 2007; Vines, Wanderley, Krumhansl, Nuzzo, & Levitin, 2004; Wanderley et al., 2005).

Furthermore, embodied approaches to music cognition suggest that performers' body movements are important in the cognitive process of creating an expressive music performance. Sloboda (1996) suggested that musicians might use bodily gestures and feelings as a way of storing representations of expressive musical playing, while Juslin (2003) proposed that principles of biological motion are one of five factors that comprise musical expressivity. Furthermore, Juslin, Friberg and Bresin (2001) implemented this idea in their computational model of expressive music performance. In relation to this, two previous studies have explored the effects of suppressing performer gesture on the music produced, by asking performers to undergo an *immobile* playing condition. While Wanderley et al. (2005) found that the *immobile* condition disrupted performers' "sense of global timing" (p.101), Thompson and Luck (2012) found no such effect. Thus, the impact of suppressing a performer's natural bodily gesture on the music they produce is a ripe topic for further exploration. No previous study has empirically tested the effects of different approaches to performer body movement on listener ratings of audible expressivity.

To this purpose, the current study asks the following research question:

What is the role of musicians' expressive body movement in the process of creating an expressive sounding music performance?

This question was addressed by recording 3 violinists performing 8 short melodies, under the performance conditions "*immobile*" and "*visually expressive*". Each melody was intended to convey one of the emotions *happy, sad, tender* or *scary*, and the resulting audio recordings were presented to listeners, who rated the expressivity, and emotional content of the performances. The emotional content ratings were included to test the validity of the emotion labels given to the melodies, and to consider how the emotional content of the melody might mediate effects of performance condition on expressivity ratings. The hypotheses were:

H1: Each melody will yield significantly higher emotion ratings for the intended emotion, compared to the other three emotion ratings.

H2: There will be an effect of performance condition on listener ratings of audible expressivity.

This study aimed to inform music performance practice and pedagogy, while contributing to the discussion of the embodiment of expressive musical performance.

Methods

Participants

Forty participants (mean age = 26.4, *SD* = 6.88), were recruited via word of mouth, and social media advertising. All participants were either students or staff of the University of Jyväskylä, and were of varying nationalities. Thirty-one participants reported being able to sing or play a musical instrument to some extent, while nineteen considered themselves to be musicians.

Stimuli

The stimuli were audio recordings of performances by 3 violinists. The violinists performed 8 melodies under 2 performance conditions: *immobile* (keeping as still as possible while still playing expressively) and *visually expressive* (showing the musical expression visually while also taking care of the musical sound). The melody set included 4 short melodies, composed for scientific purposes (Thompson, Vuoskoski, and Clarke, 2016; Vieillard et al., 2008), and 4 longer melodies, chosen from classical violin repertoire. Each violinist performed all melodies, under both performance conditions, resulting in 48 performances in total. Audio was recorded using instrument mounted microphones (DPA, d:vote, 4099), and edited and mixed using ProTools (version 11.0.3) software.

Procedure

Audio stimuli were presented to listeners using Max/MSP software. Participants undertook the listening task individually, listening through AKG K141 Studio headphones, in a quiet room. Participants rated the expressivity (*not expressive at all – very expressive)* of each performance, as well as the perceived emotional content for *happy, sad, tender* and *scary* (*absent – present*), on 7-point linear scales. Stimuli were presented in a randomised order, and 8 seconds of natural forest sound were played between each stimulus to minimise carryover effects.

Analysis

Statistical analyses were conducted using SPSS software. Two factorial repeated measures ANOVAs were carried out; one for the emotion ratings, and one for the expressivity ratings. For the emotion ratings, the factors entered were emotion (4), performer (3), and performance condition (2). For the expressivity ratings, the factors entered were performance condition (2) and performer (3). Each melody was entered as a repeated measure.

Results

Emotion Ratings

The aim was to validate the perceived emotional content of the melodies used in the experiment by testing Hypothesis 1. Listeners gave a rating for each emotion category *sad, happy, tender* and *scary,* for each performance. If a melody clearly conveyed the intended emotion it was expected that there would be a significant main effect of emotion for that melody, and that pairwise comparisons would show the intended emotion to be rated as significantly higher than the other 3 emotions. This approach allowed for the identification of mixed emotions.

Table 1 displays results of Mauchly's test of sphericity for effect of emotion, showing that for all melodies sphericity was violated. Therefore, the Greenhouse-Geisser corrected degrees of freedom are reported. Table 2 displays results for the main effect of emotion on ratings for the 8 melodies. A statistically significant (p < .001) main effect of emotion was found for all melodies, with large effect sizes (above .65) for all except the short-scary melody, whose effect size was small (.16).

Table 1. Mauchly's test of Sphericity for each melody from repeated measures factorial ANOVA on emotion ratings

Melody	χ^2	df	p
Short-sad	17.401	5	<.05
Long-sad	12.871	5	<.05
Short-happy	45.023	5	<.001
Long-happy	41.892	5	<.001
Short-scary	26.093	5	<.001
Long-scary	26.447	5	<.001
Short-tender	14.407	5	<.05
Long-tender	39.549	5	<.001

To further understand these findings, estimated marginal means were plotted and Bonferroni pairwise comparisons were conducted. Estimated marginal means showed that each melody was generally rated highest in the intended emotion category, apart from the short-tender melody, which received higher ratings for *sad* than for *tender*. Pairwise comparison results showed that the highest rated emotion was also the expected emotion, and was rated significantly higher, at p <.05, than all other emotions, for all melodies except the tender-short and short-scary melodies. For the short-tender melody, *tender* and *sad* ratings were not significantly different from each other, but were significantly higher than *happy* and *scary* ratings, at p <,001. For the short-scary melody, *scary* and *sad* emotions were not significantly different from each other, but were significantly higher than *happy* and *tender*, at p<.05. These results are displayed in Figures 1-4.

Table 2. Main effect of emotion category for each melody from repeated measures factorial ANOVA on emotion ratings

Melody	F	df	p	η^2
Short-sad	186.30	2.23, 84.80	<.001	.75
Long-sad	126.533	2.44, 92.74	<.001	.67
Short-happy	203.279	2.13, 80.94	<.001	.78
Long-happy	238.367	2.19, 83.30	<.001	.81
Short-scary	12.079	2.03, 72.28	<.001	.16
Long-scary	122.571	1.99, 75.71	<.001	.70
Short-tender	117.097	2.48, 94.22	<.001	.68
Long-tender	94.901	1.82, 69.22	<.001	.65

There was a significant main effect of performer on emotion rating for the short-happy melody, $F(2, 76) = 4.10$, $p = .02$, $\eta^2 = .02$, the long-happy melody, $F(2, 76) = 6.17$, $p = .003$ $\eta^2 = .03$, the short-scary melody, $F(2, 76) = 11.47$, $p < .001$, $\eta^2 = .04$, the long-scary melody, $F(2, 76) = 7.14$, $p = .001$ $\eta^2 = .04$, the long-sad melody, $F(2, 76) = 8.67$, $p <. 001$, $\eta^2 = .04$ and the long-tender melody, $F(2, 76) = 7.01$, $p = .002$ $\eta^2 = .04$. This means that the overall ratings of perceived emotion in the melodies were mediated by performer differences.

Estimated marginal means showed that Performer 1 most often received the highest overall emotion ratings.

There was a significant effect of performance condition on emotion rating, $F(1, 38) = 4.66$, $p = .037$ $\eta^2 = .01$ for the long-sad melody only, with slightly higher ratings in the *visually expressive* condition ($M = 3.19$, $SE = .10$) than in the *immobile* condition ($M = 3.06$, $SE = .10$) meaning that for the long-sad melody, the *visually expressive* performance condition resulted in higher ratings of overall emotion conveyed.

There was a significant interaction effect of performer and emotion for the short-happy melody, $F(6, 228) = 6.12$, $p < .001$ $\eta^2 = .05$, the long-happy melody, $F(6, 228) = 6.66$, $p < .001$ $\eta^2 = .07$ the short-scary melody, $F(6, 228) = 5.16$, $p < .001$ $\eta^2 = .05$ the long-scary melody, $F(6, 228) = 3.57$, $p = .002$ $\eta^2 = .03$ the long-sad melody, $F(6, 228) = 3.39$, $p = .003$ $\eta^2 = .03$ and the long-tender melody, $F(6, 228) = 4.72$, $p < .001$ $\eta^2 = .06$, meaning that the effect of emotion category on the ratings was mediated by individual performer differences. There was also a significant interaction of performance condition and emotion, $F(3, 114) = 3.41$, $p = .020$ $\eta^2 = .02$ for the short-happy melody only, meaning that the effect of emotion category on ratings was mediated by the performance condition. As the specific details of these interactions were not important to the research question, they were not explored further.

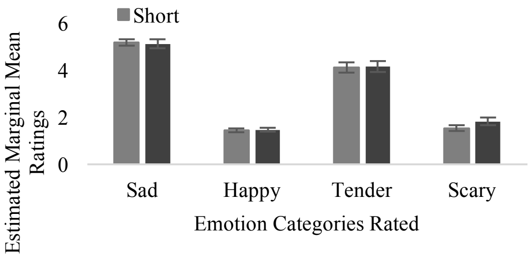


Figure 1. Estimated Marginal Means plot for effect of emotion category on emotion ratings for sad melodies. Error bars denote one standard deviation around the mean.

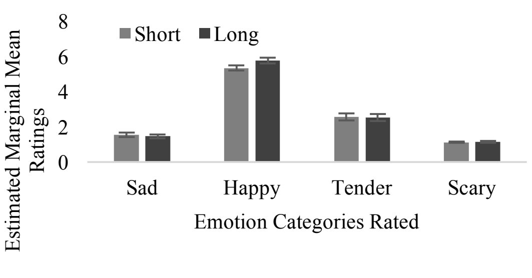


Figure 2. Estimated Marginal Means plot for effect of emotion category on emotion ratings for happy melodies. Error bars denote one standard deviation around the mean.

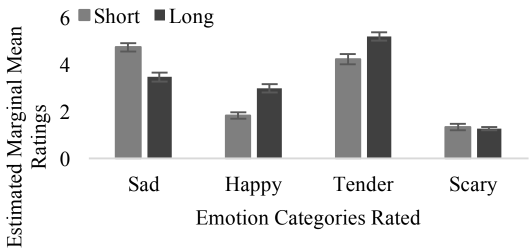


Figure 3. Estimated Marginal Means plot for effect of emotion category on emotion ratings for tender melodies. Error bars denote one standard deviation around the mean.

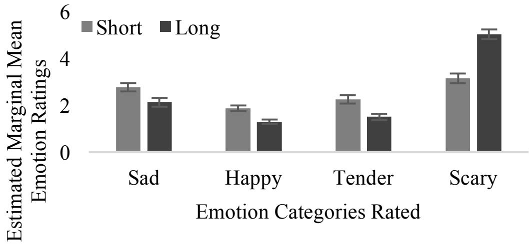


Figure 4. Estimated Marginal Means plot for effect of emotion category on emotion ratings for scary melodies. Error bars denote one standard deviation around the mean.

Expressivity ratings

This analysis tested Hypothesis 2. Results showed significant effects of performance condition for the short-sad melody, $F(1, 39) = 5.10$, $p = .03$, $\eta^2 = .03$, the short-happy melody, $F(1, 39) = 6.78$, $p = .013$, $\eta^2 = .02$ and the long-sad melody, $F(1, 39) = 7.70$, $p = .008$, $\eta^2 = .04$. For the short-sad melody, the *visually expressive* condition ($M = 5.00$ $SE = .16$) was rated higher than the *immobile* condition ($M = 4.68$, $SE = .13$). For the long-sad melody, the *visually expressive* condition ($M = 5.36$, $SE = .14$) was also rated higher than the *immobile* condition ($M = 4.97$, $SE = .18$). For the long-happy melody, the *visually expressive* condition ($M = 4.80$, $SE = .22$) was rated lower than the *immobile* condition ($M = 5.13$, $SE = .18$). For all other melodies, there was no significant effect of performance condition.

For the effect of performer, Mauchly's test of sphericity was violated for the long-happy melody, $\chi^2 (2) = 6.59$, p = .037, and for the long-scary melody, $\chi^2 (2) = 6.39$, p = .041, so, for those melodies, the Greenhouse-Geisser corrected degrees of freedom are reported. There were significant effects of performer on expressivity ratings for all melodies, which are displayed is Table 3. Bonferroni pairwise comparisons showed that significant differences between performers were different for different melodies, however Performer 1 received the highest ratings of expressivity for 6 of the 8 melodies.

There was a significant interaction effect between performer and performance condition, $F(2, 78) = 6.95$, $p = .002$, $\eta^2 = .06$ for Melody G only, showing that the effect of performance condition on expressivity ratings was different for each performer. As the specific details of these interactions were not important to the research question, they were not explored further.

Table 3. Main effects of performer on expressivity ratings for each melody

Melody	F-test	df	p	η^2
Short-sad	4.812	2	0.011	0.05
Long-sad	7.495	2	0.001	0.06
Short-happy	11.819	2	<.001	0.08
Long-happy	24.669	1.725	<.001	0.14
Short-tender	3.586	2	0.032	0.04
Long-tender	16.62	2	<.001	0.15
Short-scary	7.255	2	0.001	0.06
Long-scary	22.633	1.732	<.001	0.13

Note: **Greenhouse-Geisser corrected degrees of freedom are reported for long-happy and long-scary melodies**

Discussion

Emotion Ratings

The purpose of this analysis was to test the validity of the emotion labels given to the melodies used in the experiment, by testing Hypothesis 1. Results showed that Hypothesis 1 was supported for 6 of the 8 melodies, and therefore, the emotion labels given to these melodies can be considered valid. This means that these melodies would be suitable for use in future studies wishing to convey the emotions *happy, sad, tender* or *scary*. The short-sad melody was shown to convey mainly equal amounts of sadness and tenderness, while the short-scary melody, was shown to convey mainly equal amounts of scariness and sadness. The main effect of emotion category on the rating given was large for all melodies except the short-scary melody, which indicates that for this melody the difference between ratings of the four emotions was small, while for the others it was large. The short-scary melody can therefore be considered the most emotionally ambiguous melody, and the short-tender melody can be considered the second most emotionally ambiguous melody.

Results also indicate the presence of perceived mixed emotions in all melodies, in that there was a rating of higher than 1 for more than one emotion in every melody. This is consistent with previous evidence that music elicits mixed emotional reactions (e.g Hunter, Schellenberg, & Schimmack, 2010). In addition, the finding that both of the *happy* melodies and both of the *sad* melodies were unambiguously recognised, is in line with previous research that the basic emotions of *happy*, *sad, angry* and *fearful* are the most clearly conveyed in music (Gabrielsson & Juslin, 1996; Hailstone et al., 2009; Laukka & Gabrielsson, 2000) and that happiness and sadness may be the easiest emotions to communicate through music (Juslin & Laukka, 2004). For the two emotionally ambiguous melodies, alternative labelling of *angry* as *scary,* and *tender* as *loving* might have yielded better results. In addition, it should be noted that all the long melodies were unambiguously recognised for emotional content, implying that emotion recognition was easier in the longer melodies.

The main effect of performer showed how the different musicians influenced how high or low the emotion ratings were in general, across all emotion categories. This can be thought of as a measure of how expressive the performer was without reference to a specific emotion category, and can provide a comparison for the expressivity ratings. Consistent with previous findings that the music performer, as well as the structural features of the music, influences the emotion conveyed (Juslin, 2000), these results suggest that the performers influenced the strength of the emotional messages in the music. Overall, the results indicate that Performer 1 conveyed the strongest emotions in their playing.

Significant interaction effects of performer and emotion were shown for all melodies except the short-tender and short-sad melodies. This means that the effect of emotion category on ratings was mediated by performer differences, essentially showing that the performers had an influence on the perceived emotional content of the melodies. Again, this is consistent with the findings of Juslin (2000) that performers can alter the emotional content of a melody through acoustic cues in their performance.

Finally, there was a significant interaction between performance condition and emotion category for the long-sad melody only. This interaction, although it was only of small effect size, shows that the performance condition affected the perceived emotional content of the melody. This exploratory finding is highly relevant to the research question as it implies that the performers' approach to body movement changed some expressive aspect of the way this melody was perceived. Although this finding does not directly support Hypothesis 2, it does inform the research question, in that use of expressive body gesture did seem to have an influence on the emotional quality of the performances.

Expressivity Ratings

The aim of this analysis was to test Hypothesis 2. The two *sad* melodies both showed significantly lower ratings of expressivity in the *immobile* condition compared to the *visually expressive* condition with very small effect size. For the long-happy melody, ratings were significantly higher in the i*mmobile* condition compared to the *visually expressive* condition, again with a very small effect size. This effect in the long-happy melody could be explained by the technical requirements of the melody, as this was probably the most technically demanding performance, necessitating fast, staccato playing and at least one left-hand position shift. Expressive movements during staccato playing can create technical difficulties due to the changing position of the violin, and this could be the reason that this melody was rated as less expressive in the *visually expressive* condition. Therefore, these results suggest that when a violin melody includes fast, staccato playing, suppressing expressive gesture can help to improve expressivity, through reducing the technical demands of the playing. This finding is a good example of how technical and expressive aspects of musical performance can intertwine (Auer, 1960; Sloboda, 1996), and the importance of considering issues of instrument technique when studying expressive playing.

As both sad melodies exhibited significantly lower ratings in the *immobile* condition compared to the *visually expressive* condition, it can be concluded that, for the 3 performers who took part in this study, the audible expressivity of their performances of *sad* music was heightened when the performers made use of expressive body movement, compared to when they inhibited body movement. These results suggest that violinists playing sad, slow, legato music can increase the audible expressivity of their playing by focusing on a visually expressive performance. However, due to the very small sample size of performers, replication of these results with a different sample is required in order to generalise this finding. In relation to Sloboda's (1996) conception of how performers use knowledge of body motion to create expressive playing, and Juslin's (2003) GERM model of musical expression, this finding is very relevant as it suggests that bodily gesture while performing can be important to the performer's cognitive process of creating expressive sounding sad music. However, it should be noted that the effect size was very small, and the specificity of this effect to sad music only was unexpected.

In relation to previous research on the effects of the *immobile* condition, the results for the *happy, tender* and *scary* melodies support the findings of Thompson and Luck (2012) that performers were able to play at a normal level of expressivity under the *immobile* condition, while the results for the sad melodies support the findings of Wanderley et al. (2005) that the *immobile* condition impaired performers' expressive playing ability. For the long-sad melody, there was also a significant interaction between effect of performance condition and performer, highlighting that individual performer differences are important in this issue. Thus, these findings do not clarify the effects of the *immobile* condition, but further illuminate the complexity of this topic. Further research is needed to explore these issues in different performers, different instrumentalists, and different types of music.

Additionally, there were significant main effects of performer on expressivity ratings for all melodies, with Performer 1 as the most expressive performer overall. This result can be compared with the emotion ratings results, which also showed that Performer 1's performances received the highest perceived emotion ratings. Thus, both ratings indicate that performer 1 can be considered the most expressive performer. Similarly, the effects of performance condition on both expressivity and emotion ratings can be compared. The long-sad melody was the only melody to show a significant effect of performance condition on emotion ratings, but for expressivity ratings, there was also a significant effect of performance condition for the short-sad and long-happy melodies. As the long-sad melody also exhibited the largest effect size for effect of performance condition on expressivity ratings, the rating of expressivity may have been a more sensitive measure than the ratings of emotional content for detecting changes in performers' expressive playing. Overall, the emotion ratings showed a similar pattern to the expressivity ratings, suggesting that the two ratings were measuring similar constructs, as would be expected.

Conclusion

The results of this study contribute empirical findings to the theoretical discussion of how whole-body movement is involved in creating expressive sounding music. In this experiment, it was shown that when performers focused on increasing expressive gesture, compared to inhibiting it, the perceived audible expressivity of their performance increased, but only when they were playing *sad* melodies. In addition, for one *happy* melody, which was particularly technically difficult, the opposite effect occurred, in that focusing on increasing expressive gesture resulted in lower ratings of expressivity. Thus, the results demonstrate that different approaches to body movement can affect ratings of audible expressivity, and that the effect is mediated by the emotional content and technical demands of the music, as well as individual performer differences. These results offer some support for Juslin's (2003) psychological model of expressivity in music, by supporting the idea that knowledge of biological motion is an important aspect of creating expressivity in music. Precisely why the effect was only present in sad music is not known, and warrants further study.

This study also yielded some useful findings on the methodological approach of this experiment. Firstly, the *visually expressive* condition was a novel approach, implemented here for the first time. As performers did not report any confusion or difficulty with the *visually expressive* condition, this study showed that the instructions were effective and non-disruptive to the performance process. Secondly, the validity of using ratings of expressivity as a measure of the strength of emotion conveyed in a performance was supported, as ratings of emotional content showed similar patterns to ratings of expressivity. It was further suggested that the expressivity rating was a more sensitive measure to changes in a musical performance than the emotional content rating.

The findings presented here have provided a valid contribution to the topic of embodied music cognition, and how knowledge of motion contributes to expressivity in music performance. Further research could also explore how different movement approaches result in acoustic changes to performance sound, using music information retrieval techniques. These findings can also inform music performance practice and pedagogy, in promoting thoughtful attention to expressive musical skills, and whole-body movement during musical practice and performance.

Acknowledgments. *This study was carried out as part of a masters' thesis at the University of Jyväskylä.*

References

Broughton, M., & Stevens, C. (2009). Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychology of Music*, *37*(2), 137–153. https://doi.org/10.1177/0305735608094511

Dahl, S., & Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Perception: An Interdisciplinary Journal*, *24*(5), 433-454. DOI: 10.1525/mp.2007.24.5.433

Davidson, J. W. (1993). Visual Perception of Performance Manner in the Movements of Solo Musicians. *Psychology of Music*, *21*(2), 103–113. https://doi.org/10.1177/030573569302100201

Davidson, J. (2002) Communicating with the body in performance. In Rink, J (ed.) *Musical Performance, a guide to understanding* (pp.144 – 152). Cambridge: Cambridge University Press.

Demos, A. P., Chaffin, R., & Logan, T. (2017). Musicians body sway embodies musical structure and expression: A recurrence-based approach. *Musicae Scientiae*, 1029864916685928. https://doi.org/10.1177/1029864916685928

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of music*, *24*(1), 68-91. https://doi.org/http://dx.doi.org/10.1108/17506200710779521

Hailstone, J. C., Omar, R., Henley, S. M. D., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology*, *62*(11), 2141–55. https://doi.org/10.1080/17470210902765957

Hunter, P. G., Schellenberg, E. G., & Schimmack, U. (2010). Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts, 4*(1), 47-56. http://dx.doi.org/10.1037/a0016873

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: relating performance to perception. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(6), 1797–1813. https://doi.org/10.1037/0096-1523.26.6.1797

Juslin, P. N., Friberg, A., & Bresin, R. (2001). Toward a computational model of expression in music performance: The GERM model. *Musicae Scientiae*, *5*(1), 63–122. https://doi.org/10.1177/10298649020050S104

Juslin, P. N. (2003). Five facets of musical expression: A psychologist's perspective on music performance. *Psychology of Music*, *31*(3), 273–302. https://doi.org/10.1177/03057356030313003

Juslin, P. N., & Laukka, P. (2004). Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. *Journal of New Music Research*, *33*(3), 217–238. https://doi.org/10.1080/0929821042000317813

Laukka, P., & Gabrielsson, A. (2000). Emotional expression in drumming performance. *Psychology of Music*, *28*(2), 181-189. https://doi.org/http://dx.doi.org/10.1108/17506200710779521

Lindström, E., Juslin, P. N., Bresin, R., & Williamon, A. (2003). "Expressivity comes from within your soul": A questionnaire study of music students' perspectives on expressivity. *Research Studies in Music Education*, *20*(1), 23–47. https://doi.org/10.1177/1321103X030200010201

MacRitchie, J., Buck, B., & Bailey, N. J. (2013). Inferring musical structure through bodily gestures. *Musicae Scientiae*, *17*(1), 86-108. https://doi.org/10.1177/1029864912467632

Sloboda, J. (1996) The acquisition of musical performance expertise: Deconstructing the "talent" account of individual differences in musical expressivity. In Ericson, K. A. (Ed). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games.* (pp.107-126). New York, NY: Psychology Press.

Thompson, M. R., & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*, *16*(1), 19–40. https://doi.org/10.1177/1029864911423457

Thompson, M.R., Vuoskoski, J.K., & Clarke E.F. (2016). The relative contributions of composition and visual auditory performance cues to emotion perception: comparing piano and violin performances In T. Zantos (Ed.), *Proceedings of the 14th International Conference on Music Perception and Cognition,* San Francisco, CA.

Tsay, C.-J. (2013). Sight over sound in the judgment of music performance. Proceedings of the National Academy of Sciences of the United States of America, 110(36), 14580–5. https://doi.org/10.1073/pnas.1221454110

Van Zijl, A. G. W., & Luck, G. (2013). Moved through music: The effect of experienced emotions on performers' movement characteristics. *Psychology of Music*, *41*(2), 175–197. https://doi.org/10.1177/0305735612458334

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720–752. https://doi.org/10.1080/02699930701503567

Vines, B., Wanderley, M., Krumhansl, C., Nuzzo, R., & Levitin, D. (2004). Performance gestures of musicians: What structural and emotional information do they convey? *International Gesture Workshop,* 2004. Heidelberg: Springer. 3887–3887. https://doi.org/10.1007/978-3-540-24598-8_43

Wanderley, M.M. (2002) 'Quantitative Analysis of Non-Obvious Performer Gestures', in I. Wachsmuth and T. Sowa (eds) Gesture and Sign Language in Human–Computer Interaction, pp. 241–53. Berlin: Springer. https://doi.org/10.1007/3-540-47873-6_26

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The Musical Significance of Clarinetists' Ancillary Gestures: An Exploration of the Field. *Journal of New Music Research*, *34*(1), 97–113. https://doi.org/10.1080/09298210500124208

The Sound of Leadership: Effective Public Speaking Draws on Ethological Signals Associated With Authority

Bradley Alan Almond,[1] Joshua Albrecht[2]

[1] *Department of Management and Marketing, Texas A & M University – Central Texas, United States*
[2] *Music Department, The University of Mary Hardin-Baylor, United States*

[1]brad.almond@tamuct.edu, [2]jalbrecht@umhb.edu

Abstract

In this paper we present the results of a study of leader speech based upon the respondent ratings of good leadership within a pairwise comparison tournament of short audio excerpts of TED talks. While we find significant agreement among respondents with respect to which clips were better than others, we do not find unequivocal support for our hypotheses that mid-level auditory cues consistent with ethological assumptions provide significant influence over and above cultural factors such as regional accent, gender, and ethnicity. Three possibilities for these results are briefly considered and explored, including the influence of semantic content, low-level auditory information, or the artifactual influence of data aggregation across participants.

Introduction

Acoustic cues embedded within speech carry significant amounts of socially relevant information (Palmeri *et al.*, 1993). It is possible, upon hearing only a speaker's voice in a short utterance, to infer several sociolinguistic characteristics of the speaker, including gender (Perry *et al.,* 2001), regional identity (Clopper *et al.,* 2006), age (Waller *et al.*, 2015), height and weight (Krauss *et al.*, 2002), and even personality type (Scherer, 1979).

Humans are incredibly social, and so when they adapt to their environment, they are primarily adapting to successfully navigating the social interactions that guide most of their day-to-day reality (Mathew & Perreault, 2015). If there are adaptive and survival benefits in perceiving social characteristics from the sound of the speaker, then an ethological assumption would be that those attributes important for social interaction would be encoded into the speech signal.

Social status, as reflected in the degree to which an individual is considered to be a leader, appears to be one of those attributes that are acoustically-coded in speech. In a corpus study of speeches given by corporate CEOs, leaders of non-profit organizations, and university professors, Weninger *et al.* (2012) found that various aspects of leadership could be modeled using acoustic features alone at significantly above chance levels, reaching 72.5% accuracy. Likewise, Gregory and Gallagher (2002) found that the winners of presidential elections can be predicted by measuring spectral information in their speeches below 500 Hz. Rosenberg and Hirschberg (2005) found that fundamental frequency was able to predict charisma ratings, with intensity and speaking rate marginally correlated with charisma.

One potential explanation for some of these effects may be the way that physiological changes are related to other characteristics correlated with leadership. For example, Waller *et al.* (2015) found that listeners associated speech rate with age, with slower speech being associated with older age. Hummert *et al.* (1999) found that vocal volume (intensity) increased with age for male speakers. Fundamental frequency of speaking also gradually decreases as speakers age as a result of gradually lengthening vocal cords (Xue & Hao, 2003). Because social status is correlated with age, it is possible that when listeners judge leadership, they are evaluating mid-level acoustical correlates, such as speech rate, intensity, and fundamental frequency.

Research in signal processing and music information retrieval differentiates among four levels of audio information (Bello, 2018). The highest level of information relates to simple generic classifications (as opposed to measurements) such as style, form, or artist, each of which may be shared by many different audio selections. Mid-level information relates to quantifiable features such as beats, pitch, and onsets, each of which may be measured or differentiated by the average listener without the aid of a computer. Low-level information relates to complex computer-generated quantifiable features such as spectral flux and cepstrum that are unique to a particular signal or source and of which auditors are not consciously aware. The examples in the preceding paragraph demonstrate that more idiosyncratic lower-level acoustical features are often translated and classified by hearers according to a higher-level of socially constructed or culturally shared information.

A similar translation of lower-level acoustic parameters to a higher level of perception was documented in Li, Logan, and Pastore (1991). They discovered that the time distance between the heel and toe tap as a ratio of total leg length was the dominant factor in being able to discern gender by the sound of footfalls alone. By digitally varying the temporal distance between heel and toe tap in audio recordings of people walking, they were able to manipulate auditors' perceptions of the gender of the person walking. However, although listeners are attending to foot length, they are unable to provide estimates of foot length, suggesting that listeners are translating lower-level acoustic parameters into a higher-level representation of gender without having conscious access to the processes by which they are doing so. It is possible that similar translations are happening between the mid-level percepts of pitch height, sound intensity, and speech rate and the high-level conception of leadership.

If relatively lower-level parameters are the operative component of the audio signal, then listeners who are able to attend to these features more significantly may perceive leadership differently. It is well known that musicians tend to be able to discern finer distinctions in these lower-level acoustic domains. Moreover, there are clear parallels to the perception of authority in music. For example, Huron, Kinney, and Precoda (2006) found that melodies transposed down in pitch are perceived as more heavy or threatening, whereas melodies transposed up in pitch are perceived as more polite or submissive. Shanahan and Huron (2014) likewise found that opera composers encode sociability into their parts, with higher-pitched voices like tenors and sopranos assigned to more pro-social roles (like heroes) and lower-pitched voices like baritones, basses, and altos associated with roles given to less sociability, like villains. These results suggest that, at least for the domain of pitch, acoustic correlates of age and social status apply to perceptions of speech and music.

Hypothesis

The cross-domain consistency between speech and music support an ethological interpretation of acoustic correlates of leadership, which have implications for both speech and music. In this paper, we test the assumption that ethological signals associated with authority and positively correlated with age predict perceptions of leadership. Formally, we test the following hypotheses:

H1: Speakers with lower fundamental frequency are perceived as stronger leaders.

H2: Speakers with slower speech rates are perceived as stronger leaders.

H3: Speakers with greater intensity are perceived as stronger leaders.

Method

With a complex social character trait like leadership, there are likely a large number of factors that contribute to perceptions of the trait. Given the large number of leadership positions held by men, gender alone may account for a significant percentage of perception of leadership. As mentioned above, age may also play a role. Regional affiliations may also be significant, such that speakers reflecting a listener's culture or language background may be perceived as having stronger leadership.

For this study, we wanted to examine mid-level acoustical correlates with perceptions of leadership over and above some of these other social determinants. One way to accomplish this would be to eliminate variability in these domains. However, this approach would limit the generalizability of the findings to only, for example, middle-aged male English speakers. If acoustic elements of leadership are indeed ethological signals, then they should be perceptible in a broad sample of cultures, genders, and ages.

We therefore opted to examine a wide range of speakers, and sought instead to control for these variables before testing our hypotheses. One strategy would be to use carefully controlled and balanced stimuli created in an experimental setting. However, this approach would suffer from a lack of ecological validity. We therefore opted to use speeches from a naturalistic setting.

Sample

In order to try to minimize the effect of confounding variables, like speech setting, historical speech patterns, and subject matter, we chose to randomly sample talks from www.TED.com, a repository of relatively recent talks centered around technology, entertainment, and design. TED speakers represent a wide range of backgrounds, and each speaker is an established or emerging leader in his or her respective field, selected by means of a thorough pre-screening and review process. TED speakers include heads of state, heads of corporations, influential scientists, artists, and educators, and winners of nearly every major award of distinction including Nobel, Pulitzer, Grammy, Oscar, Tony, Medal of Honor, OBE, and so on. In addition, TED speakers typically employ an epistemic modality of confidence and enthusiasm delivered in a rhetorical mode of persuasion and narrative rather than one of analysis and exposition (Caliendo & Compagnone, 2014), and are therefore more likely to be perceived as leaders as opposed to merely experts.

For our study, we randomly sampled 100 English-language TED talks from the website, each from a different speaker. After talks were selected, we downloaded .mp3 files of the audio only, and then randomly sampled an 8-second excerpt from within each talk. As we wanted each excerpt to contain only the speaker's voice, any excerpt featuring other noises such as laughter, applause, music, or the voice of a second speaker was thrown out and replaced with another excerpt from the same speaker. To limit the risk of participant bias in favor or against a particular speaker on the basis of semantic content rather than acoustical properties, we also eliminated and replaced any samples that contained trigger words or speech content that might be deemed offensive or polarizing.

Of the 100 excerpts included in the final sample, 60 were given by men and 40 by women, 64 were given by Americans and 36 by non-Americans, and 88 were given by white speakers whereas 12 were non-white. Each speaker was only represented once in the database.

Automatic feature extraction

The MIR toolbox version 1.7 was used to estimate many low-level acoustic features of the corpus using automatic feature extraction (Lartillot & Toiviainen 2007). The *mirfeatures* tool was employed, which returned 378 acoustic features for each audio excerpt. A full list of features in the MIR toolbox can be found in the User's Manual (Lartillot 2017).

For purposes of our hypotheses, only those variables related to the mid-level feature of intensity (dynamics and fluctuation) were utilized.

Additional features

Our original hypotheses involved the mid-level features of fundamental frequency and speech rate, but these features are not included in the MIR toolbox output. To calculate fundamental frequency, Sonic Visualiser (v. 3.0.3) was used to estimate the pitch of the fundamental (Cannam, *et al*., 2010). Specifically, the "Fundamental Frequency (failsafe)" plug-in from the "LibXtract" library by Jamie Bullock was used (Bullock, 2007). This plug-in identifies windows of small numbers of samples and assigns a fundamental frequency based on the spectral energy. Due to noise artifacts in the recordings, automatic detection of the fundamental can be problematic. Since most human speech falls in the range of 85-300 Hz, any windows in which the fundamental estimation fell outside of the 50-400 Hz range was discarded, and the remaining estimations were averaged together to attain an estimate of fundamental frequency for each sample.

In order to estimate speech rhythms, two measures were used, total number of syllable onsets and nPVI. Although nPVI is not related to speech rate, it has been used extensively to analyze speech rhythms in the last few years and has been shown to significantly predict elements of speech like language-family (Grabe & Low, 2002). It accomplishes this by looking at pairs of syllables to determine how even the distribution of syllable length is. To determine number of onsets and nPVI, the authors subdivided the set of 100 samples and analyzed syllable onsets for each recording. To check for reliability, each author also analyzed 5 (10%) of the other author's samples. Of the shared samples, the correlation of nPVI scores between authors was .917, which was significantly correlated at p = .0002. Given the strong correlation, no further validation was deemed necessary.

Content analysis

Although the focus of this study was on mid-level acoustic correlates to perceptions of leadership, and participants were instructed to ignore semantic meaning, it is possible that semantic content may have influenced participant responses. To examine the effect of semantic content on participant responses, the texts for each sample were transcribed and automatic content analyses were performed on each sample.

The program used to perform content analysis was the Linguistic Inquiry and Word Count program (LIWC2015), consisting of over 90 variables that measure properties of texts (Pennebaker, *et al*. 2015). Specifically, LIWC2015 measures the degree to which texts suggest logical argument, the relative social status of the text, levels of authenticity in the text, the overall positive or negative emotional tone of the text, the degree of anxiety, anger, sadness, and positive emotion of the text, social words, affiliation, achievement, power, reward focus, and risk prevention focus, among dozens of other dimensions.

Participants

To examine whether there were any differences between musicians and non-musicians in their approach to interpreting leadership using mid-level acoustic domains, two samples of participants were used for this study. Twenty-seven undergraduate music majors in their first year of study at the University of Mary Hardin-Baylor (UMHB) and forty-one upper-level and nontraditional undergraduate and graduate students at Texas A&M University – Central Texas (TAMUCT) were recruited to participate in the study.

Of the 27 UMHB students, 11 were male and 16 were female. Mean age was 19.1 (sd = .95). As a continuous metric of musical sophistication, participants also completed the Goldsmith's Musical Sophistication index (Müllensiefen *et al.,* 2014). The mean Gold-MSI for general musical sophistication was 96.6 out of 126 total (sd = 11.6). Average scores of subsets of the Gold-MSI measure included 48.8 out of 63 for active engagement (sd = 6.8), 48.2 out of 63 for perceptual abilities (sd = 7.9), 37.6 out of 49 for musical training (sd = 4.5), 34.6 out of 42 for emotions (sd = 4.4), and 34.3 of 49 for singing abilities (sd = 7.4).

Of the 41 TAMUCT students, 22 were female and 19 were male. Mean age was 34.1 (sd = 8.3). The mean Gold-MSI for general musical sophistication was 64.6 out of 126 total (sd = 19.8). Average scores of subsets of the Gold-MSI measure included 32.7 out of 63 for active engagement (sd = 10.2), 43.9 out of 63 for perceptual abilities (sd = 8.3), 18.4 out of 49 for musical training (sd = 10.2), 29.2 out of 42 for emotions (sd = 6.4), and 25.6 of 49 for singing abilities (sd = 8.5).

As a means of ensuring differences between groups, t-tests were performed for age and each level of the Gold-MSI. UMHB students were significantly younger (p < .0001), and significantly higher on all musical sophistication indices at p < .05. However, the difference in perceptual abilities was significant at p = .04, which fails to reach significance under a Bonferroni correction.

Procedure

At UMHB, participants completed the study in a computer lab consisting of eight workstations with up to eight participants at a time. At TAMUCT, participants completed the study one at a time at a campus library workstation.

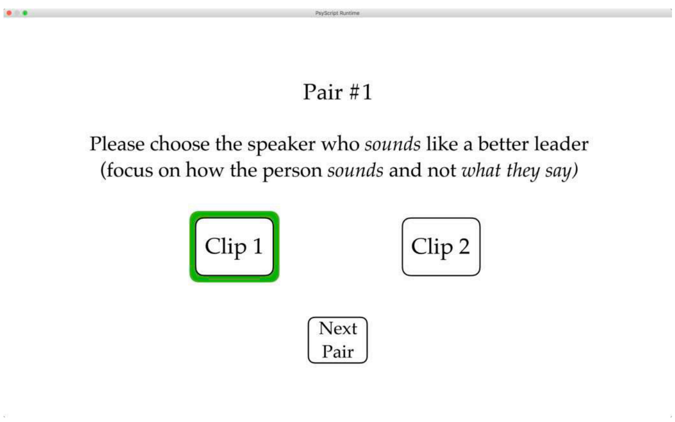


Figure 1. The sample interface, showing a forced-choice paradigm in which participants chose the speaker that sounded like a better leader. Clips were highlighted to indicate when it was playing

In this study, participants engaged in a forced-choice paradigm in which they chose the speaker that sounded like a better leader in a head-to-head comparison. A sample interface is shown in Figure 1. Two clips were randomly chosen from the list of 100 possible clips for a head-to-head comparison. Participants heard the first 8-second clip, followed by a 1.5 second pause, and then the second clip. Clips were illuminated during playback so that participants would know which clip they heard.

After listening to both excerpts, participants were instructed to select the excerpt that sounded like a better leader. For this study, participants were explicitly told to ignore whether they agreed with the speaker, liked the topic better, or anything relating to the semantic content. In other words, participants were told to focus only on the sound of the speaker and not what the speaker was saying.

Before the study began, participants completed one practice trial with the experimenter and were allowed to ask any questions they had. They were then given 50 head-to-head comparisons, with the 100 clips randomly chosen for each comparison series. Once all 100 clips were sampled, the list was reshuffled and another 50 comparisons were given, for a total of 100 forced choices. Once the 100 trials were completed, participants filled out the Gold-MSI and then a post-experiment questionnaire.

Results

Leadership scores

The most complete methodology for determining the clips with the greatest degree of leadership would have a large number of participants compare all possible pairwise combinations of clips. However, with 100 clips there would be 4,950 pairings. For 100 pairings per participant, this would require 50 participants to get one comparison per pairing. In order to get 30 comparisons per pairing, we would need 1,500 participants, an untenable number.

Instead, our approach to analysis is similar to a bracket for a sports league. Although every team is not able to play every other team in the league, win-loss record is used to place each team within the bracket. Consequently, we consider each comparison that an excerpt participated in to be a 'competition' and then we measured how many of those competitions were won by each clip. These win counts became the dependent variable predicted by the various acoustic features used in our analysis.

Regression models

As a preliminary test of test our hypotheses we conducted a series of linear regressions using SPSS. Win count was the dependent variable, and three dichotomous variables were utilized as controls: gender, whether or not the speaker had an American accent, and whether or not the speaker was ethnically white. Hypothesis 1 tested whether fundamental frequency provided significant explanatory power for win count in the presence of these controls. Hypothesis 2 tested the explanatory power of syllable onsets in the same conditions, and hypothesis 3 tested that of the dynamics and fluctuation variables from the MIR toolbox. Hypotheses 1 and 2 returned insignificant results. While the F statistics for both models indicated significant explanatory power for the overall model ($F = 4.747$, $p = .002$, and $F = 5.173$, $p = .001$, respectively), the p values for the focal independent variables were insignificant in both cases ($t = .396$, $p = .693$, and $t = 1.257$, $p = .212$, respectively). Hypothesis 3 included a set of 8 independent variables from the MIR toolbox related to intensity. For this hypothesis test, the three controls were entered as a block, and the 8 independent variables were entered according to a forward stepwise procedure. The procedure identified only one significant focal independent variable, *fluctuation_peak_PeakMagMean* ($F = 7.484$, $p < .001$, $t = 3.049$, $p = .003$).

Additional analysis

In contrast to the findings of Rosenberg & Hirschberg (2005), we did not find a significant relationship between mid-level acoustical properties (e.g., fundamental pitch and onsets) and perceptions of leadership. There are a great many differences between the design of that study and this one that could account for this discrepancy, but one critical difference could be that in the present study we allowed respondents to collectively define what "good leadership" sounds like by their aggregated choices, while Rosenberg and Hirschberg themselves pre-selected audio clips by listening to them and deciding to include only those that sounded charismatic. Thus, their sample consisted solely of clips of good leadership. To approximate these conditions, we re-tested our hypotheses using only the 50 clips with the best win-loss record. With this new test, fundamental frequency was found to be significant in the presence of controls ($F = 5.578$, $p < .001$, $t = 2.943$, $p = .005$), suggesting that fundamental frequency may not be a robust determinant of auditor perceptions of leadership quality across a full range of leadership exemplars.

In any case, the effect of mid-level auditory information was not in accord with our hypotheses, and this is particularly puzzling given that in post-experiment interviews most respondents reported that their selections were, in fact, based on mid-level auditory cues such as tone, pitch, speed, cadence, stresses, pauses, and so on. There are at least three possible explanations for this result.

One possibility is that in spite of our efforts to limit the influence of the semantic content of the clips on auditor preferences and choices, the actual words of the clip (rather than the sounds) could have exerted a significant influence in this fashion. As an exploratory investigation of this, we ran a forward stepwise regression procedure in SPSS using all of the variables generated by the LIWC program mentioned earlier. The resulting model included five significant independent variables ($F = 7.488$, $p < .001$ for overall model; all $p < .05$ for independent variables). These independents included variables relating to cognitive processing and clout, which could have influenced our respondents by functioning as proxy indicators of intelligence and leadership, respectively.

Another intriguing possibility is that, similar to the findings of Li, Logan, & Pastore (1991), our respondents may be making their selections on the basis of low-level auditory information of which they are only unconsciously and tacitly aware (Polanyi, 1983). Again, as a preliminary and purely exploratory consideration of this possibility, we ran a forward stepwise regression in SPSS using the 357 variables calculated by the MIR toolbox as independent variables and win counts as the dependent variable. This procedure yielded a model with 22 significant predictor variables (all $p < .05$) and an adjusted R-square of .735. All but one of the predictors were spectral variables of some kind, and 17 of these were related to Mel-frequency cepstral coefficients (MFCC) or their first- or second-order derivatives (DMFCC or DDMFCC), which have been shown to be effective in speaker verification in artificial intelligence applications (Hossan, Memon, & Gregory, 2011).

A third possibility is that by aggregating responses across participants we obscured and confounded what may have been clear preferences and stable selection criteria within individual respondents. Further analysis will include by-participant examination of the effect of auditor demographics on choice patterns and implicit preferences. It is likely that the latent model of "good leader" systematically varies from respondent to respondent, but further analysis will be needed to determine the nature and extent of these differences.

Conclusions and Future Research

In this paper we have presented the results of a study of good leader speech based upon the respondent ratings within a pairwise comparison tournament of short audio excerpts of TED talks. While we found significant agreement among respondents with respect to which clips were better than others, we did not find unequivocal support for our hypotheses that mid-level auditory cues consistent with ethological assumptions provided significant influence over and above cultural factors such as regional accent, gender, and ethnicity. Three possibilities for these results were briefly considered, including the influence of semantic content, low-level auditory information, or the artifactual influence of data aggregation across participants. In addition to the by-respondent analysis suggested above, a truly context-free study could be accomplished via the extraction of MIDI files from the audio excerpts of leader speech. These clips could then be rendered using the same MIDI instrument to create context-free clips of leadership speech. Respondents who were asked to rate which clip sounded like a good leader would then have only variation in mid-level auditory information to work with, since all three other sources of variation (cultural, low-level auditory, and semantic) would be effectively controlled for.

Acknowledgements. Special thanks to Ikeoluwa Oyewole for her meticulous and dedicated administrative support of this study. Thanks to the UMHB faculty development grant for funding for travel to ICMPC.

References

Bello, J. (2018, June 9). *Low-level features and timbre*. Retrieved from http://www.nyu.edu/classes/bello/MIR_files/timbre.pdf.

Bullock, J. (2007). LibXtract: A lightweight library for audio feature extraction. In *Proceedings of the 2007 International Computer Music Conference*, Sweden.

Caliendo, G., & Compagnone, A. (2014). Expressing epistemic stance in University lectures and TED talks: a contrastive corpu-based analysis. *Lingue Linguaggi*, *11*, 105-122.

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic Visualiser: An open source application for viewing, analyzing, and annotating music audio files. In *Proceedings of the ACM Multimedia 2010 International Conference* (pp. 1467-1468). Firenze, Italy.

Clopper, C.G., Levi, S.L., Pisoni, D.B. (2006). Perceptual similarity of regional dialects of American English. *Journal of the Acoustical Society of America, 119*, 566-574.

Grabe, E., & Low, E.L. (2002). Durational variability in speech and the rhythm class hypothesis. *Laboratory Phonology, 7,* 515-546.

Greggory, S.W., & Gallagher, T.J. (2002). Spectral analysis of candidates' nonverbal vocal communication: Predicting U.S. presidential election outcomes. *Social Psychology Quarterly, 65*, 298-308.

Hossan, M., Memon, S., & Gregory, M. (2011). A novel approach for MFCC feature extraction. In (Ed.), *Proceedings of the 4th International Conference on Signal Processing and Communication Systems*.

Hummert, M.L., Mazloff, D., & Henry, C. (1999). Vocal characteristics of older adults and stereotyping. *Journal of Nonverbal Behavior*, *23* (2), 111-132.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review*, *1* (3), 170-177.

Krauss, R.M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology*, *38*, 618-625.

Lartillot, O. (2017). *MIRtoolbox User's Manual*. Retrieved from https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7.pdf

Lartillot, O., & Toiviainen, P. (2007). A MATLAB toolbox for musical feature extraction from audio. In (Ed.), *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-01-08)*. Bordeaux, France.

Li, X., Logan, R., & Pastore, R. (1991). Perception of acoustic source characteristics: Walking sounds. *The Journal of the Acoustical Society of America, 90,* 3036 .

Mathew, S., & Perreault, C. (2015). Behavioural variation in 172 small-scale societies indicates that social learning is the main mode of human adaptation. *Proceedings of the Royal Society B*, *282*: 20150061.

Müllensiefen, D., Gingras, B., Steward, L., & Musil, J. (2014). The musicality of non-musicians: An index for measuring musical sophistication in the general population. *PLoS ONE, 9* (2), e89642.

Palmeri, T.J., Goldinger, S.D., & Pisoni, D.B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*: 308-328.

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.

Perry, T.L., Ohde, R.N., & Ashmead, D.H. (2001). The acoustic basis for gender identification from children's voices. *Journal of the Acoustical Society of America, 109*, 2988-2998.

Polanyi, M. (1983). *The Tacit Dimension*. Gloucester, MA: Peter Smith.

Rosenberg, A. & Hirschberg, J. (2005). Acoustic/prosody and lexical correlates of charismatic speech. *Proceedings of the International Speech Communications Association* (pp. 513-516). Lisbon, Portugal.

Scherer, K.R. (1979). Personality markers in speech. In K.R. Scherer & H. Giles (Eds.), *Social Markers in Speech* (pp. 147-209). Cambridge: Cambridge University Press.

Shanahan, D., & Huron, D. (2014). Heroes and villains: The relationship between pitch tessitura and sociability of operatic characters. *Empirical Musicology Review, 9* (2), 141-153.

Waller, S.S., Eriksson, M., & Sörqvist, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in Psychology, 6*, 978.

Weninger, F., Krajewski, J., Batliner, A., & Schuller, B. (2012). The voice of leadership: Models and performances of automatic analysis in online speeches. *IEEE Transactions on Affective Computing*, *3* (4), 496-508.

Xue, S.A., & Hao, G.J. (2003). Changes in the human vocal tract due to aging and the acoustic correlates of speech production: a pilot study. *Journal of Speech, Language, and Hearing Research, 46*, 689-701.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Expert Teacher: Investigations on Pedagogical Practices Used in Music Performance

[1]Anderson César Alves
[1,] *Music School, Federal University of Rio de Janeiro, Brazil*
[1] andces@uol.com.br

Abstract

This paper aims to comprehend the development of *expert* performance among Brazilian's clarinet players, and it has inspired me to investigate the factors relating to the development of musical expertise in light of expert teachers and investigate how pedagogic and artistic expertise are linked. The implication of this research on teaching expertise has frequently been discussed and analyzed in the field of education, and identifying those high pedagogical practices in the field of music research could help to understand how musical expertise can be developed based on high pedagogical models. Berliner (1988) took an interest on the subject and adapted the heuristic model of development proposed by Dreyfus & Dreyfus (1980) to the development of pedagogical expertise. The aim is to identify the pedagogic practices used by expert teachers directly related to the development of expert performance among clarinet players. The main hypothesis is that the development of expert performance in music is related to the access to the expert teachers. The methodology will be the concurrent analysis protocol with three clarinet teachers, and it consists of video observation and recording of their musical classes, in order to understand the procedures involved in the teachers' pedagogical actions (Ericsson, 2006). The pedagogical practices used in their classes will be analyzed, and the data will be compared to the literature review in pedagogy and cognitive psychology in order to comprehend the development of expert performance in music. Results of my data collection are still ongoing, and I propose to establish a dialogue between the results of my research with education literature, highlighting the data could confirm the narrow link between the role of expert teachers in the development of expert music performance in their students.

Introduction

There are Brazilian clarinet players regarded as *experts* acting as teachers. That has led me to investigate the practices used by clarinet teachers who have turned certain students into high class soloists. Expertise can be defined as the ability acquired by the practical exercise of performing well, in qualitative terms, in a particular task of a specific domain (Frensch & Stenberg, 1989, p. 158). It is related to processes of knowledge accumulation and mechanisms that monitor cognitive functions, as a way to perform a set of tasks in an efficient way. It is a long-term process, fruit of experiences and extensive practices (Feltovich; Prietula; & Ericsson, 2006). Experts have the skills to restructure, reorganize and refine their representations of knowledge and they know the required mechanisms to make use of those skills in their work environments (Ericsson; Lehman, 1996).

In the early 1980s, research on expertise in several domains began to progressively influence studies on teaching. In the ensuing literature, research on people who had shown qualitatively different performances seemed relevant to the investigations related to the field of education, as the results indicated that expert learning was directly related to expert teaching (Siedentop, & Eldar, 1989). Said emerging literature on the similarities of expertise in different fields seemed relevant to researches on education, because, by analogy, it was as if those researches were describing the effective qualities of those teachers. That way, research on expertise was able to provide information that allowed an investigation into the learning processes of expert teachers (Berliner, 1988, 1994, 2004).

The objective of this article is to identify the pedagogical practices used by expert clarinet teachers that are directly related to the development of musical expertise on clarinet players. The method used for data generation is simultaneous protocol analysis (Ericsson, 2006), which consists of a structured observation and video recordings of classes with the intent of understanding the procedures involved in the pedagogical actions of the teachers. Three teachers recognized by their peers as expert clarinet teachers were chosen. The criteria were exclusively the high level of their classes in at least 10 years of teaching.

Expert Teacher

Research by Ericsson, Römer & Krampe (1993) and by Sosniak (2006) discussed that almost every single person who showed exceptional performances had studied with great masters, i.e., renowned teachers in their specific domains. Research on *expertise* in different domains began to be developed in the 1980s, especially in the field of pedagogy. In those researches, the characteristics of individuals who had shown a superior level of performance were of interest to researchers on the field of education, as the qualities of the students were related directly to those of efficient teachers.

It was of interest to Berliner (1988), who adapted the heuristic development model proposed by two Brothers who teach at the University of California, the philosopher Hubert Dreyfus and the computer scientist Stuart Dreyfus, for the development of pedagogical *expertise* (1980). Based on their model, Berliner (1988) wrote about five stages of teacher development: novice, advanced beginner, competent teacher, proficient teacher and *expert*. In the first stage, novice, the teachers display a rational and relatively inflexible behavior. They tend to follow the rules and procedures they were instructed to follow, and the expected performance is within a certain limit. The second stage is the advanced beginner, where episodic knowledge and case knowledge are built. In the third stage, the "competent teacher", he makes conscious choices on what he will do, establishing priorities and making decisions regarding planning. Their objectives are well-defined, and they know the means to reach their goals. In the fourth stage, the "proficient teacher" has developed the intuition and know-how through accumulated experience, recognizing similarities that allow him to accurately foresee events. Lastly, in the fifth stage, the "expert teacher", which Berliner describes as intuitive, the teacher has a non-analytical approach, which means he makes decisions in non-deliberate fashion.

A research program was created in the USA in 1987 to evaluate teachers. It is called *National Board of Professional Teacher Standards* (NBPTS). The program depends on a national system of volunteers and aims to evaluate and certify expert teachers. The program has 63 members, more than half of which are teachers. Candidates send portfolios and videos of their classes to an evaluation center to be certified. Reviewers were trained to evaluate the characteristics of their expertise, analyzing and coding numerically the data from the interviews. They looked for evidence of knowledge organization and reorganization, links between the knowledge of the teachers in other school disciplines, and the link between the knowledge of the teachers and previous and future knowledge of their students. Thirteen consistent characteristics were chosen to evaluate the performance of the expert teacher: better use of knowledge; extensive knowledge of pedagogical content; better strategy to resolve problems; better adaptation and modification of goals of several students; better decision making; more challenging objectives; better atmosphere in the classroom; better perception of events taking place in the classroom; a higher sensibility to the context; more attention to the learning aspect, including giving feedbacks to students; frequent testing of hypotheses; respect to students; and passion for teaching.

Berliner (2004) also defines that one of the ways to evaluate the performance of the expert teacher is directly related to their students' level of performance, so they were also evaluated, and the following criteria were taken into consideration: high motivation for learning, feeling of self-efficacy; deep understanding of the subject; and high levels of success (Berliner, 2001, 2004). The results showed that certified teachers had better results than the others in every criteria. As for the students, there was no significant difference between those of certified teachers and uncertified teachers, since all selected teachers were highly experienced and well-prepared

Another significant item is the fact that time and experience have an important role in the development of the teacher's expertise. The teacher acquires experience after five years of teaching, but it takes seven to ten years to become an expert teacher, which is approximately ten thousand hours acting in their field (Berliner, 1988, 2004). An expert teacher with 7 years of experience has given 7,000 hours of classes. If said teacher came from a regular university course, he will have 1,000 more hours of experience in teacher-student relation.

Although research on teacher expertise has been carried out in small scale, some similarities were found between the expert teacher and the expert in general. Among them are the fact that high performance teachers outperform the others, in their domain and in a particular context, the fact that they develop automation through repetition of operations that are necessary to reach their goals; they are more sensitive to the demands of the tasks and to the social situation when solving problems; they find significant patterns in their domains; they are quicker and more accurate in recognizing their students' skills; they use richer sources of personal information to deal with the problems they face; and they are more intelligent and flexible when it comes to teaching (Berliner, 1994, 2001, 2004). According to Berliner (2001), research on teacher expertise up to that point have the same difficulties as in other fields, such as the fact that research methods are always qualitative and intensively focused on a small group of individuals. Another difficulty is the impossibility to observe the performer for a decade of intense practice. All results are based on observations during short periods of time (Ericsson; Smith, 1994).

Expertise on Performance Pedagogy

Duke & Simmons (2006) showed that the study of expertise on teaching has been widely discussed and analyzed in the field of education, and expertise on that area is defined as an indication of good teaching. Identifying those pedagogical practices in the field of music research has been a problem for educators and professionals involved in teacher evaluation, as the interactions between teachers and students are different from the conventional teaching method. According to the authors, this could be the cause of the scarcity of research on the expertise of music teachers.

They recorded and analyzed 25 class hours of three music teachers that are internationally renowned as performance teaching experts, with the intent of describing the processes of teaching and honing musical expertise in their students' performance. Analyses of the data from the video recordings revealed similar strategies, which were broken into three categories: "goals and expectations", "effective changes" and "relayed information".

In the "goals and expectations" category, the level of challenges of the repertoire was determined by the proficiency of the student. The teachers have also developed the skill of categorizing the chosen repertoire in auditory imagery. The decisions they make are appropriate to the process of resolving problems of performance, comparing the development of the student. In the "effective changes" category, accuracy and excellence were the goals of every performance. The strategic planning for the performance was defined by the teacher, and if the goals were not met, new goals were created. Small breaks for mental and physical rest helped to correct technical problems performances had. There was some freedom in interpretative choices, but within certain limits. In the "relayed information" category, the auditory imagery of the repertoire was observed in the evaluation of the teachers. The close relationship between physical movements and their effects on sounds was analyzed. Self-feedback was used as a tool to improve performance, and imitation was also used by teachers as a pedagogical tool.

That research highlighted important pedagogical tools related to performance teaching, such as: level of challenges; categorization of the repertoire in mental images; appropriate decisions to resolve problems; comparisons between stages of the development of the performance; access to feedback; motivational processes; accuracy and excellence when performing the repertoire, with pieces played from beginning to end; strategic planning of the performance, where technical flaws were related to the excess of physical movements; the importance of taking breaks for physical and mental relaxation; permission for taking liberties with interpretative choices; and imitation.

Model for Evaluating Pedagogical Expertise on Musical Performance

After reviewing the literature, I was able to see that there are many studies on musical expertise, but very few when it comes to focusing on the expertise of the teachers´ performance. The lack of research on the expertise of teachers of any instrument, in special, has led me to seek out the literature of related fields and, through analogy or approximation, use that data on research on expert teachers of musical performance.

This article is based on my doctoral research, which is still in progress, and through the data I will obtain it will be possible to think up a model that could define the characteristics of expert teachers of musical performance. Table1 represents the general structure on which the hypothesis that I will discuss and validate is based. It is divided on 4 columns. In the first column, there is the data from the American teacher evaluation platform (NBTPS, 2012), which defines the characteristics which are taken into consideration when evaluating and certifying expert teachers. In the second column, I included the data from Dreyfus (1980) on the development of skills and, in the third column, the data from Berliner (1988) on the pedagogical expertise of the teacher. It is important to know that Berliner used the data from Dreyfus as well, taking the information to the field of pedagogy. When analyzing and comparing the data and definitions of the expert teacher, it was possible to confirm that the characteristics of said teachers related to one another organically, and I believe that by collating the qualities discussed by the theories I will be able to use the model I propose here.

Table1 – Descriptive Model of the Qualities of the Expert Teacher

Dreyfus (1980)	**Berliner (1988)**	**National Board for Professional Teaching Standards (2012)**	**Expert Teacher on Musical Performance**
Deep holistic understanding	Tacit knowledge in problem solving	Extensive knowledge of pedagogical content	Orientation and strategies for performance preparation
Intuitive way of solving common problems	In problematic situations, deliberative analytical processes are used	Better strategy for problem solving	Analytical thinking about performance
Acquired skill to go beyond existing interpretations	Decision making	Better use of knowledge	Acquired skill to manipulate important ideas on the content
Does not depend on rules and guidelines	Planning	Better decision making	Patterns of recognition, problem categorization and solving in different situations
Reaches excellence easily	Defined goals	Adaptation and modification of goals	Levels of performance challenges
Analytical approximations used only on new situations	Intuitive patterns of recognition	Monitoring of learning	Didactic-pedagogical strategies
Acquired skill to solve problems intuitively	Intuitive understanding of the situation	Testing of hypotheses	Thinking about performance (Praxis)
Capable of going beyond the existing patterns, creating their own interpretation	Decision making	Challenging goals	Teaching how to learn
Holistic understanding of complex situations, alternating between analytical or intuitive approaches	Problem solving	Perception of events in the classroom	Strategies for problem solving
Has a general overview and alternative approaches	Patterns of recognition	Sensibility to the context	Problem solving
Masterful knowledge of the discipline in his field	Access to significative sources of information	Giving feedback to the students	Holistic approach

Discussion and Conclusion

The lack of a model on the qualities that can define expert teachers of musical performance made me think of the paradigm I showed in the fourth column of Table1. The model I created to define the qualities of the expert teacher of musical performance can be used as an evaluation tool and promote the development of new studies.

Table 1 will also be used as an important tool for data generation. Collation and analysis of the data will be utilized to justify my model and found very consistently my hypothesis on the honing of the expert musical performance considering the access to expert teachers.

My goal is to build a model that can define the characteristics of expert teachers of musical instruments. I base the present study on models from studies on expert teachers of pedagogy by NBPTS (2012), on the development of skills by Dreyfus and Dreyfus (1980), and on pedagogical expertise by Berliner (1988, 2001, 2004). The lack of a specific model for expert teachers of musical performance has led to the elaboration of a model that can guide the investigation and analysis of data from observations of pedagogical practices in said field.

After analyzing the data from studies of other fields, it was possible to come up with a model for evaluating expert teachers based on their pedagogical practices for musical performance preparations, such as: orientation and strategies for performance preparation, an analytical thinking about their own performance, acquired skill to manipulate important ideas on the content to be taught, or on the programmatic content of cognitive psychology, like patterns of recognition, problem categorization and solving in different situations. There was also the concern about the levels of motivation — which is often related to the challenge of a piece of music — and the pedagogical strategies used by the teachers when playing a piece. Another important element is the praxis related to thinking about the performance as a whole. An expression from the field of education that can be used by music teachers is "teaching how to learn", because it can contribute to the building of metacognitive processes on students. Based on Ericsson *et al.* (1993), it was possible to identify that the expert musician has developed a superior skill of problem solving; through observation, I will investigate the strategies teachers use to solve problems quickly and concisely. I will also take into consideration the type of holistic approach those teachers adopt in their pedagogical practices.

References

Berliner, D. C. (1988). *The development of expertise in pedagogy*. American Association of Colleges for Teacher.

Berliner, D. C. (1994). Expertise: the wonder of exemplary performances. In Mangieri, J. N. & Block, C. C. (Ed.), *Creating a powerful thinking in teachers and students* (pp. 141-186). Fort Worth, TX: Holt, Rinehart and Wiston.

Berliner, D. C. (2001). Learning about and learning from *expert* teachers. *International Journal of Educational Research,* 35(5): 463-482.

Berliner, D. C. (2004). Expert teachers: their characteristics, development and accomplishments. In Obiols, R. B. & Martínez, A. E. et al (Eds.). *De la teoria a l'aula: formación del professorat i ensenyament de las Ciências Socials* (pp. 13-28). Barcelona: Universidad Autonoma de Barcelona.

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Berkeley, University of California: Operations Research Center.

Duke, R. A., & Simmons A. L. (2006). The Nature of Expertise: Narrative Descriptions of 19 Common Elements Observed in the Lessons of Three Renowned Artist-Teachers. *Bulletin of the Council for Research in Music Education*, 170, 7-19.

Ericsson, K. A., Ralf. T. H. K., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review,* 100 (3): 363–406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task. *Annual Review of Psychology*, 47,273-305.

Ericsson, K. A (2006). Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance on Representative Tasks. In A. Ericsson, N. Charness, P. J. Feltovich, & R R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp.223-241).Cambridge: Cambridge University Press.

Frensch, P. A., & Sternberg, R. J. (1989). Expertise and intelligent thinking: when is it worse to know better? In Robert. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 157-188). New Jersey: Lawrence Erlbaum Associates.

Feltovich, P. J., Michael. J. P., & Ericsson, K. A. (2006). Studies of Expertise from Psychological Perspectives. In Anders Ericsson, N. Charness, Paul. J. Feltovich, & Robert. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 41-68). Cambridge: Cambridge University Press.

National Board for Professional Teaching Standards – NBPTS (2002). *What teachers should know and be able to do*. Arlington: NBPTS.

Siedentop, D., & Eldar, E. (1989) Expertise, experience, and effectiveness. *Journal of Teaching in Physical Education*, 8,254-260.

Sosniak, L. A. (2006). Retrospective interviews in the study of expertise and expert performance. In A. Ericsson, N. Charness, P. J. Feltovich, & R R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp.287-301). New York: Cambridge University Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Musical Scales and Timing: Implications from Music Psychology for Eastern-European Ethnomusicology and vice versa

Rytis Ambrazevičius

Department of Audiovisual Arts, Kaunas University of Technology, Lithuania
Department of Ethnomusicology, Lithuanian Academy of Music and Theatre, Lithuania

rytis.ambrazevicius@ktu.lt

Abstract

Musical scales and timing have long been topics attracting tremendous attention among East-European ethnomusicologists, resulting in vast studies. These topics are also relatively frequent in the contemporary cross-cultural research of music perception. In addition, basics of music perception (such as categorization and the emic/etic problem) and a number of psychoacoustical phenomena (such as masking and roughness, relevant for the present study) are well known. The studies of music perception, however, are largely unknown in Eastern-European ethnomusicology and vice versa. The present paper examines how findings in music perception could contribute to Eastern-European ethnomusicological research and vice versa.

In the paper, certain paradigms and attitudes characteristic of Eastern-European ethnomusicological research on musical scales and timing are discussed. The transcriptions are considered indispensable data for the revelation of perceptual phenomena. Understanding of categorization and collision of emic systems is shown to result in dissolution of "aural ghosts", such as false categorization of pitch and time, "chromaticisms", and ostensible Ancient Greek modes found in ethnomusicological studies. Psychoacoustical roughness is shown to be responsible for some cases of scale formation in *Schwebungsdiaphonie* cultures.

As a majority of cross-cultural studies of music perception employ "exotic" musical cultures, it is shown that some "exotic" elements in "non-exotic" (e.g., East-European) cultures are typically overlooked. The phenomena found in these cultures could shed a light on, for instance, questions of the universal of asymmetries in musical scale, patterns and varieties of tonal hierarchies, non-isochronous pulse, non-accented meter, ametrical structures, and so on.

Introduction

The paper starts from discussion on the general issue of categorization (meaning pitch and time domains) as introduction into the subsequent problem of false categorization resulting in "aural ghosts." This phenomenon is exemplified by the cases of ostensible "Ancient Greek" modes, chromaticisms, and fictitious rhythm interpretations found in Eastern-European ethnomusicological studies. Then the cases of "peculiar" tonal structures and psychoacoustical roughness as determinant of musical scale are considered. Finally, some issues of interpretations of meter and rhythm are discussed; here the questions concerning non-accented meter and other relatively complicated phenomena in time domain are raised. Of course, these examples reflect just few instances pointing at the benefit of a closer collaboration of music psychology and ethnomusicology.

Categorization and Collision of Emic Systems

Categorization is well known phenomenon in psychology of music and in psychology, in general. Yet, the insider's and outsider's emic systems of categories do not, in general, coincide. Therefore, the outsider's inferences about phenomena intrinsic for distant cultures may be erroneous. Projecting the problem onto the domains of pitch and time results in the following statement: since the insider's and outsider's classifications of the pitch and time continuums are not necessarily the same, the musical scales and time structures perceived by an outsider are not necessarily those that are intended by an insider. This kind of phenomenon is sometimes referred to as an "aural ghost." For instance, suppose that thirds are not differentiated into minor, major and neutral thirds in some musical culture. The implication made by European musicologists about the usage of distinct versions of the third and about the corresponding chromatic changes would be merely a kind of "aural ghost" (Chenoweth, 1972, p. 50).

The cases of collision of two emic systems were demonstrated, for instance, by researching the peculiar pentatonics in African music (Arom, 1991; 1997; Arom & Voisin, 1998; Cross, 1999; etc.). The emic systems are characteristic of individual nuances, so interpreting the musical scale, as well as other musical parameters, depends not only on emic norms, but also on the individual. This problem is originally illustrated by transcription experiments, where experts in traditional music transcribe a sound recording and then comparatively analyze the (quite different individual) transcriptions (e.g. Listopadov, 1909; England *et al*., 1964). Imagine that each expert creates his/her own (different, individual) theory on the scale that is ostensibly characteristic of the heard melody. The theories ostensibly reflect the original musical thinking of the informant — they "correspond to reality." And yet they contradict each other. No more comment is needed.

The phenomenon of categorization and the problem of collision of emic systems resulting in "aural ghosts" is usually overlooked in Eastern-European ethnomusicologies (still, to some degree, in Western ethnomusicologies as well). Often there is the conscious or unconscious attempt to wedge scales of traditional musical practice into the framework of equal temperament, which is regarded as a perfect scale. Sometimes musicologists do not differentiate between 12TET and diatonics in general. The twelve-tone system or its diatonic subset are sometimes treated as the initial systems in the development of scales, and the scales used in traditional musical practice are seen as kind of devolution of these systems (Czekanowska, 1983, p. 99; see also Wallaschek, 1893/2009, p.

153–159). This devolution purportedly results in "incomplete", "defective" or "nontempered" scales.

Ostensible "Ancient Greek" Modes

A typical example of "aural ghosts" is the identification of "Ancient Greek" or "Gregorian" modes in traditional music. A large role in the propagation of this issue was played by a certain historical aspect, especially in the case of Eastern Europe. In the 19th century and in the beginning of the 20th century, the national renaissance movements in Europe stimulated interest in folk culture and folk music. Often their roots, or at least certain analogies, were sought in ancient European civilization, thereby showing their value and encouraging national pride. The Ukrainian researcher Sokalski (1888) made a great impact on setting this notion, which attempted to wedge all folk music into the firm frame of Greek tetrachords. As a result, the "Ancient Greek" modes appeared in treatises of traditional music; the romantic tradition tended to consider any deviations from the major-minor system as manifestations of these Ancient Greek modes.

Despite the fact that diatonicism is frequent in ethnomusicological studies, some attempts of Western ethnomusicologists to get unbiased insight into the original design of musical scales of European traditional music have been made; "one single loosely-knit modal folksong scale" (i.e. in between Mixolydian, Dorian, Aeolian, and other modes), loosely-knit "anhemitonic heptatonics" anchored in fifth and/or fourth, and other similar non-diatonic hybrid structures were revealed (Grainger, 1908–1909, p. 156, 158–159; Dal, 1956; Sevåg, 1974, p. 210–211; Keller, 1984, p. 101). Yet such discussions are actually absent in Eastern-European ethnomusicologies.

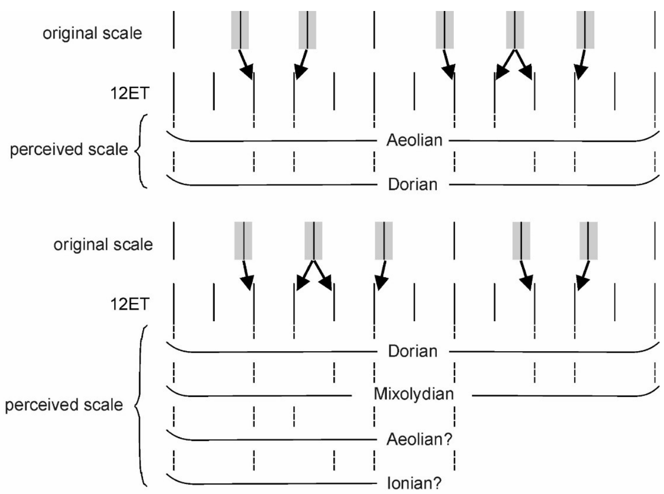


Figure 1. Most probable interpretations of hybrid (equitonic) scales. Top: hypothetical scale based on framework of fundamental, fourth and octave. Bottom: hypothetical scale based on framework of fundamental, fifth and octave. The rest of the scale degrees are characterized by wide zones of intonation.

It should be noted that quite a few ethnomusicologists have urged avoiding preconceived theoretical schemata and to look at the modes from the viewpoint of the culture's insider. Paradoxically, the researchers usually stopped short at merely proclaiming this requirement and did not allow for any scales in folk music other than diatonic scales, or, at least, those based on whole tones and semitones (Čiurlionytė, 1955, p. 11; Rubcov, 1973; etc.). Many researchers noted the significant differences of intonations from 12TET and still they considered the "deviations" not important and did not escape from the preconceived models.

Of course, one cannot fully deny the existence of the "Ancient Greek" modes in the Eastern-European traditional musics, especially if refer to relatively late recordings. However, at least for the Lithuanian case, it was shown that majority of the ostensible "Ancient Greek" modes are actually manifestations of hybrid scales (Ambrazevičius, 2006). Consider the instance of a tonal framework with relatively stable fourth or fifth anchors. When perceived, the unstable intermediate pitches are "attracted" to the closest equivalents in 12TET. In this way, the whole set of the Ancient Greek modes emerges (Figure 1).

One will probably notice that Phrygian and Lydian modes do not appear in Figure 1. Strikingly enough, the Ancient Greek modes were identified in traditional music in different degrees. The prevalence of Ionian, Mixolydian, Aeolian, and Dorian modes in contrast to the lack or negligible traces of Phrygian and Lydian modes is noted in many cases (Čiurlionytė, 1955, p. 11; 1969, p. 221; Rahn, 1990; Lippus, 1995, p. 122–124).

Ostensible Chromaticisms

Chromatic change is another phenomenon that seems to be non-existent in many cases. Figure 2 shows instances of real and fictitious chromaticisms. In the first case, the intonational versions (denoted by points) cluster into two categories, while in the second and third cases, they cluster into a single category. For example, in the second case, a musicologist biased toward 12TET will perceive major, minor, and (maybe) neutral thirds (in regard to C4); the latter one will correspond approximately to the uncertainty range. The original music language, however, operates on a single, generalized third.

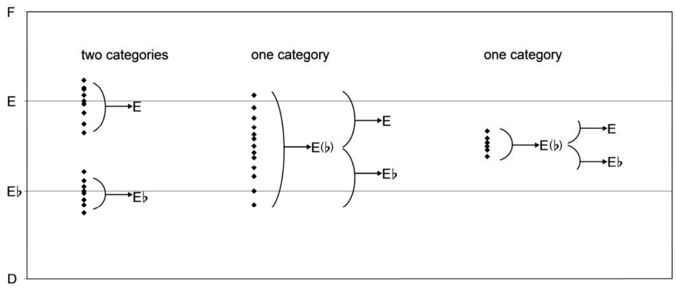


Figure 2. Schemes of real (1st) and fictitious (2nd and 3rd) chromaticisms.

Chromatic change presumes highly developed musical thinking on the part of a traditional performer, allowing him or her to play with chromatic "lights and shades of mood." I find the tendency met in various publications to consider real or seeming chromaticisms as a performer's deliberate "means of artistic expression." This tendency possibly comes from a classical conception of usage of chromaticisms. However, it would be naïve to presume that a performer of traditional music deliberately operates, say, by alternating Phrygian and Aeolian tetrachords to create certain emotional nuances. What if the alternation is merely a fictitious result of our perception caused, in reality, by a wide intonational zone? Maybe sometimes a systematic alternation of intonations for a scale degree is real, yet maybe it results from some physiological vocal qualities or unconscious manipulations (such as pitch performance rules), not from the performer's deliberate aestheticization.

Let's take an example from Rubcov's study. The intonation of a lyric song (Figure 3) is purportedly freely developing, "sometimes applying a Phrygian second, purely for expressional aims" (Rubcov, 1973, p. 72). Another example (Figure 4) wannabe shows how singers "vary pitch locations of particular [scale] degrees, sort of 'playing' with modal types" (*ibid*.). And, further, the discussed diatonic system wannabe is characteristic of an "emphatically deliberate usage of modal type providing purposeful 'play' with major-minor light-dark" (*ibid*., 73).

Figure 3. Example of "modal changes"; Russian tradition (Rubcov 1973, p. 71).

Figure 4. Example of "major-minor lights and shades"; Russian tradition (Rubcov 1973, p. 71).

"Modal changeability," "chain modes," "airy seventh," "transitional intervals," and other phenomena ostensibly reflecting chromatic thinking in Russian folk songs are found by Kastalski, Khristiansen, Shchurov, and many other Russian musicologists (Shchurov, 1998, p. 176–182). A mere glance at the transcriptions discussed, however, raises suspicions that, as a rule, "nontempered" pitches and/or wide zones of pitch intonation are actually at work.

"Aural Ghosts" in Time Domain

Figure 5. Example of artificial rhythm categorization. "Quail" (Četkauskaitė, 1998, p. 76, N6).

Similar problems are found in the interpretations of timing. The artificial rhythm categorization is especially manifested in the examples of free and irregular rhythm, *tempo rubato*: sequences of freely prolonged and shortened sounds are presented as complicated rhythm patterns. For instance, in the transcription of quail imitation three rhythm values are used: sixteenth, eighth, and dotted eighth (Figure 5). Measurements of the three *pliks* (IOIs) give 269, 299, and 255°ms, i.e. the eighth is actually longer than the dotted eighths (durations of the rest eighths – *neik neik neik* – are 252, 263, and 324°ms). Obviously, these durations cluster into a single category; the rhythm pattern in the discussed example uses only two values instead of three. The false categorization results from an attempt to transcribe "superprecisely". The number of rhythm values (categories) in the simple rhythm pattern performed *tempo rubato* is falsely increased.

Roughness / Dissonance as Determinant of Musical Scale

Quite a few musical cultures, in their polyphonies, favor dissonances (in terms of physiological acoustics) rather than consonances. This is described as various types of a psychoacoustically based "diaphony of beats" (*Schwebungsdiaphonie*); the style of performance where dyads of parts form predominantly rough sonorities or at least result in audible beats. The notion could be further extended to multipart music with a greater number of parts. This style is found, though not abundantly, throughout the world, such as in the Balkans, Indonesia, and elsewhere (Cazden, 1945; Brandl, 1989; Messner, 1989; etc.).

Lithuanian sutartinės can be regarded as a kind of *Schwebungsdiaphonie*. They are based on polyphonic and polyrhythmic patterns resulting from intertwining vocal parts. Typically, two female voices perform simultaneously, creating sequences of chords that, in their featuring seconds, are mostly dissonant.

The scales of sutartinės are considered as composed of sequences of semitones and whole tones, with some notes about the "non-temperation" of the intervals (Boiko, 1996; Vyčinienė, 2002). Yet the acoustical analysis shows that usually such categorization is a mere result of the discussed "aural ghosts." Take, for example, one sutartinė (Figure 6).

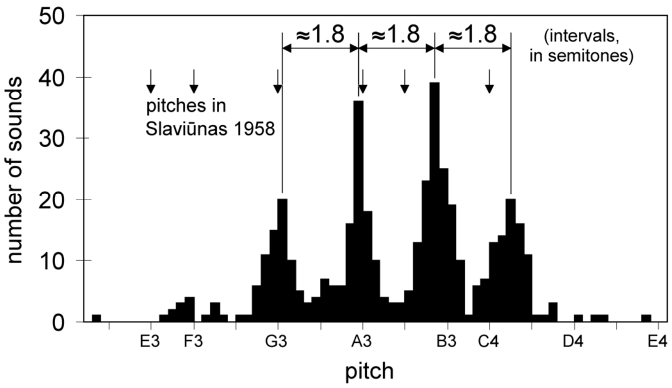


Figure 6. Histogram of pitches in the sutartinė "Myna, myna, mynagaučio lylio"; all pitches in all parts.

Actually, the tuning system has nothing in common with diatonics; there is no semitone/whole tone contrast in the sequence of intervals. The scale could be considered as "squeezed anhemitonics," because the intervals between the adjacent pitches are a bit narrower than the tempered whole tone. Figure 6 depicts the crude difference between the actual scale and the published conventional transcription ("pitches in Slaviūnas 1958"). While the shift of the entire scale is not important, as it could be attributed to a transcriber's attempts to apply simpler tonality (with fewer accidentals), to divergences in the speed of recording and playback, or to other circumstances, the very act of the "diatonicization" foregrounds an essential misinterpretation.

I have also collected the total distribution of dyad-intervals in 25 sutartinės (862 dyads all in all; Figure 7). The distribution shows that the majority of the intervals are seconds. The category of the interval is quite wide and does not split into the individual categories of minor and major seconds. As in the case of the separate sutartinė "Myna, myna, mynagaučio lylio," the seconds that are slightly narrower than a tempered whole tone (around 1.7 semitones) are the most preferred.

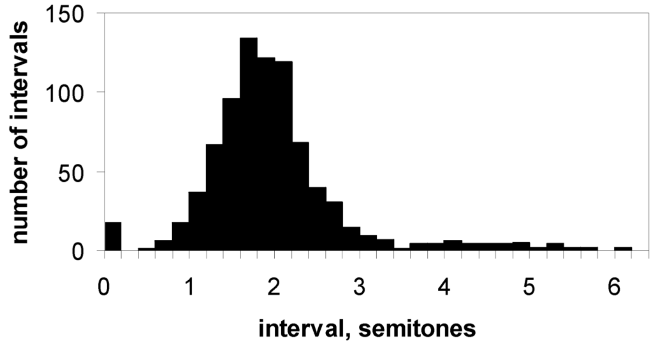


Figure 7. Intervals between the voices of the dyads in 25 sutartinės.

So, again, we come to a simple conclusion: the intervals of the second between the voices in sutartinė dyads comprise a relatively wide category centered at, approximately, 170–180 cents. What accounts for such a peculiar interval? Let's try to answer this question. Figure 8 shows the evaluations of intervals corresponding to maximum roughness / dissonance gathered from different psychoacoustical studies (Terhardt, 1968, p. 219; Fastl & Zwicker, 2007, p. 259; Leman 2000, p. DAFX-5; Rakowski, 1982; Vassilakis, 2001, p. 197–198; Hutchinson & Knopoff, 1978). The curves were composed based on the formulas and interpolations of the graphically presented results from the discussed sources. The confusion between dissonance, roughness, and its possible types probably explains why the results of the experiments show significant discrepancies. A closer examination of Figure 8 reveals that roughness is typically associated with larger interval sizes, and that sensory dissonance is associated with narrower interval sizes (see the discussion in Ambrazevičius, 2014-2016).

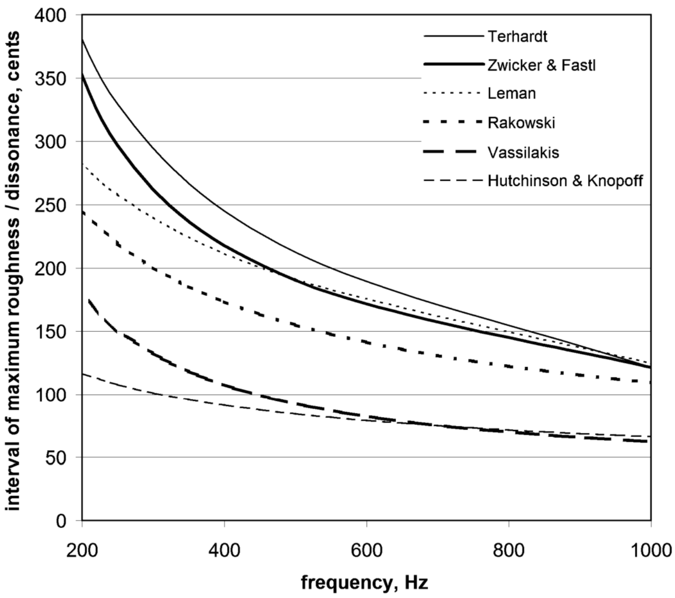


Figure 8. Dependence of maximum roughness / dissonance on the central frequency. See the body text for details.

Now let's return to sutartinės. On the one hand, because of the first formant, the most intense frequency range in the spectra of singing voices embraces roughly 400–800 Hz in the case of sutartinės (i.e. female voices). This corresponds to the second or third (or sometimes fourth) harmonics. On the other hand, as already discussed, intervals slightly narrower than the tempered whole tone are found to be the most preferred in sutartinės.

If we apply these values of the frequencies to the graphs in Figure 8, we can conclude that the singers were aiming for maximum roughness (and not dissonance). In other words, they "clashed" their voices in the intervals that would most probably create the maximal sense of roughness. It seems that specifically roughness was meant by the singers of sutartinės when describing their sonorities as "clashing" (clanging, warbling; but not "cutting" which would point to sensory dissonance and narrower intervals). Singers considered the "strong clash" to be an essential quality and marker of a congenial performance.

To summarize, the intervals in sutartinės performances are based on the psychoacoustical phenomenon of roughness (perhaps even maximum roughness). This results in the peculiar scale structures deviating considerably from 12TET; they comprise intervals slightly narrower than a tempered major second (a "squeezed" whole tone); with quite a wide range of variations.

Peculiar Tonal Structures

A polyphonic excerpt cut from the original recording of the sutartinė discussed in the previous chapter (Figure 6) was chosen for the probe tone test. The 9 sound recordings consisting of the excerpt and the succeeding probe tone were presented to two groups of listeners, "experts" (sutartinės performers) and "non-experts" (students from Lithuanian universities with some musicological background but without practical experience in sutartinės singing).

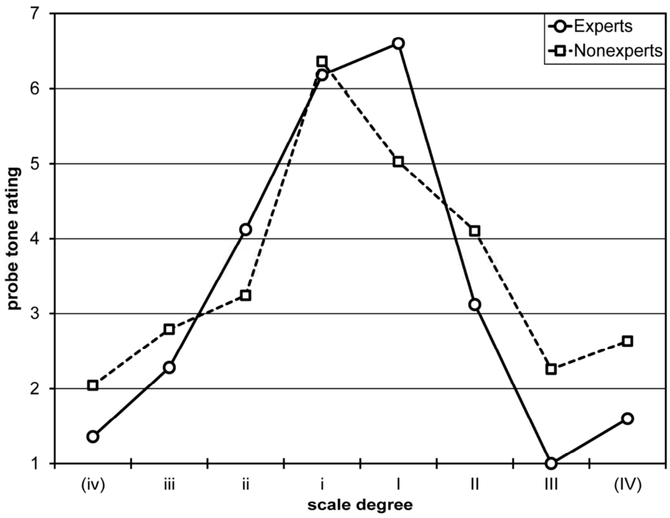


Figure 9. The average ratings given to the probe tones by the two groups of experts and non-experts, for the excerpt from the sutartinė "Myna, myna, mynagaučio lylio". The degrees that do not belong to the scale (employed only as probe tones) are marked in brackets.

The results averaged across the groups are presented in Figure 9. The tonal profiles are approximately bell-shaped. This means that the tonal structure is comprised of the tonal center ("double tonic"; marked as "i" and "I") and the scale degrees with the salience gradually decreasing towards the margins of the scale.

The "experts" give approximately the same ratings for both degrees constituting the "double tonic." The "non-experts" distinguish single central pitch. Possibly because they are more influenced by the usual scale with one tonic. Moreover, the "experts" make bigger hierarchical differences between the degrees, whereas the "non-experts" show a flatter tonal profile. Thus, the hierarchical structure is more distinctly cognized by the sutartinės singers.

Obviously, the cognitive basis of this structure is completely different from that of the major-minor system (Krumhansl, 1990; etc.).

Issues of Meter and Rhythm

Generally, the cross-cultural studies dealing with perception of time structures, and relevant for the present paper, showed fairly reliable recognition of the original qualities, yet with some influences of cultural exposure and musical training (cf. Drake & Ben El Heni, 2003; Toiviainen & Eerola, 2003; Soley & Hannon, 2010; Yates *et al.*, 2017). However, perception of mostly simple (or not too complex) accented meter is analyzed in the studies. But what about the "ethnomusicological" non-isochronous pulse, non-accented meter ("time-measuring/time counting" meter; Alexeyev, 1990, p. 88), ametrical structures, and the specific meter-phrasing relations found in ethnomusicological studies? (Of course, this question could encompass not only instances of traditional music.)

Figure 10. Song „Jurja, geras vakaras“ (Barauskienė *et al.*, 1962, p. 678).

Figure 11. Song „Oi teka bėga“ (Čiurlionytė, 1938, p. 195).

Figures 10 and 11 show a couple of Lithuanian examples. The first one is a typical case of non-accented meter, yet the "time counting" is clearly perceived. The second example presents the case when even the "time counting" is impossible as the pitches are intoned with very free, prolonged and shortened durations. The rhythm values presented in the transcription are rough approximations; they can be interpreted differently when listening repeatedly. Most probably, certain new phenomena could be revealed when considering such instances from the psychological perspective.

Discussion

Several examples combining perspectives of music psychology and ethnomusicology (mostly centering at Eastern-European ethnomusicology) were discussed. Understanding of psychological categorization and especially different categorization in different emic systems could help resolving the issues of biased perception resulting in "aural ghosts," such as ostensible "Ancient Greek" modes, chromaticisms, and fictitious rhythm patterns.

Quite surprisingly, this (perhaps simple) problem is often overlooked, and not only in Eastern-European ethnomusicologies, but in studies of music perception as well. Asymmetry in musical scale is usually considered as musical universal "getting our bearings" regarding where we are in the scale (cf. Snyder, 2000, p. 140). But what if the "getting our bearings" is achieved by other means than scale asymmetry? And what if this faculty is only partially needed or not needed at all (presumably in the case of more archaic musical thinking)? This case has evidences in music cultures characteristic of equidistant scales. Roughly equidistant scales cannot be excluded from the consideration, the same as a system of only approximate 12TET whole tones and semitones nevertheless does mean manifestation of diatonic. (See e.g. the recent study on African scales (Ross & Knight, 2017); some examples of this scale type are discussed in the current paper.)

Concerning tonal hierarchies: on the one hand, Eastern-European and other traditional musics present examples of peculiar and not major-minor profiles. On the other hand, the probe-tone technique and other psychological techniques could be employed by ethnomusicologists for objectivized investigation of these peculiar and even "normal" scales in traditional music. Ethnomusicologists should also consider the factor of psychoacoustical roughness in scale formation, for instance, in *Schwebungsdiaphonie* cultures.

Ethnomusicological studies present a wide variety of (again, "peculiar") rhythm patterns and meters, not stopping at simple accented meters. This material could be employed in psychological research, possibly resulting in revelation of certain new phenomena of music perception. Ethnomusicology could borrow variety of methods of timing analysis used in acoustical and psychological research. "Tempo curves" serve as an example.

On the one hand, the phenomena of music perception are studied with a help of psychological experiments. On the other hand, these phenomena are abundantly demonstrated via examples of traditional music, both live performances and transcriptions. This way the local and universal features could be articulated more reliably. The transcriptions serve as indispensable data for the revelation of perceptual phenomena. First, the ethnomusicological transcriptions reveal the variety of phenomena interesting from the prospective of music psychology. Second, as the results of perception, they provide a corpus of materials for the research of outsider's perception in terms of emic / etic.

A majority of the discussed questions are actually mutual interdisciplinary psychological-ethnomusicological questions and fields of study, and a closer collaboration of the two disciplines could contribute to the development of both.

References

[Alexeyev] Алексеев, Э. (1990). *Нотная запись народной музыки: теория и практика [Notation of folk music: Theory and practice]*. Москва: Советский композитор.

Ambrazevičius, R. (2006). Pseudo-Greek modes in traditional music as result of misperception. In I M. Baroni, A. R. Addessi, R. Caterina, M. Costa (Eds.), *Proceedings of the 9th International Conference on Music Perception and Cognition. 6th Triennial Conference of the European Society for the Cognitive Sciences of Music. Alma Mater Studiorum University of Bologna, Italy, August 22–26, 2006* [CD] (pp. 1817-1822). Bologna: Bononia University Press.

Ambrazevičius, R. (2014-2016). Dissonance/roughness & tonality perception in Lithuanian traditional Schwebungsdiaphonie. *Journal of Interdisciplinary Music Studies, 8*, 39-53.

Arom, S. (1991). *African polyphony and polyrhythm: Musical structure and methodology*. Cambridge: Cambridge University Press.

Arom, S. (1997). Le "syndrome" du pentatonisme africain [The "syndrome" of African pentatonic]. *Musicæ Scientiæ, 1*, 139-163.

Arom, S., & Voisin, F. (1998). Theory and technology in African music. In R. Stone (Ed.), *The Garland encyclopaedia of world music. Vol. 1. Africa* (pp. 254-270). New York: Garland.

Barauskienė, V., Kazlauskienė, B., Uginčius, B., & Aleksynas, K. (Eds.) (1962). *Lietuvių tautosaka [Lithuanian folklore]. Vol. 1. Dainos [Songs]*. Vilnius: Lietuvos muzikos akademijos muzikologijos instituto etnomuzikologijos skyrius.

Boiko, M. (1996). On the interaction between styles in Baltic folk music: Sutartinės polyphonie and East Baltic refrain songs. In U. Wegner, M. P. Baumann, & A. Simon (Eds.), *European studies in ethnomusicology: Historical developments and recent trends* (pp. 218-236). Wilhelmshaven: Florian Noetzel Verlag.

Brandl, R. M. (1989). Die Schwebungs-Diaphonie – aus musikethnologischer und systematisch-musikwissenschaftlicher Sicht. In C. Eberhardt & G. Weiss (Eds.), *Südosteuropa-Studien. Bd. 40, Volks- und Kunstmusik in Südosteuropa* (pp. 51-67). Regensburg: Gustav Bosse Verlag.

Cazden, N. (1945). Musical consonance and dissonance: A cultural criterion. *Journal of Aesthetics and Art Criticism, 4*, 3-11.

Četkauskaitė, G. (Ed.) (1998). *Lietuvių liaudies muzika. II. Aukštaičių dainos. Šiaurės rytų Lietuva / Lithuanian folk music. II. Songs of Aukštaičiai. North-eastern Lithuania* [book with CDs]. Vilnius: Lietuvos muzikos akademijos muzikologijos instituto etnomuzikologijos skyrius.

Chenoweth, V. (1972). *Melodic perception and analysis. A manual on ethnic melody*. Papua New Guinea: Summer Institute of Linguistics Ukarumpa, E. H. D.

Čiurlionytė, J. (1938). *Tautosakos darbai [Folklore studies]. Vol. 5: Lietuvių liaudies melodijos [Lithuanian folk melodies]*. Kaunas: Lietuvių tautosakos archyvas.

Čiurlionytė, J. (1955). *Lietuvių liaudies dainos [Lithuanian folk songs]*. Vilnius: Valstybinė grožinės literatūros leidykla.

Cross, I. (1999). AI and music perception. *AISB Quarterly, 102*, 12-25.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.

[Czekanowska] Чекановска, А. (1983). *Музыкальная этнография. Методология и методика [Musical ethnography. Methodology and techniques]*. Москва: Советский композитор.

Dal, E. (1956). Scandinavian folk music – a survey. *International Folk Music Journal, 8*, 6-11.

Drake, C., & Ben El Heni, J. (2003). Synchronizing with music: Intercultural differences. *Annals of the New York Academy of Sciences, 999*, 429-437.

England, N. M., Garfias, R., Kolinski, M, List, G., Rhodes, W., & Seeger, Ch. (1964). Symposium on transcription and analysis: A Hukwe song with musical bow. *Ethnomusicology, 8*, 223-277.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics. Facts and models*. Berlin, Heidelberg, New York: Springer.

Grainger, P. (1908–1909). Collecting with the phonograph. *Journal of the Folk Song Society, 3*, 147-242.

Hutchinson, W., & Knopoff, L. (1978). The acoustical component of Western consonance. *Interface, 7*, 1-29.

Keller, M. S. (1984). The problem of classification in folksong research: A short history. *Folklore, XCV*, 100-104.

Leman, M. (2000). Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (SIM). In *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7–9, 2000* (pp. DAFX-1-DAFX-6).

Lippus, U. (1995). *Linear musical thinking*. Helsinki: University of Helsinki.

[Listopadov] Листопадов, А. (1909). К вопросу о записях народных песен [On the recordings of folk songs]. *Музыка и жизнь, 1*, 5-6.

Messner, G. F. (1989). Jaap Kunst revisited. Multipart singing in three East Florinese villages fifty years later: A preliminary investigation. *The World of Music, 31*, 3-51.

Rahn, J. (1990). An introduction to English-language folksong style (II): Tonality, modality, harmony, and intonation in LaRena Clark's traditional songs. *Canadian Journal for Traditional Music*. Retrieved from http://cjtm.icaap.org/content/18/v18art4.html

Rakowski, A. (1982). Psychoacoustic dissonance in pure-tone intervals: Disparities and common findings. In C. Dahlhaus & M. Krause (Eds.), *Tiefenstruktur der Musik* (pp. 51-67). Berlin: Technische Universität Berlin.

Ross, B., & Knight, S. (2017). Reports of equitonic scale systems in African musical traditions and their implications for cognitive models of pitch organization. *Musicæ Scientiæ*.

[Rubcov] Рубцов, Ф. (1973). Основы ладового строения русских народных песен [Fundamentals of modal organization of Russian folk songs]. In Ф. Рубцов (Ed.), *Статьи по музыкальному фольклору [Papers on musical folklore]* (pp. 8-81). Ленинград, Москва: Советский композитор.

Sevåg, R. (1974). Neutral tones and the problem of mode in Norwegian folk music. In G. Hilleström (Ed.), *Studia Instrumentorum Musicae Popularis III* (pp. 207-213). Stockholm: Musikhistoriska Museet.

[Shchurov] Щуров, В. (1998). *Стилевые основы русской народной музыки [Stylistic fundamentals of the Russian folk music]*. Москва: Московская государственная консерватория.

Slaviūnas, Z. (1958). *Sutartinės. Daugiabalsės lietuvių liaudies dainos [Sutartinės. Lithuanian multipart folk songs]*. Vol. 1. Vilnius: Valstybinė grožinės literatūros leidykla.

Snyder, B. (2000). *Music and memory. An introduction*. Cambridge: The MIT Press.

[Sokalski] Сокальский, П. (1888). *Русская народная музыка, великорусская и малорусская в ее строении мелодическом и отличия ее от основ соременной гармонической музыки [Russian folk music, its melodic structure and differences from the principles of contemporary harmonious music]*. Харьков.

Soley, G., & Hannon, E. E. (2010). Infants prefer the musical meter of their own culture: A cross-cultural comparison. *Developmental Psychology, 46*, 286-292.

Terhardt, E. (1968). Über akustische Rauhigkeit und Schwankungsstärke. *Acustica, 20*, 215-224.

Toiviainen, P., & Eerola, T. (2003). Where is the beat? Comparison of Finnish and South-African listeners. In R. Kopiez, A. C. Lehmann, I. Wolther, & C. Wolf (Eds.), *Proceedings of the 5th Triennial ESCOM Conference. 8-13 September 2003* (pp. 501-504). Hanover: Hanover University of Music and Drama.

Vassilakis, P. N. (2001). Perceptual and physical properties of amplitude fluctuation and their musical significance. (Doctoral dissertation). University of California, Los Angeles, USA.

Vyčinienė, D. (2002). *Lithuanian polyphonic songs sutartinės*. Vilnius: Vaga.

Wallaschek, R. (1893/2009). *Primitive music: An inquiry into the origin and development of music, songs, instruments, dances, and pantomimes of savage races*. New York: Cambridge University Press.

Yates, Ch. M., Justus, T., Atalay, N. B., Mert, N., & Trehub, S. E. (2017). Effects of musical training and culture on meter perception. *Psychology of Music, 45*, 231-245.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Music Protolanguage. Comparison to Motherese in Human Mother - Infant Interactions and animal

Alessandra Anastasi

Department of Cognitive Sciences, Psychological, Educational and Cultural Studies, University of Messina, Italy

anastasia@unime.it

Abstract

The purpose of this paper is to demonstrate the theory according to which the origin of the language is detectable in the continuity of morphological structures which influenced the use of vocalizations of prototypical music and language. According to the Darwinian Biolinguistics Hypothesis, the acquisition of language implies that some of its elements are the result of clear biological constraints as well as natural ones, which have ensured its continuity. Assuming that the language has evolved on the basis of structures that allow vocalism, I will try to illustrate how the inception of a protolanguage is found in the very first forms of communication between mother and child amongst non-human primates (motherese). This special relationship is often marked by a form of simple communication, high intonation and singsong which ensure the continuation of the mother-child bond. Just like in the blabbering of babies, the prosodic unit produced during animal parental care is the result of a natural bond that determines which sound goes after another, apart from those physical restrictions that limit the production of speech sounds. The existence of forms of infant-direct speech in some primates (e.g. *Macaca fuscata*) suggests, in addition to the social function of the infant-specific vocalization and thanks to the melodic intonation and rhythmic signal, some information on the emotional condition of the speaker. These data suggest that there may be a continuum, in the ability to transmit information, that becomes particularly evident in some mother-child communication. Therefore, it can be assumed assume that spoken language may have developed first after the vocalizations used by the mother-child communication across non-human primates, and later by female hominids who interacted with their children.

Introduction

The study of musical perception and cognition as a privileged subject, which served as base to develop hypotheses on the origin of language, boasts great supporters but also bitter enemies. When Pinker (1997) borrowed the concept of spandrel (Gould & Lewontin 1979) to describe music as an evolutionary accident that language dragged behind, he ignored the mechanism of biological cause-effect of the music. The famous linguist considered music useless, an auditory cheesecake, a by-product of natural selection whose only purpose was to tickle the true adaptive functions. In his perspective, music would not have a lasting purpose equal to verbal language; in fact, the disappearance of music would not imply any vital loss for our lifestyle. Yet, recent studies (Trehub, 2003) showed how music processing helps the little sapiens to get ready for language (prosody): the evolutionary origin of music would thus be founded, because its capillary presence among human beings satisfies the biological criterion of being diffused in a species, additionally implying the presence of specialized processing structures. The ability to acquire musical skills can therefore be seen through the perspective of an evolutionary adaptation; the latter, shaped by natural and sexual selection, ensured the universality of music also through its capacity to reinforce interpersonal relationships and group cohesion (Miller, 2000). The first step of this process may be the inception of a communication system that is not yet referential, similar to singing and now better known as motherese, able to create a sort of emotional contagion that consists in the tendency to mimic and automatically synchronize expressive behaviours, movements and vocalizations of another person (Hatfield et al. 1994, Falk 2009); eventually, this phenomenon affected the nature of our language. Vocalizations produced by the first hominids were made possible using an apparatus common to all mammals (and still noticeable in nature nowadays). Subsequently, our biological nature structurally conditioned the morphological variants that bound us first to a form of musical vocalization (the motherese itself), and then to the actual language.

The prosodic and musical style of language

Music communication is the bridge between animal communication and human language. The data of ethological, paleoanthropological and psychobiological origin clearly showed that speaking of "specialty" of the human language is incorrect. Obviously, the vocalization strategies shown by non-human primates vary according to the structural peculiarities of each species: it is evident that each animal species has its own "phonatory equipment", applied to different ecological-social contexts (Struhsaker 1967; Mitani & Nishida 1993). We could think, for example, about the social function of Indri's song, constituted by specific contextual information (Maretti et al., 2010). However, the use of ethological comparison not only allows to identify anatomical differences and phylogenetic derivations of central and peripheral morphological traits responsible for the human vocal production (Fitch 2010), but also to understand how the biological constraints of a species-specific vocality are employed to build social relations (Anastasi 2014). The similarities in the communication between human and non-human primates suggest the existence of deep evolutionary roots found in the prosodic/musical components of speech. The prosodic elements of verbal language constitute the common component for word and sound; the spoken language contains its own musical component, which is manifested through intonation, rhythm and intensity. The introduction of prosody within the human language mainly originates from the right side of the brain and, oppositely to what many people think (linguists included), the tone of voice is not excluded from language, as it proved crucial in the communication of our ancient, non-verbal progenitors (Falk 2004). For this reason, prosody should be analysed also because of its evolutionary role. According to Falk (2009), at some point during the evolution, mothers began to maintain a contact with their children not only with physically, but also vocally: the result would have been a special language which took the name of motherese (Dissanayake 2004, Newman 2004). Compared to the actual language, motherese is slower and repetitive, with a higher tone and a simpler vocabulary; it is a fairly musical language, which gives a certain intensity and modulation to the tone of voice of the language of adults. Precisely like a lullaby or a nursery rhyme sung to children, the motherese melody firstly conveys an emotional meaning, only then leaving room for the actual language. From the harmony of the sounds emitted by the mother and processed by the child as a prosodic component it is possible to comprehend single words, up to the most complex activities of comprehension and sentence construction. According to Newman (2004) the motherese paradigm can be applied to the other mammals as well and, according to his view, motherese also includes vocalizations or songs used by primate mothers use to attract the attention of their offspring. Since the brain areas responsible for the vocalization between mother and child are essentially those used for language, Newman hypothesizes that spoken language may have developed from the vocalizations which hominid females sent to their children. At this point, the definition of motherese could include the vocalizations and songs which primates use to communicate with their offspring. Masataka (2003) offers an interesting approach as he provides new perspectives on the language and music evolution. From his observations it emerges that, even though human beings in their first years of life emit language-like sounds, the process is still divided in two phases: the first, occurring between the sixth and the eighth week, consists in the emission of vowel-like sounds. Newborns start producing sounds through mutual exchanges with their caregivers. In this case, the quality of the adults' vocal responses affects the vocalizations of the newborn in development. The second phase, on the other hand, starts with the babbling inception at about the eighth month. The same mechanism would occur in Japanese macaques (Macaca fuscata) which, via COOS (callbacks) emission, manage to maintain contact not only with the children but also with the remaining members of the group, without any physical contact. In other words, the call becomes an actual tool of social cohesion. The ethological comparison becomes thus fundamental at the time to explain to explain how the first aspect of the vocal control was, probably, the prosodic language, which also constitute the actual base for song. Although equipped with minimal vocal control, the first hominids would have been able to develop a range of sounds, later evolved into a primitive song. The latter is more difficult to define as it assumed a central role in the culture of sapiens; but it can still be approached within a prelinguistic context, as it is a precursor of the language itself. The hypothesis of the care and breeding of newborns seems, to date, the most accredited explanation if one attempts to explain the evolution of music by referring to its adaptive function (Fitch 2005). The frequent evidence scattered across the world of primates seem to suggest that the prosodic vocalizations, from which the music would derive, played a major role. Whether lullabies for children or vocalizations to keep in touch with offspring, the analogies between music and language cannot certainly be ignored.

Biological Constraints

The thesis according to which the biological constraints of language – seen as structural and morphological conditioning (Pennisi & Falzone 2014) – allowed the *sapiens'* typical way to use language to decode and represent their knowledge about the world (Hagoort et al., 2004), has engendered in these years a discussion about the species-specificity of articulated language. As described by Balari and Lorenzo (2015), language is clearly the result of a biological evolution which defined the organs responsible for the verbal articulation and the hearing through neural development, in order to obtain the neural control of certain tasks linked with the articulation and processing of vocal sounds. The vocal, visual, tactile and species-specific signals characterize the social life of many animal species and, specifically, the parental care of many non-human primates whose infantile vocalizations seem to contain a greater amount of information that varies, probably, depending on the socio-ecological context and the characteristics of different primate species (Maestrepieri & Call 1996). These observations suggest that spoken language may have developed from vocalizations used first during mother-child interaction in non-human primates, and later by hominid females who addressed their children. In this perspective, according to Sarah Hrdy (2009), the growth of children within a social context broader than the family would represent the social success of the human species. In a very distant way from Falk (2009), who hypothesized that the roots of verbal communication of our species are not to be sought in the coordination between men during hunt, but rather in the exclusive relationship between mother and son, Hrdy's study (2000) gave a decisive and revolutionary approach when it proved that infanticide is systematic in some primates. Therefore, maternal care is all but obvious, although it is subject to the female's decisions about the survival possibility of herself and her offspring. If we look at the nature of human language as a consequence dictated by morphological and cognitive structures, we must bear in mind that some of its components seem to imply clear biological constraints that allowed a sort of conditioning of those proto-musical and linguistic vocalizations. Our species has therefore inherited the necessary skills to manifest language through the involved morphological structures, which in turn that allow us to emit a wide range of sounds to meet satisfy our communication needs. Studying our communication system inevitably requires a basic knowledge of how sounds are produced, as well as the anatomy which allows them. The human language uses rapid variations in many acoustic parameters to produce an astonishing number of words. The underlying mechanism of this process is very similar both in man and other mammals: the air exhaled through the lungs allows the vocal cords in the larynx to oscillate; in turn, the vocal oscillation determines the pitch of the produced sound. The generated acoustic energy travels along the vocal tract (pharynx, oral and nasal cavity) where it is filtered and, finally, outwards through the nostrils and lips. Frequencies that tend to lose intensity during their 'travel' inside the vocal duct are called formant frequencies, their value determined by the length and shape of the vocal tract. They can be quickly modified during the speech via coordinated movement of various joints (e.g. tongue, lips and soft palate), through which a wide variety of sounds is produced (Fitch 2000). The morphological body structures act as constraints for the possible vocal performances. The performance of each animal species, therefore, is defined by the biological constraints that allow it, in relation to the environment in which the species lives. There are obviously species-specific differences across vocal performances produced by each species. An exceptional case is that of the pygmy marmoset (*Cebuella pigmea*), known to be the smallest monkey in the world, which communicates with its conspecifics emitting guttural sounds with an acute voice (Elowson et al., 1998). The characteristic of their vocal activity is a more intense vocalic peak during the call production; such peak is characteristically made of a rhythmic and repetitive call string, which generally emerges during the first two weeks of life. Vocal activity peaks occurred just before weaning, (roughly six to eight weeks). According to what Elowson and collaborators described (1998), this behaviour can be compared to the infant babbling in human primates which, as we know, represents the first manifestation of an emerging linguistic ability. The *sapiens*, instead, starts from some universal non-specific language sounds (babbling) and improves until the production of a sequence of sounds, specific to the language of the parents, with a much higher quantity and combinational potential than the initial babbling. This determines a communication which is totally species-specific, bound by a common evolutionary base and specific body structures, and determined by the performativity that these structures allow.

Music Protolanguage

As a human and universal behaviour, music cannot be separated from biology. It is precisely this shared link that brings us to its second mode of expression: singing. What we commonly call singing is nothing but the musical product of our phonatory organ; the emission of sounds ordered by rhythm and height and their combination generate, in fact, the melody of the vocal expression. Singing appears to be the most ancient and instinctive musical instrument, and having developed naturally is bound by our biological and cognitive system, just like language. The hypothesis of the existence of a musical protolanguage induces to explore analogies and differences between singing and verbal language. Compared to the compositional theory on the birth of language, according to which it is possible that our ancestors had a relatively wide range of words (albeit with a limited grammar), the perspective of Wray (2002) on phonetic segmentation as a matrix of language seems to be the most convincing. The development of the compositional language would arise from the segmentation of the initially non-compositional protolinguistic signals of primates, pretty much like children resorting to the segmentation of phonetic units during language acquisition. It should be noted, however, that children segment a signal which is already syntactically articulated and semantically compositional (Mithen 2006). The articulated language, result of the coordination of peripheral anatomical structures, is crucial to produce sounds that are proper to a language (motherese included); however, this does not imply that man is unique only because he has the means to produce language. The primates, which are paradoxically not so phylogenetically close to us, are endowed with a vocal tract that allows them to vocalize and elaborate rather complex songs; this is the clear demonstration of how anatomical precursors are functional precursors within species. To believe that species capable of producing songs are natural precursors of language proves the possibility (once a parallel between ontogenetic and phylogenetic development is established) of foreseeing that all those components belonging to our peripheral structures first emerged as wholly independent organs. Only after a subsequent re-functionalization process they were ready to produce vocalizations and songs in non-human primates first and, with the appearance of more evolved life forms, as musical protolanguage and language in the sapiens later. In such a scenario, singing can be included in the forms of communication between human beings and it is precisely here that it we can find a possible explanation in evolutionary terms: singing could have developed to convey concrete meanings before the birth of verbal language, taking the form of a protolanguage on which the language itself would be installed, at least from the melodic/prosodic point of view. Music would therefore be closely linked to language, probably representing a part of the language itself. In what appears to be a “linguistic mosaic”, we identified traits similar to other species; that is, the vocal tract morphology, which allows us to interpret the song as a natural precursor of language and species-specific traits that marked the differences within the human phylogenesis (prosody, syntax, grammar). What emerges from this discussion is that, although human language can be qualitatively different from any form of animal communication, in terms of recursion and referentiality, the appearance of a musical protolanguage places the development of human vocalization in a context that appears to comply with the functional criteria of natural selection proposed by Tinbergen (1951). The evolutionary phylogenesis of the transformation into a musical man would be found in those response mechanisms included in the first modes of communication. The human being does not choose to communicate with language, but this is the binding element to communicate with his conspecifics, and his ability to segment, combine and decode the vocal production is the basis of what Mithen (1999) defines “cognitive fluency”. Based on this observation, the fork leading from the musical protolanguage of the hominids to the language-cognition interface is the turning point of the species-specificity of our language.

Conclusions

The similarities between human and animal singing show how they both carry a message and reflect an emotional state. Many behavioural characteristics related to music can be interpreted as physical attachment: this is the case of lullabies and the relationship developed between mother and child; or singing and collective music, which crystallize the group's coordination and bonding. Studies on non-human primates revealed the existence of conceptual tools which have produced the idea that in these animals there is a referential communication system (Premack 1985). The most radical hypothesis on this matter is that while animal vocalizations do not carry the referential power of our words (both in objective terms and mental states), they can still represent a sufficiently advanced system to justify classification as a precursor. Considering that the vocalizations of primates can represent a form of protolanguage, it is opportune to implement a precise conceptualization of the human primates’ morphogenetic instruments, as well as a conceptualization of the selective pressure types emerged during evolutionary history. The human musical ability seems to be born from an array of characteristics with varied shapes, characterized by the complex social and environmental challenges faced by our species during its evolution. The mosaic of vocal evolution, in which it is possible to identify the link between animal communication and the musically organized communication of human beings, suggests the means through which every element of the array of characteristics may have evolved. Contemporary studies provided us with an explanatory framework on the functioning of non-human primates' communication, as well as the undeniable analogies in the voice biomechanics that constitutes the prelinguistic substratum of communication. Studying music through an interdisciplinary approach, as is typical of cognitive science, helps us to weave the biological and cultural dimension of an element: language itself, whose evolutionary history, is, in some ways, still shrouded in mystery.

References

Anastasi, A. (2014). Il ruolo della comunicazione nella costruzione della dimensione sociale. In A. Falzone, S. Nucera, F. Parisi (Eds.), *Le ragioni della natura. La sfida teorica delle scienze della vita* (pp. 29 - 40). Messina-Roma: Corisco Edizioni.

Balari, S., & Lorenzo, G. (2015). The end of development. *Biological Theory*, *10*(1), 60-72.

Dissanayake, E. (2004). Motherese is but one part of a ritualized, multimodal, temporally organized, affiliative interaction. *Behavioral and Brain Sciences*, 27(4), 512-513.

Elowson, A. M., Snowdon, C. T., & Lazaro-Perea, C. (1998). Infant 'babbling' in a nonhuman primate: Complex vocal sequences with repeated call types. *Behaviour*, *135*(5), 643-664.

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese?. *Behavioral and Brain Sciences*, *27*(4), 491-503.

Falk, D. (2009). *Finding our tongues. Mothers, Infants, and the Origins of Language*. New York: Basic Books.

Fitch, W. T. (2000). The phonetic potential of nonhuman vocal tracts: comparative cineradiographic observations of vocalizing animals. *Phonetica*, *57*(2-4), 205-218.

Fitch, W. T. (2005). The evolution of music in comparative perspective. *Annals of the New York Academy of Sciences*, *1060*(1), 29-49.

Fitch, W. T. (2010). *The evolution of language*. Cambridge, UK: Cambridge University Press.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B*, 205(1161), 581-598.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441.

Hatfield, E., Cacioppo, J. & Rapson, R.L. (1994). *Emotional contagion*. Cambridge, UK: Cambridge University Press.

Hrdy, S. B. (2000). *Mother Nature: Maternal Instincts and How They Shape the Human Species*. New York: Pantheon Press.

Hrdy, S. B. (2009). *Mothers and Others: The Evolutionary Origins of Mutual Understanding*. Cambridge, MA: Harvard University Press.

Maestripieri, D., & Call, J. (1996). Mother-infant communication in primates. In *Advances in the Study of Behavior* (Vol. 25, pp. 613-642). Academic Press.

Masataka, N. (2003). *The Onset of Language*. Cambridge, UK: Cambridge University Press.

Miller, G. (2000). Evolution of music through sexual selection. In N. Wallin, B. Merker & S. Brown (Eds.), *The Origins of Music* (pp. 329-360). Cambridge, MA: MIT Press.

Mitani, J. C., & Nishida, T. (1993). Contexts and social correlates of long-distance calling by male chimpanzees. *Animal Behaviour*, *45*(4), 735-746.

Mithen, S. (1999). *The Prehistory of the Mind: The Cognitive Origins of Art, Religion and Science*. London: Thames and Hudson.

Mithen, S. (2006). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. USA: Harvard University Press.

Newman, J. D. (2004). Motherese by any other name: Mother-infant communication in non-hominin mammals. *Behavioral and Brain Sciences*, *27*(4), 519-520.

Pennisi, A., & Falzone, A. (2014). Residuals of intelligent design in contemporary theories about language nature and origins. *Humana. Mente*, *27*, 161-183.

Pinker, S. (1997). *How the Mind Works*, New York: Norton.

Premack, D. (1985). "Gavagai!" or the future history of the animal language controversy. *Cognition*, *19*(3), 207-296.

Struhsaker, T. T. (1967). Auditory Communication among Vervet Monkeys (Cercopithecus aethiops). In S.A. Altmann (Ed.), *Social Communication Among Primates* (pp. 281-324). Chicago: Chicago University Press.

Tinbergen, N. (1951). *The Study of Instinct*. Oxford: Clarendon Press.

Trehub, S. E. (2003). The developmental origins of musicality. *Nature neuroscience*, 6(7), 669.

Wray, A. (2002). *The transition to language. Studies in the evolution of language*. Oxford: Oxford University Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

A Statistical Approach to Measure the Usage of Notes Over Different Musical Eras

Sophie Jane Andrews[1]

[1] *Chatham High School, USA*

[1]sophie.j.andrews@gmail.com

Abstract

The Yale Classical Archives Corpus (YCAC) is a repository of music compositions by over 500 western European classical composers from the 16th century to the 20th century. The corpus represents each composition in a computer-readable format. This information allows us to study the note distribution in each composer's pieces. We focus on 18 major composers defined as "Greats" by YCAC. We observe several trends and provide supporting statistical evidence. First, composers use more sharps and flats as the music eras progress from Baroque to Impressionist. Debussy (Impressionist), the latest among the 18, uses almost as many sharps and flats (45%) as natural notes, whereas Vivaldi (Baroque), the earliest among the 18, uses sharps and flats half as often (22.8%) as Debussy. Second, most composers use more sharps and flats later in their composing years. Wagner is the most pronounced example, as the compositions during his later period use extreme chromaticism extensively. Third, earlier composers use a smaller set of notes in comparison to later composers. These three observations illustrate that musical language gets richer, more complex and more sophisticated over time.

Introduction

Music notes in a composition are like letters in a piece of writing. In English, "e" is the most commonly used letter in the 26-letter alphabet. In fact, the twelve most popular letters "e t a o i n s r h l d c" are found in around 80% of English words, whereas the two least popular letters "q" and "z" each appears in 0.2% of words.

A musical alphabet is the collection of notes that can be used for a composition. For example, the 88 keys on the piano form the alphabet for piano pieces. What do we know about the musical alphabet? Is there a most popular note? Is there a subset of popular notes that make up most of the composition, and is there a subset of notes that rarely appear?

Does a composer have his personal favorite note? Do favorite notes change over a composer's composition life? Did Baroque composers use the same set of notes as the Romantic composers a century later? How does the musical alphabet evolve over time? In this work, we explore some of these questions in the context of western European classical music.

We use the Yale Classical Archives Corpus (YCAC) from Yale University [YCAC]. The archive provides a collection of compositions from western classical composers from the 16th to 20th century. Each composition is recorded in a computer-readable format, which allows for a computational and algorithmic approach in music analysis. Christopher White and Ian Quinn have a sequence of work using the YCAC corpus. The poster [WhiteQuinn 2014] shows important statistics for the YCAC including the number of composers, pieces, chords, top genres, pieces per time period etc. The paper, "The Yale-Classical Archives Corpus" [WhiteQuinn 2016], offers a detailed description of the wealth of information in the YCAC files, followed by a sample case study whose purpose is to use the YCAC data to verify a hypothesized trend in the usage of two chords. The paper, "A Corpus Sensitive Algorithm for Automated Tonal Analysis" [White 2015], creates a model that learns the chord vocabulary from a corpus like the YCAC database and introduces an algorithmic key finding scheme. The paper "Corpus-Derived Key Profiles Are Not Transpositionally Equivalent" [WhiteQuinn2017] tests the popular assumption that the frequency distributions of pitch classes in all keys are transpositionally equivalent, and finds evidence of the opposite from the YCAC database.

In this study, we count the number of occurrences of each note in a composer's pieces, and study the resulting *note distribution*. Not surprisingly, for the same composer, the note counts differ significantly from note to note. We also compare the note distribution for different composers and for compositions from different eras. The main theme of our findings is:

With the progression of music eras, composers use a larger set of notes and more sharps and flats. Most composers also use more sharps and flats towards the end of their composing years.

This illustrates that the music gets richer, more complex and more sophisticated with the progression of time.

Data

The YCAC archive consists of one metadata file and computer-readable MIDI files that store 8980 pieces/movements by 505 composers from several periods of western European music history [WhiteQuinn 2014].

Metadata. The metadata is a CSV (comma-separated version) spreadsheet in which each row corresponds to a composition, and up to 18 fields of information are associated with each piece. These fields include its title, the composer, the year(s) it was written, the instrumentation (e.g. piano, voice, orchestra, etc), the genre (e.g. symphony, etude, solo, duet, etc), and the key (e.g. G major, E minor, etc). For example, the following line contains 8 fields from a row in the metadata file:

Hungarian Dance No. 5 in F# minor | Brahms | WoO1no5 | 1858-1868 | piano | SoloDanceOneMovement | F# minor

Brahms composed the piece *Hungarian Dance No. 5 in F# minor* for piano from 1858-1868. The piece has the key of *F#* minor and the genre of solo dance one movement. Its catalog number is WoO1no5. Some fields in the metadata file may be missing. For example, none of the pieces by Byrd (b. 1540) have year of composition, which could be because Bryd was too early to have such records.

Music data. A MIDI file records notes and their time information for a composition. More precisely, each row of a MIDI file contains a *salami slice* or *slice*, which is a collection of notes. A new slice is created whenever a note enters or leaves the music [WhiteQuinn 2014]. Up to 12 fields of information are associated with each slice. In addition to the set of notes that make up a slice, we also pay attention to its *offset*, which is the starting time of the slice measured in quarter notes.

In a MIDI file, notes are represented by *C, D, E, F, G, A* and *B*. For a sharp note the "#" symbol is used, and for a flat note the "-" symbol is used. Finally, an integer from -1 to 9 is used at the end to denote the octave number. The following are the first two slices in *Brahms' Hungarian Dance No. 5*.

2 <music21.chord.Chord *C#4 F#2*>
2.5 <music21.chord.Chord *C#4 C#3 F#3 A3*>

The first slice consists of 2 notes, *C#* from the 4th octave and *F#* from the 2nd octave. The second slice consists of 4 notes, *C#* from the 4th octave, *C#* from the 3rd octave, *F#* from the 3rd octave and *A* from the 3rd octave. The first slice starts at offset 2 and stops at offset 2.5, which is when the second slice starts. Since the offset is measured in the number of quarter notes, the duration of the first slice is an eighth note.

In addition to the offset, other fields for the slice include highest pitch, lowest pitch, prime form, scale-degrees in relation to the global key, local key information, etc. However, we do not use these fields for the purpose of this study.

The Greats. In this study, we focus on the 18 "Greats" from Vivaldi (b. 1678) to Debussy (b. 1862). YCAC refers to a group of 19 composers who have a large number of compositions in the archive as the "Greats" [WhiteQuinn 2014]. Byrd (b. 1540) is a "Great" we choose not to include here since he is a chronological outlier. The 18 Greats that we study are from the following 4 eras of western classical music according to Encyclopedia Britannica 2017. For each era, we list the composers, their birth years and their composing years. The information of the latter is extracted from the metadata file.

1. *Baroque*
   Vivaldi (b. 1678, 1705-1729), Telemann (b. 1681, 1701-1768), Bach (b. 1685, 1703-1750), Handel (b. 1685, 1705-1749), Scarlatti (b. 1685, 1705-1755)

2. *Classical*
   Haydn (b.\ 1732, 1760-1801), Mozart (b. 1756, 1739-1791), Beethoven (b. 1770, 1782-1826)

3. *Romantic*
   Schubert (b. 1797, 1814-1828), Mendelssohn (b. 1809, 1824-1846), Chopin (b. 1810, 1817-1847), Schumann (b. 1810, 1831-1862), Liszt (b. 1811, 1833-1886), Wagner (b. 1813, 1832-1882), Brahms (b. 1833, 1851-1896), Saint-Saens (b. 1835, 1855-1913), Tchaikovsky (b. 1840, 1868-1893)

4. *Impressionist*
   Debussy (b. 1862, 1880-1915)

Since Debussy is the only impressionist composer and he is close in birth year and composing years to late romantic composers, we group Debussy with the Romantic era and call it Romantic+.

Findings

We write Python programs to process the metadata and MIDI files, and we visually display our results on graphs made with the Python plotting library Matplotlib.

We count the number of slices (or equivalently the number of times) a note appears in a composer's pieces. For example, *NoteCount*(*G*3, Vivaldi) counts the number of times the note *G*3 appears in the pieces composed by Vivaldi. In general, for every note in the corpus (octave -1 through 9, 12 notes per octave) and each of the 18 Greats, we use

$$NoteCount(n,c)$$

to denote the number of times note *n* appears in the compositions by composer *c*. Of the total note count by a composer *c*, we can compute the percentage that *c* uses note *n* as follows:

$$\frac{NoteCount(n,\mathrm{c})}{\sum_{\mathrm{all}\ n} NoteCount(n,\mathrm{c})}$$

In addition to counting the number of times a note appears using *NoteCount*, we also measure the total duration of a note by using the offset information from the MIDI files. We will see these two ways of counting notes have no qualitative difference in terms of the trend we have stated in the Introduction. (See Figure 3 for an example.) Therefore, we describe most of our findings in terms of note counts rather than total duration.

Observation 1: Composers use more sharps and flats as the music eras progress from Baroque to Classical to Romantic+.

We compare the note usage by Vivaldi from the Baroque era, the earliest among the 18 Greats, to the note usage by Debussy from the Impressionist (or Romantic+) era, the latest among the 18 Greats. We plot *NoteCount*(*n*, Vivaldi) and *NoteCount*(*n*,Debussy) in Figure 1. The *x*-axis is the notes from octaves 2 to 6, with 12 notes per octave

C, C#D-, D, D#E-, E, F, F#G-, G, G#A-, A, A#B- and *B*.

For notes that are enharmonically equivalent such as *C#4* and *D-4*, we treat them as the same note when counting. We also do not count double sharps and double flats. This is the same practice as in [WhiteQuinn 2017]. We do not plot the notes in lower octaves, -1 through 1, and higher octaves, 7 through 9, as the note counts in these octaves are low. We mark the count of

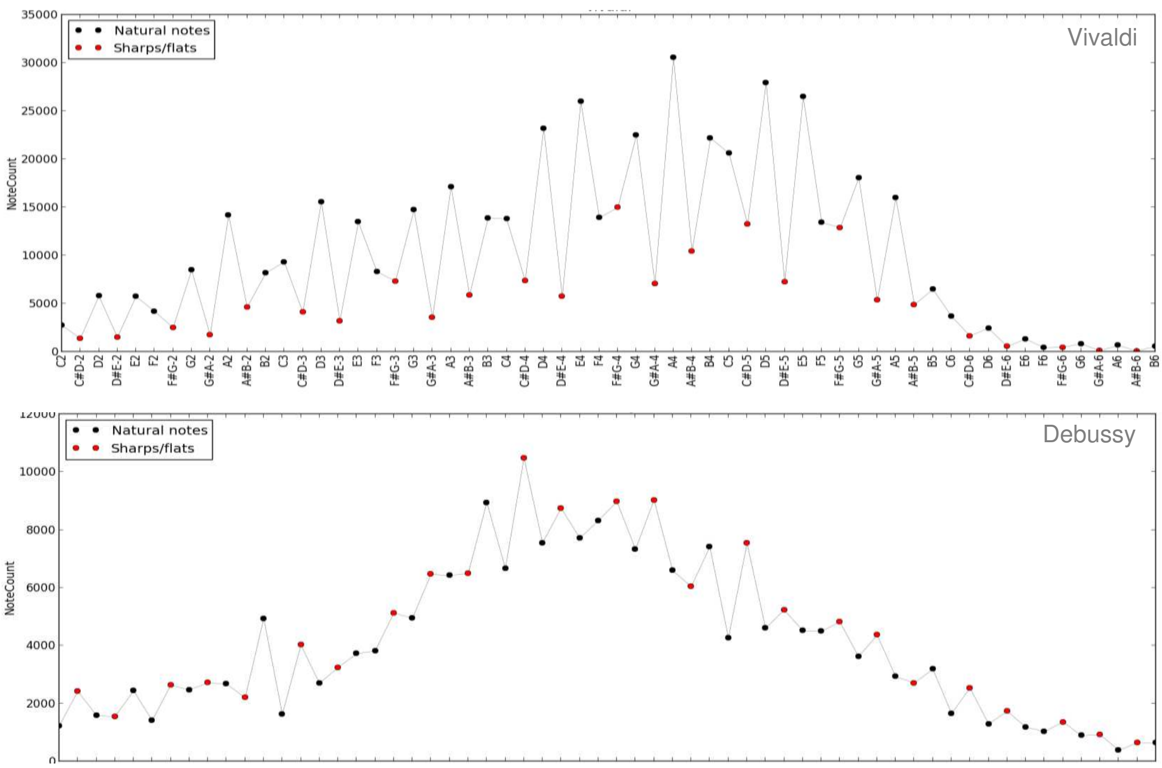


Figure 1: Note distribution of Vivaldi (b. 1678, earliest) vs distribution of Debussy (b. 1862, latest). Red dots are sharps and flats.

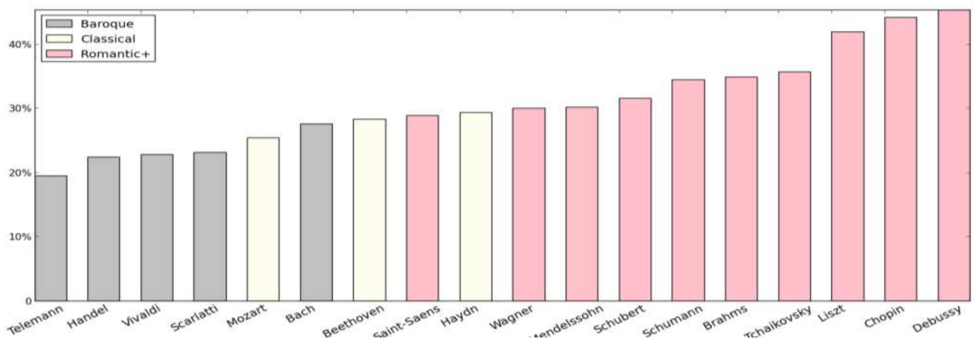


Figure 2: Percentages of sharps and flats by 18 Greats.

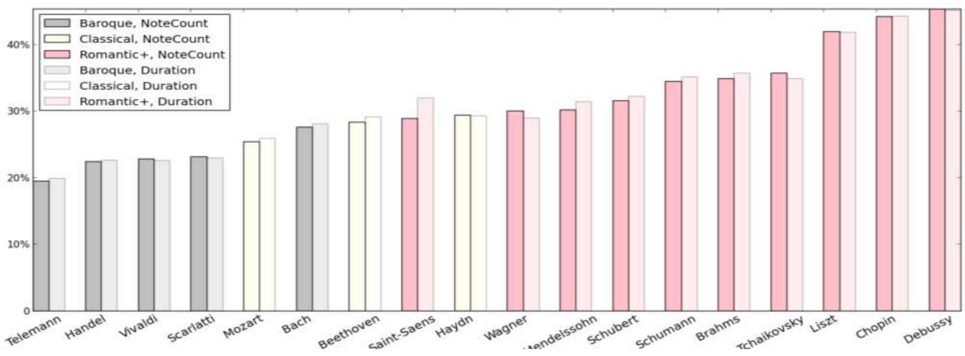


Figure 3: Percentage of sharps and flats. Dark shade: percentage based on NoteCount; Light shade: based on duration.

a sharp or flat note with a red dot. (Note that we are not counting accidentals relative to a key, but the notes corresponding to the black piano keys.) For Vivaldi, sharps and flats make up 22.8% of the total note count. On the other hand, in Debussy's compositions, almost 45% of the total note count comes from sharps and flats, which doubles the percentage of Vivaldi.

In Figure 2, we plot the percentage of sharps and flats for each composer. On the *x*-axis, we sort the composers in the order of increasing percentage. As we can see, the lowest percentage is Telemann and the highest percentage is Debussy. The Baroque composers cluster at the left end of the chart, and their sharp/flat percentages n from Telemann's 19.5% to Bach's 27.6%. Following Baroque composers are Classical composers, whose sharp/flat percentages range from Mozart's 25.4% to Haydn's 27.4%. The Romantic composers occupy the right half of the chart. Their percentages range from Saint-Saens' 28.9% to Chopin's 44.2%.

Figures 1 and 2 are statistical evidence of Observation 1. An interesting question is why these trends occur. Two likely reasons are the shift towards equal temperament and the greater use of chromaticism. Equal temperament involves spacing the half-steps in an octave equally so that all keys (including those with many sharps and flats) sound roughly in tune. Other tunings can make keys with fewer sharps/flats (e.g. C major) sound more in tune, but at the expense of much worse tuning for keys with many sharps/flats. Bach's *Well-tempered Klavier* showcases the benefits of equal temperament by including pieces in each of the 24 major and minor keys. After Bach equal temperament became more widely used which leads to greater use of keys with many sharps/flats. According to [Britannica 2017], "Equal temperament tuning was widely adopted in France and Germany by the late 18th century and in England by the 19th".

Another likely reason for the trends in Observation 1 is the greater use of Chromaticism, i.e. the greater use of accidentals. According to [Britannica 2017], "Chromaticism has been employed in rich variety as expressive and structural means. Chromatic modulation between distantly related keys, an occasional feature in the music of Johann Sebastian Bach, Joseph Haydn, and Wolfgang Amadeus Mozart, was increasingly used by early Romantic composers, including Franz Schubert and Frederic Chopin." With the progression of time, music became more chromatic with a greater use of accidentals. This in turn leads to a greater use of sharps and flats.

The paper [QuinnWhite 2017] computes pitch-class distribution (by duration) over the corpus. For pitch class, the same note/pitch from different octaves is treated the same when counting the total duration for the class. For example, notes *G*0 through *G*9 are all in the pitch class *G*. One chart in [QuinnWhite 2017] shows that, over the YCAC corpus, the

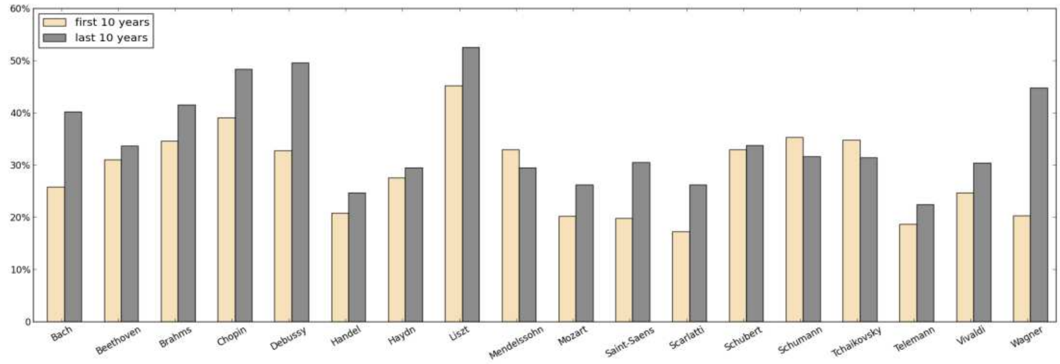


Figure 4: Percentage of sharps and flats by 18 Greats: first 10 years vs. last 10 years.

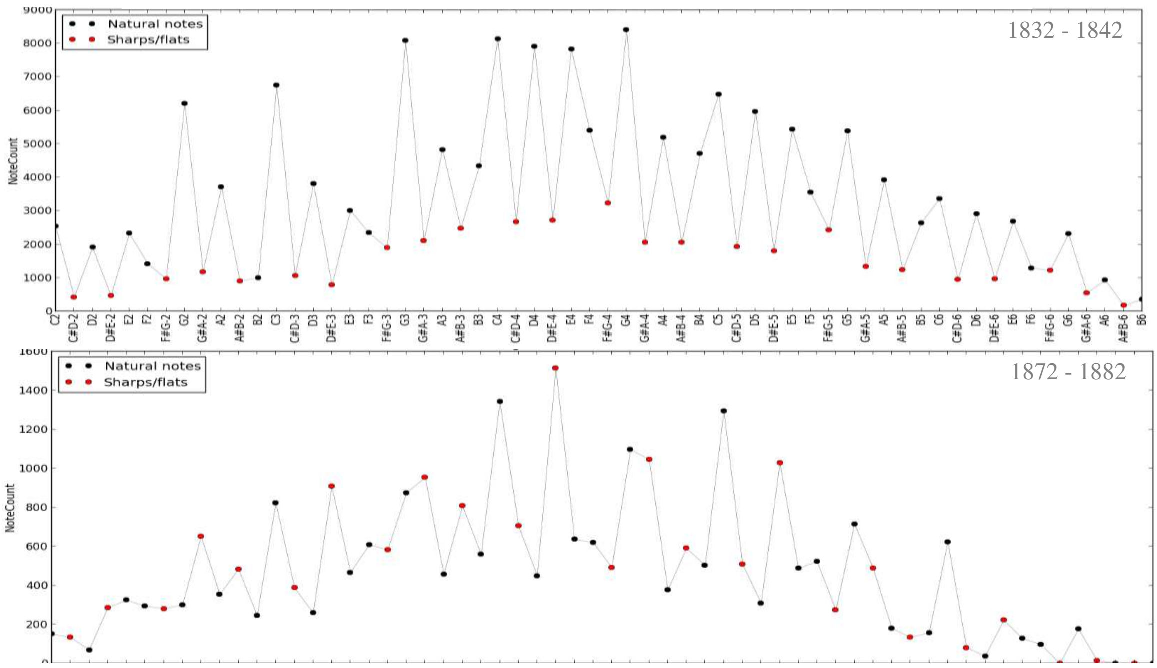


Figure 5: Note distribution by Wagner: first 10 years vs last 10 years.

total duration for each pitch class *C, D, E, F, G, A* and *B* is generally longer than for the pitch class *C#D-, D#E-, F#G-, G#A-* and *A#B-*. This is consistent with our findings for the 18 composers, although in our note distribution the same note from different octaves is treated differently. We also compute the note distribution for each composer, not for the entire corpus collectively.

In Figure 3 we compare counting by the number of note occurrences against counting by note duration. For each composer, the plot contains two bars. The one with a dark shade, which is identical to Figure 2, is the percentage of sharps and flats based on *NoteCount*; the bar with a light shade is percentage based on note duration. The two bars are close in height. Therefore, the trend of more sharps and flats with the progression of time holds regardless of how notes are counted. For the rest of the paper we only show plots based on *NoteCount*.

Observation 2: Most composers use more sharps and flats later in their composing years.

Even with the same composer, his late compositions tend to have more sharps and flats than his earlier ones. The bar chart of Figure 4 shows, for each composer, the percentage of sharps and flats that appear in compositions during his first 10 years of composing life (beige bars) and the percentages in the last 10 years (grey bars). Most composers use more sharps and flats during the last 10 composition years, with the exception of Mendelssohn, Schumann and Tchaikovsky who show a small decrease.

The most dramatic example is Wagner. As we can see, his grey bar more than doubles his beige bar in height. In Figure 5 we compare *NoteCount*(*n*,Wagner) during years 1832 through 1842 and years 1872 through 1882. Again, the red dots indicate sharps and flats. So what happened to Wagner's composition style later in life? Britannica 2017 states that he used more chromaticism. In particular, **"**Chromaticism became an outstanding aspect of the style of the dramatic composer Richard Wagner. In his opera Tristan und Isolde (1857–59) Wagner developed a continuously chromatic harmonic vocabulary in which the music frequently progressed toward new keys yet repeatedly postponed key-strengthening cadences."

Observation 3: Composers use a larger set of notes in their compositions with the progression of the music eras from Baroque to Classical and to Romantic+.

This observation is illustrated by Figure 6, which shows, among the 10 Romantic and Impressionist composers, it takes 2.2 notes on average to make up 10% of the total note count. The average number to make up 30%, 50%, 70% and 90% of the total note count are 9.1, 17.5, 28.5 and 45.9 notes respectively. These numbers are shown in the pink bars in Figure 6. We plot the corresponding numbers for the Baroque era (grey bars) and the Classical era (white bars), and observe an increasing trend in the number of notes it takes to make up a certain percentage of the total note count. This is shown by the increasing bar heights in every percentage.

Here are a few extra facts that support Observation 3.

- Scarlatti, one of the earliest composers among the 18, used the least number of notes in his compositions: 57 notes in total and 31 notes to make up 90% of his total note count.
- Saint-Saens, one of the latest among the 18, used the most number of notes in his compositions: 102 notes in total and 49 notes to make up 90% of the total note count.
- Bach used a total of 94 notes, which is significantly more than other Baroque composers: 74 notes for Vivaldi, 63 for Telemann, 68 for Handel and 57 for Scarlatti. However, Bach only used 36 notes to make up 90% of his total note count. The count of 36 is comparable to other Baroque composers. This means that Bach used many notes very few times.

Conclusion

In this study, we use the YCAC archive to count the number of appearances of each note in the 18 Great composers' pieces. We observe a clear trend that a larger set of notes are used with the progression of time. This is reflected by the usage of a larger number of the notes and more sharps and flats.

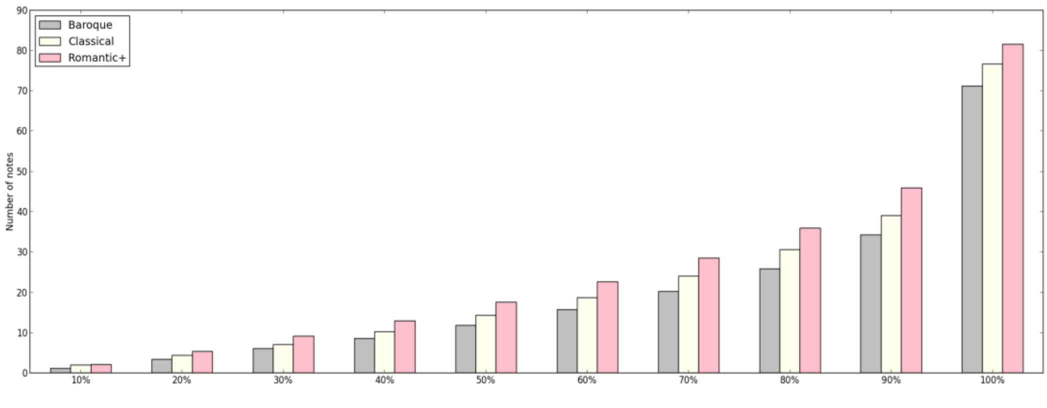


Figure 6: Number of notes that make up 10% to 100% of the total note counts by Baroque, Classical and Romantic+ composers.

Acknowledgements. I would like to thank Olivia Winn for pointing me to the YCAC database and for suggesting interesting questions to pursue in this study. I would like to thank Professor Christopher White from UMass Amherst for his encouragement and his suggestions on improving an early draft of this paper.

References

Encyclopedia Britannica (2017) *Chromaticism*. Retrieved from https://www.britannica.com/art/chromaticism.

Encyclopedia Britannica (2017) *Western music*. Retrieved from https://www.britannica.com/art/Western-music.

YCAC (2017) *Yale Classical Archives Corpus, Yale University*. Retrieved from http://ycac.yale.edu/.

Quinn, I. & White, C. W. (2017) Corpus-derived key profiles are not transpositionally equivalent. *Music Perception: An Interdisciplinary Journal*, 34(5).

White, C. W. (2015) A corpus sensitive algorithm for automated tonal analysis. *Mathematics and Computation in Music*.

White, C. W. & Quinn, I. (2014, August.) Compiling and processing the Yale Classical Archives Corpus. Poster presented at the International Conference on Music Perception and Cognition, Seoul, Korea.

White, C. W. & Quinn, I. (2016) The Yale-Classical Archives Corpus. *Empirical Musicology Review*, 11(1):50-58, http://emusicology.org/article/view/4958.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Cognition of South Indian Percussion

Jay Appaji,[1] Zachary Wallmark,[2] and W. Jay Dowling[3]

[1,2] *Southern Methodist University, USA*
[3] *University of Texas at Dallas, USA*

[1]jay.appaji@gmail.com, [2]zwallmark@smu.edu, [3]jdowling@utdallas.edu

Abstract

The *mridangam* is a double-headed pitched drum prominently featured in South Indian (Carnatic) music. Carnatic music utilizes a series of looped percussive patterns that often feature rhythmic accents between pulses. While previous studies in rhythmic memory have dealt with Western rhythms (Iversen, Repp, & Patel, 2009), few have focused on non-Western rhythms. Studies on the cognitive representation of rhythm suggest that listeners' attention is directed toward the downbeat of a rhythm, then orients other parts of the rhythm in reference to the downbeat in a "hierarchical" sequence (Fitch, 2013). We designed an exploratory study that measured listeners' ability to encode and decode rhythmic samples through two experiments. The first required listeners to determine the degree of similarity between two randomly selected rhythmic samples separated by a variable delay time and was intended to determine the extent to which memory for isolated Carnatic rhythms is modulated by delay between stimuli, timbre, and by musical congruence. The second experiment required participants to listen to a "target" rhythmic stimulus and successfully choose that same stimulus from a set of three answer choices consisting of a random ordering of the target, "similar," and "different" lure rhythms and was intended to measure the effect of stimuli ordering with musical congruence as it pertains to encoding rhythm in memory.

Introduction

Rhythm refers to the organization of events in time (Iversen, Repp, & Patel, 2009). It encompasses several factors including timbral quality, musical organization, and accenting. The encoding of rhythm in memory can be described as an instance of *perceptual learning.* Perceptual learning refers to an organism's development of perceptual sensitivity toward "important dimensions and features" of the environment (Goldstone 1998). This process contains two major steps. The first is *imprinting,* where receptors are developed for certain types of stimuli or parts of stimuli. The second step is *differentiation,* where aspects of stimuli that were previously unidentifiable are now recognizable and made available as criteria for comparison. Through this process, previously experienced stimuli can be parsed with ever greater perceptual acuity. As a result, attention can be optimized and delegated to novel elements stimuli. This process can be generalized to music cognition and perception (Kersten, 2015). Kersten suggests that music perception is an extension of worldly functions, and that everyday experience is filled with acoustic events. For example, the human auditory system adjusts to locate the source of sounds, and also looks for pitch-rhythm alignment cues to determine musical qualities.

Perceptual learning is relevant to rhythmic memory because every stimulus contains a multitude of information. Iversen, Repp & Patel (2009) revealed that the primary concern of listeners upon hearing a rhythm was finding the location of the "beat." It is important to not confuse "beat" with "pulse," as Iversen describes beat as a cognitive interpretation, whereas pulse is the the physical location of onsets which drive tempo of the music. The study tested listeners' response to a repeated isochronous (i.e., even subdivision of beats) tone pattern that contained physical accents on different beats. Confirming Repp (2005), they reported three main findings: first, time-location of the beat receives the most attention, which increases its position in the mental hierarchy. As a result an auditory "image" of a physical accent can be formed, which allows listeners to envision the beat without a physical accent being present. Second, they note the presence of "covert motor rhythm," a simulation of the beat without the physical accent being present. Finally, they discuss a "hybrid model" of rhythm perception which involves generating an internal rhythm and creating an auditory image to help keep track of the beat.

The findings of Iversen, Repp & Patel (2009) are especially important in establishing a baseline perceptual schema for rhythm. South Indian percussion employs a variety of meters and subdivisions; however, in this study we chose stimuli which were set in an eight-beat meter, known as *adi tala.* This is in contrast to similar studies conducted in the realm of cross-cultural rhythm cognition, which emphasized the metrical organization of rhythm across participant groups (Hannon & Trehub, 2005). Although we used isochronous rhythms in this study, most rhythmic stimuli contained salient accented phrases that occurred between beats or offset within the beat itself. This is the main contrast between Carnatic and Western classical rhythmic stimuli.

The *mridangam* (Fig. 1) is a two-sided pitched hand drum. The drum consists of two skin membranes stretched over the barrel. On one side of the drum is a lower-tension head, known as the *thoppi,* consisting of a combination of cow, buffalo, and goat skin, intended for creating tenor and bass tones. The mridangam player plays this side with his or her non-dominant hand. On the other side of the drum is the higher-tension head, known as the *vallanthallai,* consisting of a similar combination of skins. This head, however, contains a patch of iron oxide rice powder which creates the pitched element of the drum. Each drum is tuned to a note of the scale. The mridangam is played by creating different hand positions and striking the heads in different locations. Both heads are played together to generate the looped rhythmic patterns that characterize its sound. The mridangam is typically used as an accompanying instrument with a Carnatic vocalist, who sings at the same pitch to which the drum is tuned.

Figure 1: First author playing the *mridangam*

The mridangam's construction enables timbral flexibility, which is another influential factor driving perceptual learning and memory encoding. Alluri (2012) describes timbre as essential to sound categorization and mental representation of our surroundings: it is thus the element of music most important to human survival. Timbre is relevant to cross-cultural rhythmic perception and cognition because there exists such a wide multitude of rhythmic mediums in the world. Alluri discusses *timbral emergence*, whereby familiar timbres merge to form a unique new timbre. Listeners hearing culturally unfamiliar musical styles will in theory be more equipped to encode such a stimulus into memory.

Fitch (2013) describes rhythms as possessing a hierarchical, "tree-like" structure in which the downbeat serves as the "root" and other accented beats and salient features form the "branches," events that occur in relation to the downbeat. Fitch's theory is angled towards explaining rhythmic hierarchy from a metrical perspective, in which a listener hears a salient, recurring moment in the music as a downbeat while viewing everything else as a function of the downbeat. Although this pertains mainly to the metrical structure of the rhythm, perhaps it can be extended to the pitch differences heard in the different strokes of a mridangam pattern. Could such differences and the pitch counter of the pattern serve as a hierarchy in itself?

Monahan, Kendall, & Carterette (1987) suggested that grouped pitches are more readily encoded in memory. Cooper and Meyer (1960) found that accenting is the basis for grouping. Given the stark differences in pitch and timbre between strokes on the mridangam, perhaps an effective coding schema might be to identify pitch-based accents in a repeated pattern. Mridangam patterns typically mark downbeats and other "strong" beats with a stroke that combines both a bass and pitched stroke. Intermediate beats are typically filled with higher frequency strokes played on the *vallanthallai*. Despite the fact that maintaining the beat using finger counts and hand claps is a core component of Indian musical performance, many musicians (vocalists in particular) rely on the described rhythmic timbre structure to keep time. Because Carnatic music is highly improvisational, this timekeeping technique is especially helpful to prevent rhythm errors on the part of the vocalist. While such a schema is obtained naturally by experienced practitioners of the musical style, it remains to be seen if those who lack exposure to Carnatic music can use similar timbre-based cues to develop a rhythmic hierarchy in memory. The effect of timbral cues can also be measured outside of a cross-cultural context, by comparing retention of stimuli when produced on different mediums.

After discussing the roles of beat perception and timbral quality on rhythm encoding, it is important to acknowledge their cultural specificity. Hannon and Trehub (2005) measured the response of infants, 12-month olds, and adults to Balkan rhythms. The rhythmic stimuli were in both isochronous and non-isochronous meters. The 12-month old and adult groups were exposed to Balkan folk-dance melodies prior to the study. The authors found that 12-month olds were more adept at detecting rhythmic changes in unfamiliar stimuli than infants were. These results speak strongly to the effect of enculturation on cognition of unfamiliar rhythm. Moreover, adults performed more accurately on the isochronous condition of the experiment. In fact, participants' mean accuracy on the non-isochronous condition was less than chance. The authors attribute this effect to the fact that adult participants were likely assimilating the non-isochronous rhythms into the framework of an isochronous meter. The findings in this portion of the study indicate that prolonged exposure to one musical style can serve as a detractor when assessing unfamiliar musical styles. While this study aimed to measure the effect of enculturation, however, it is difficult to determine the extent to which familiarity serves as a memory aid or detractor.

Our study explores the effect of timbral quality, musical congruence, and delay time on the encoding schema of culturally unfamiliar rhythms. We designed two experiments to measure these effects. George and Coch (2011) indicated that musical training can result in enhanced short-term (working) memory. In Experiment 1, participants possessed only Western classical music experience, and in Experiment 2, participants possessed either Western classical or Carnatic music experience, and were group accordingly. Experiment 1 investigates the role of musical congruence, delay time, and timbral quality on recognition accuracy; Experiment 2 measures recognition accuracy as affected by musical congruence and recency.

Experiment 1

Participants (N = 48, M years musical training = 6, ages 18–22) were presented with two listening tests consisting of 36 trials of rhythmic stimuli. One test consisted of natural rhythmic stimuli recorded on the mridangam and the other consisted of mechanical rhythmic stimuli identical in structure to the natural set, but synthesized.

A total of 81 Carnatic rhythmic stimuli (8 s) were recorded by the first author on the mridangam. An identical set of 81 mechanical rhythmic stimuli were generated on a woodblock MIDI sound using Cakewalk Audio software. Each rhythm was set in the eight beat cycle, *adi tala*. Stimuli were divided into three categories: (1) *target,* a reference rhythm separated from two lure types, (2) *similar,* rhythms designed to have a high level of musical congruence to the respective target, and (3) *different*, rhythms designed to have a low degree of musical congruence to the respective target. The similar rhythms were created using musical techniques such as embellishing strokes in the target and replacing silence with sound, among others. The different rhythms were created to be very dissimilar to the target. Figure 2 illustrates the difference between these three stimuli categories.

Participants were randomly assigned an ordering of the two tests (Natural-Mechanical or Mechanical-Natural). Each test consisted of 36 trials. Each trial contained a pairing of a target rhythm with either a similar or different lure. The two rhythms were separated by a variable delay time of either 3 seconds, 6 seconds, or 12 seconds. Listeners were instructed to rate the similarity of the two rhythms on a 4 point Likert-scale, where 1 corresponded to "sure same" and 4 corresponded to "sure different." The same procedure was used for both natural and mechanical trials.

Results: A repeated-measures ANOVA (2x3x2: 2 trial types (natural, mechanical), 3 delay times (3s, 6s, 12s), and 2 lures (similar, different)) was calculated. Figure 3 illustrates the results from Experiment 1: there was no difference between natural and mechanical excerpts, $F(1,47) = 0$, $p = .99$, suggesting that timbral quality is not involved in recognition accuracy. However, there was a significant interaction between trial type and lure $F(1,47) = 15.55$, $p < .001$, $\eta_p^2 = .25$. In the mechanical condition, when a target was paired with a different lure, participants were more accurate, whereas in the natural condition, participants were more accurate during trial pairings with a similar lure. delay time was also found to be significant, $F(2,94) = 6.47$, $p < .01$, $\eta_p^2 = .12$. Accuracy declined as delay time increased, confirming established findings that elapsed time detracts from memory recall (Dosher, 1999). Results from Experiment 1 revealed that the most favorable combination for recognition accuracy was when a target rhythm was paired with both low-congruence (different) lures and a short delay time. Additionally, post-hoc analysis (Bonferroni) revealed a significant difference between the shortest delay time (3s) and longest delay (12s), $p < .001$.

Experiment 2

Participants ($N = 34$, ages 18–30) were presented with a listening test consisting of 20 trials of mridangam patterns. Participants consisted of two independent groups: *naïve* (no previous exposure to Carnatic music; $N = 24$, M years training in western music = 9.7 years), and *expert* (members of the Dallas Carnatic music community, $N = 10$, M years training in Carnatic music = 10.7). The listening test included a memory task and recorded participant confidence after each trial.

A total of 60 Carnatic rhythmic stimuli (4 s) were recorded on the mridangam. Stimuli were shortened versions of a subset of recordings also used in Experiment 1. These rhythms were then equally assigned (20 each) into the three conditions (target, similar, different) as they were in Experiment 1.

Twenty listening trials were administered. Each trial included a target rhythm followed by a random ordering of three answer choices comprising the *target* itself along with *similar* and *different* lures. After listening to all three stimuli, participants were instructed to choose the position (1, 2, or 3) of the target rhythm. Following each trial, participants rated their confidence in their selection on a 6-point Likert scale where 1 corresponded to "not confident at all" and 6 corresponded to "very confident." At the conclusion of the listening tests, participants rated their attentiveness on a 6-point scale of increasing attentiveness.

Results: A multi-level logistic regression model was constructed to predict accuracy (binary) from confidence, experience, and ordering variables. To account for repeated measurements, "participants" were modeled as random

Figure 2: Example set of target (T), similar lure (S), and different lure (D) Carnatic rhythms

effects. Results of fixed effects for naïve participants are displayed in Table 1, expert participants in Table 2. For both participant groups, confidence was a statistically significant predictor of accuracy. In both groups, different and similar lure types in position 1 and 2 were significant negative predictors of accuracy. While this was expected, different and similar lure types did not differ in predictive strength. As shown in Figure 4, while naïve participants underestimated

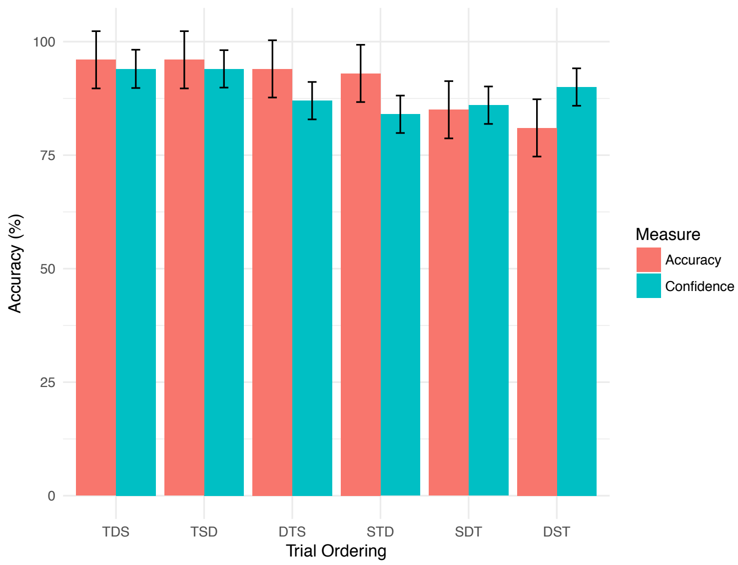


Figure 4: Mean accuracy and confidence by trial ordering for naïve group

accuracy in the four highest performing trial orderings, they overestimated their performance in the trial orderings

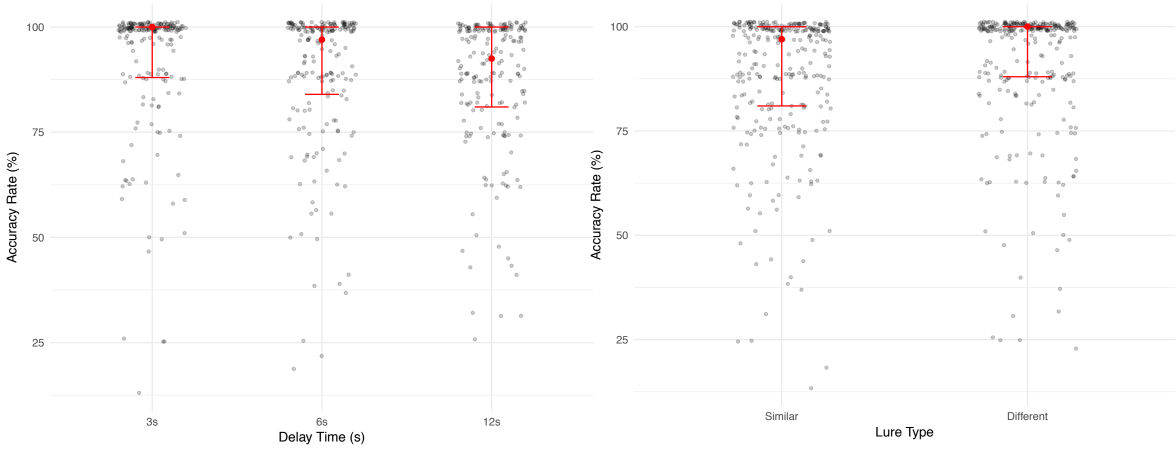


Figure 3: Median accuracy rates by Delay time (left); Median accuracy rates by Lure Type (right) (Errorbars = IQR)

(Similar-Different-Target (SDT), Different-Similar-Target (DST)) with the lowest accuracy. By way of contrast, for the expert group the mean accuracy rates was over 99%, with confidence ratings accurately reflecting this strong performance. Therefore, a ceiling effect was observed in this group. Of the 200 responses to the memory question made by members of the expert group, there were only two incorrect responses.

Conclusion

Results from both experiments converge in highlighting the role of recency in encoding of short-term memories for rhythmic patterns: listeners' recognition accuracy was highest where serial position of the correct answer was closest to the test stimulus either in delay time (Exp. 1) or serial position (Exp. 2). Both experiments also suggest that in addition to delay time, recognition accuracy is closely related to the order in which samples of varying musical congruence are presented.

In Experiment 1, timbral quality did not appear to have an effect on recognition accuracy. This indicates that listeners' mental hierarchy (Fitch, 2013) may be based upon rhythmic contour alone (Monahan, Kendall, & Carterette, 1987). Perhaps a more robust way to determine the extent of the usage of rhythmic contour in encoding would be to eliminate the tonal differences in the mechanical set of a Carnatic rhythms while still preserving the musical structure.

Experiment 2 indicated that serial position serves as a form of distracted delay time between samples, and is congruent with the expectation that increased delay time would decrease recognition accuracy. The results of Experiment 2 exhibited a strong ceiling effect in the case of the expert group. In the case of the naïve group, unfamiliarity with the musical tradition combined with the isochronous meter of the stimuli could have served as a strong memory enhancer for listeners. This could be attributed to the fact that intended distractors (high-congruence lure types) are not perceived as such, hence not serving their intended purpose. This runs contrary to the findings of Hannon and Trehub (2005), who found that those without exposure to an unfamiliar musical tradition were at a disadvantage compared to those who had exposure. This effect can be seen when put into the context of the expert group, whose results were essentially 100% accurate, confirming the advantageous influence of enculturation on rhythmic memory encoding.

Administering the listening test on a true naïve listener set may provide some insight into the extent of both non-specific and specialized musical experience and enculturation on recognition accuracy. It is possible that salient perceptual elements found in isochronous Carnatic rhythm have already been established in the mental hierarchy of listeners who possess substantial training in Western classical music. Another topic for further study would be to

Table 1: Summary of fixed effects from multi-level logistic regression for naïve group

Predictor	*b*	*SE*	*Z*	*p*
Confidence	0.94	0.16	6.00	< .0001
D in Position 1	-1.52	0.58	-2.62	0.01
S in Position 1	-1.37	0.62	-2.20	0.03
D in Position 2	-1.35	0.56	-2.41	0.02
S in Position 2	-1.49	0.58	-2.55	0.01

Table 2: Summary of fixed effects from multi-level logistic regression for expert group

Predictor	*b*	*SE*	*Z*	*p*
Confidence	0.90	0.14	6.34	< .0001
D in Position 1	-1.27	0.54	-2.35	0.01
S in Position 1	-1.27	0.60	-2.12	0.03
D in Position 2	-1.14	0.53	-2.14	0.03
S in Position 2	-1.51	0.56	-2.67	0.01

recreate the listening test using non-isochronous rhythms, as meters of fives and sevens are common in Carnatic music. Isochronous meters also proved easier for listeners to count (Hannon & Trehub, 2005). This would treat metrical structure as a variable, and could be useful in determining the extent to which meter is an element of the mental hierarchy.

References

Alluri, V. (2012). *Acoustic, neural, and perceptual correlates of polyphonic timbre.* PhD diss, University of Jyväskylä.

Cooper, G. W., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.

Dosher, B. A. (1999). Item interference and time delays in working memory: Immediate serial recall. *International Journal of Psychology, 34*(5-6), 276–284.

Fitch, W. T. (2013). Rhythmic cognition in humans and animals: Distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, *7*(68). doi:10.3389/fnsys2013.00068

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia, 49*(5), 1083–1094.

Goldstone, R. L. (1998). Perceptual Learning. *Annual Reviews of Psychology,49*, 585-612. Retrieved May 22, 2018, from http://invibe.net/biblio_database_dyva/woda/data/att/9a2b.file.pdf

Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12639–12643.

Iversen, J. R., Repp, B. H., & Patel, A. D. (2009). Top-down control of rhythm perception modulates early auditory responses. *Annals of the New York Academy of Sciences*, *1169*(1), 58–73.

Kersten, L. (2015). Music and cognitive extension. *Empirical Musicology Review, 9*(3–4). doi: http://emusicology.org/article view/4315/4188

Monahan, C. B., Kendall, R. A., & Carterette, E. C. (1987). The effect of melodic and temporal contour on recognition memory for pitch change. *Attention, Perception, & Psychophysics*, *41*(6), 576–600.

Repp, B. H. (2005). Rate limits of on-beat and off-beat tapping with simple auditory rhythms: 2. The roles of different kinds of accent. *Music Perception, 23*(2), 165–188.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Role of Structural Tones in Establishing Mode in Renaissance Two-part Counterpoint

Claire Arthur,[1] Peter Schubert[1], and Julie Cumming[1]

[1,] *Schulich School of Music, McGill University, Canada*

[1]claire.arthur@mcgill.ca, [2]julie.cumming@mcgill.ca, [3]peter.schubert@mcgill.ca

Abstract

This project investigates mode via a corpus study of two-part counterpoint. Music theorists have often claimed that it is not simply the finalis (i.e., final note) that determines mode, but that "structural tones" within a melody are largely responsible for establishing the mode. Specifically, these theorists have pointed to notes forming "turning points" (outlines) and, to a lesser extent, leaps, as key "ingredients" for the careful construction of a melody in a given mode. However, to the authors' knowledge no empirical investigation has been made of this claim. Taking as an assumption that endpoints of melodic leaps and outlines will carry greater structural significance, we hypothesize that these structural features alone will be predictive of a piece's modal label. In addition, we hypothesize a correlation between pitches creating perfect vertical intervals and modal labels (Schubert, 1993). To carry out this analysis, we assembled a corpus of 44 Renaissance duos, with the size of the corpus being restricted to the limited set of data with reliable ground-truth data. We use multinomial regression modelling as well as behavioural experiments to evaluate the predictive power of leaps, outlines, and vertical intervals to predict mode. While our results are tentative due to our small sample size, we propose that the authentic/plagal distinction historically linked to monophonic chant is not clearly marked in polyphonic music. Vertical intervals appear the best predictor of mode compared with melodic leaps and outlines, however, simple pitch class distributions predict mode just as well while using a simpler model.

Introduction

The nature of mode in polyphonic music is as hotly debated by theorists now as it was in the sixteenth century (Powers and Wiering 2017, Powers 1981, Wiering 2001, Judd 2000). Is mode a classification system applied to works post-compositionally, or did composers deliberately "compose in the modes"? And if they did, what did that mean? Which musical features best create a sense of mode? And, does the plagal-authentic distinction (e.g., dorian versus hypodorian) apply to polyphonic works, given this distinction is traditionally associated with the melodic range of a single voice?

Music theorists have often claimed that it is not solely the *finalis* (i.e., final note) that determines mode, but that "turning points" (outlines) and leaps within a melody are largely responsible for establishing the mode by highlighting the boundary pitches of the principal interval species (e.g., Marchetto, Tinctoris, Aron, Zarlino; see Powers and Wiering 2017, II.4.ii-iii, III.2.i). While most discussion of mode has focused on the melodic behaviour of the voices, some Renaissance theorists suggested that the perfect vertical intervals formed between parts in a polyphonic texture may have modal significance (e.g., Ramis de Pareia; see Figure 2). If the theorists are correct, then it should be possible to predict the mode from the frequency distribution of notes forming leaps, outlines, or perfect vertical intervals, even without knowledge of the final and range of the individual voices. Accordingly, we present a corpus analysis of a set of 44 Renaissance contrapuntal duos with the aim of testing the theoretical assumption that melodic leaps, outlines, and perfect vertical intervals will be used in a way that highlights the principal interval species of a mode.

We evaluated each of the above features (melodic leaps, outlines, and vertical intervals) for their potential to predict the mode of a piece using both statistical modelling as well as behavioural experiments with early music experts. To anticipate our results, the first finding is that the distinction between authentic and plagal modes is minimal in these polyphonic duos; by extension, this finding may apply to polyphony more generally. The second finding is that melodic information is less useful for predicting mode than vertical interval information. This is especially surprising, since most discussion of mode in the Renaissance focuses on melodic behaviour.

Background

Modal Significance of Leaps and Outlines

Modes are described by medieval and Renaissance theorists as principally depending on their ranges and their finals. They are grouped in pairs, each pair (which we call "mode family") having one of four finals, D, E, F, or G. The first pair of modes, for instance, share the same final, D, but the range of the first mode is from D to D, the range of the second from A to A. Modes whose range is from final to the note an octave above are called authentic; those that go from the fifth degree below to the fifth degree an octave above are called plagal. These criteria were well established by the beginning of the eleventh century (Cohen 2002). However, the modal tradition is a monophonic one, originating in Gregorian chant.

Even dating back to the monophonic tradition, however, there are theorists who describe other ways that composers may create a sense of mode through melodic emphasis. Marchetto of Padua (ca. 1317), for example, illustrates the importance of highlighting "important" notes (i.e., the bounding notes of a species; see Figure 1) through melodic outlines and leaps. Marchetto's examples are monophonic, but his theories were later applied to the individual lines of polyphonic pieces in the Renaissance. Followers in this tradition include Tinctoris (1476), Aron (1525), Glarean (1547), Zarlino (1558), and Aiguino (1581).

Most of the skips and outlines in Figure 1 are between D and A, but there is one outline between G3 and D4 that contradicts the basic principle. Most melodies do not strictly adhere to the principle, and do in fact contain emphases on notes other than the two mode-defining ones. Our study aims to find out to what extent composers adhered to "correct" modal behaviour. Since we are focusing here on polyphonic works, we have chosen to ignore the traditional feature of range—which is complicated by the presence of multiple lines—in favor of the behaviour of the melodic line, specifically focusing on skips and outlines that occur during the course of a melody.

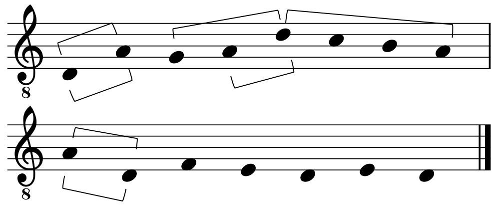


Figure 1. The first mode according to Marchetto (ca.1317). Square brackets added above and below indicate melodic outlines and leaps, respectively.

The Role of Vertical Intervals

In addition to melodic leaps and outlines, certain vertical intervals may also carry significance for the interpretation of mode. Ramis de Pareia (1482) comments on the significance of perfect vertical intervals when improvising a line against a plainchant. He shows two examples of note-against-note counterpoint above a Dorian cantus firmus. He comments that the second one exhibits the "organization" of the Dorian mode (Ramis 1482, 128). We can see in Figure 2 that there are three vertical D octaves and two fifths below D. Thus the added (upper) line emphasizes D4 melodically, but also creates five perfect vertical intervals involving D.

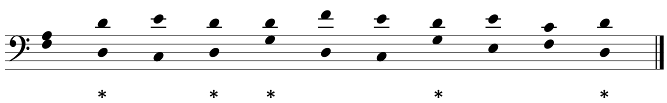


Figure 2. "Dorian organization" of the upper line according to Ramis de Pareia.

Tinctoris, too, discusses the role of perfect vertical intervals for highlighting a mode (Tinctoris 1444, III.5). Some scholars (e.g., Alexis Luko) believe that Tinctoris' reference to the use of perfect vertical intervals applies merely to those involved in cadential activity, however, his example does not make this clear. Here we follow Ramis, whose perfect intervals can occur anywhere in the phrase. Thus, as we explain in the following section, in our study we count all perfect vertical intervals as (potentially) modally significant, regardless of whether they occur at cadences or not.

Methodology

The Corpus

In order to examine a claim made by music theorists about mode, we first needed to establish a corpus of pieces of music in which the modal identification was not debatable, and that have overall similarities of form and structure. Accordingly, the pieces in our corpus come directly from treatises on mode by Glarean (1547), Zarlino (1558) and Pontio (1588) wherein the theorists ascribed each piece an explicit modal label. In addition, we add to our corpus a didactic modally labelled collection of pieces by Orlando di Lasso (1577) with a clear didactic purpose (Schubert 1995, Powers 1981). The result is a corpus of 44 pieces, all for two voices. We chose to focus on duos because several theorists include what could be called "modal cycles" of duos in their treatises. Zarlino and Pontio composed their duos as illustrations of the modes for their treatises, and Lasso composed his duos and arranged them in modal order.

Zarlino and Glarean advocated a new system of twelve modes, while Lasso and Pontio used the traditional eight modes. In order to evaluate the modal language across these collections, we therefore excluded the examples from Glarean and Zarlino for modes 9 to 12. In some cases, the duos were presented in the transposed version of the mode (such as duos with a G final and a Bb, or "G-Dorian). We transposed those pieces back to the traditional final of the mode for our dataset, so that they would be more comparable with other pieces in the same mode. Each piece was transcribed in the original note values from facsimiles of the original sources using notation software, and then exported as a musicXML file for use in our computational analysis.

Unfortunately, with the exception of Zarlino, the theorists represented here did not always use an equal number of duos to illustrate each mode in their treatises (see Appendix). For instance, neither Pontio nor Glarean provides duos for mode 4. Likewise, in the Lasso collection, there is only one duo to represent each of modes 3, 4, and 5 (while there are two duos in each of the remaining modes). In addition, the length of each piece in our corpus is quite varied, ranging from eight measures (Pontio, mode 2) to 72 measures (Glarean, mode 2: Vaqueras). The larger numbers of duos in modes 1 and 2, and the paucity of mode 4 duos, corresponds to our sense of the representation of these modes in the larger repertoire of sixteenth-century polyphony (see Table 1).

Table 1. Summary of the number of pieces (duos) in the corpus according to their ground truth modal labels.

Mode	1	2	3	4	5	6	7	8	Total
Number of Duos	7	8	5	2	5	6	5	6	**44**

The Analysis

The pitches, and size and direction (ascending or descending) of a leap or melodic outline may be paramount in establishing or contributing to the sense of mode. It is assumed that the endpoints of leaps and outlines function as "structural tones" in that they are salient features of a musical surface that might draw attention to a particular note and therefore carry greater structural significance. In addition, as explained in the Background section above, we have reason to believe that notes forming perfect vertical intervals may similarly function as structural tones. Thus, using the symbolic music analysis toolkits music21 and VIS, we sought to reduce each piece to a set of tables that carried information about the pitch content of leaps outlines, and perfect vertical intervals. Our hypothesis is that the frequency distributions of these notes can be used to predict the mode of a piece. As discussed further in the Results section, we were unsure whether a "leap" should be defined as a minor third or greater, or a perfect fourth or greater. As such, we tested the accuracy of our predictions based on both definitions. For the sake of clarity, we define an outline to contain a string of notes moving in the same direction, and the notes described as "forming the outline" are those notes where the melody changes direction (i.e., turning points).

We use two methods to evaluate our hypotheses: multinomial regression modelling, and behavioural experiments. While statistical regression is a useful tool, we were curious about the model's performance in classifying the modes compared to that of early music experts. In particular, we felt that the experts would be able to use their existing knowledge of mode in the Renaissance to an advantage. Given the small size of the corpus, the model will have a limited set of training data, while our experts have had the equivalent of a lifetime of training data. It is important to note that although our experts helped to curate the corpus, they were not familiar with most of the pieces, and it was assumed that identifying a piece based on the limited information in the tables would be highly unlikely. We initially carried out a total of two regression analyses and three experiments.

Study 1: Melodic Considerations

Recall that we wanted to know whether the usage of leaps and outlines alone might suffice to deduce the mode of a piece. Accordingly, our first analysis consisted of independently tabulating the ending pitch class of any leap or outline in either the upper or lower voice of each duo. This was important because we wanted to evaluate the predictive power of leaps and outlines independently. This resulted in total of forty-four tables, one for each piece in the corpus. An example of such a table is shown in Table 2. Notice that all information about the size, direction, or register (i.e., octave placement) of the leaps and outlines has been eliminated. These tables were evaluated by both our human experts and a regression model.

Table 2. Example table of leaps and outlines used in Study 1. The table counts instances of pitch classes forming either the end point of a leap or an outline in either the upper or lower voice throughout a single piece in the corpus.

		A	B	C	D	E	F	G
Upper Voice	Leap Endpoint	8		3	9	4	5	3
	Outline	8		5	6	5	1	3
Lower Voice	Leap Endpoint	6	3	6	11	4	4	3
	Outline	8	3	2	6	9	2	4
	Total	30	6	16	32	22	12	13

Early music experts Schubert and Cumming were given the complete set of tables with the tallies of pitch classes forming the ends of leaps and outlines (see Table 2) for each piece from the corpus. The order of the tables was randomized, and the experts were challenged to guess the mode of each piece based on this limited amount of information.

The same set of tables were used in various multinomial regression models. Separate regression models were trained with each of the five possible rows of the table (i.e., leap counts from upper or lower voice, outline counts from upper or lower voice, and totals) to see which set of data might best predict the mode. Counts for each pitch class were used as predictor variables. Rather than conduct a stepwise regression, which would iteratively test all possible combinations of parameters (i.e., pitch classes), we chose to limit the possible interactions, capping at all three-way interactions. This was primarily done to avoid overfitting the models to our small dataset. In addition, we tested several models with interactions based on theoretical predictions. For instance, instead of only calculating all pairwise interactions (which in this case would be every possible pair of pitch classes), we defined a set of interactions based on pitches related by thirds or fifths. These provide a theoretically-driven alternative to the all-pairwise-interactions model while dramatically simplifying the model (i.e., there are only 7 possible pairs of fifths and 9 possible pairs of thirds, while there are 28 possible pairs of pitch classes.) If these theory-driven parameters proved to be highly predictive, then it would provide some insight as to whether third or fifth relations are the more important indicators of a mode.

To evaluate the models, we conducted a standard "train-and-test" procedure. In the training phase, the models are fed a portion of the corpus along with the "ground truth" (the theorists' modal labels) for learning the relationships among the various parameters. In the testing phase, the model makes a "best guess" about the true mode of each of the remaining pieces in the corpus, based on the relations it observed in the training phase. In this case, since we do not have an equal number of pieces in each mode, each test set consists of one piece randomly selected from each mode (or mode family) and each training set is the remainder of the pieces in the corpus. (Sometimes referred to in MIR literature as the "leave one out" procedure.) This train-and-test procedure is repeated one hundred times, and the model estimates are taken from the averaged performance.

Study 2: Additional Melodic Information

While the success or failure of both the regression models and the behavioural experiment provide novel insights about the amount or type of information required to execute a modal assignment, it cannot directly test the claims raised by theorists on the use of leaps and outlines to highlight the mode of a piece. Recall that Marchetto points out the role of the pitches, species, size, and direction of the interval, as well as the range in which the leap or outline occurs. Since direction, size, species, and range are missing from our first set of tables (Study 1), we sought to evaluate the predictive power of a set of tables that would retain this information. Our hypothesis was that if this additional information aligned with the theoretical predictions, it would dramatically improve our experts' modal predictions. Therefore, for each piece in the corpus we created a table as shown in Table 3.

In this second study, two tables were created for each piece—one for leaps, and another for outlines—in order to retain the independent contributions of leap and outline information. Table 3 provides an example of the "leap table" for one piece. Unlike in Study 1, here the tallies from both voices are collapsed into a single table. However, since this set of tables contains information about pitch height, some information about the contributions of each voice can be derived.

Table 3. Example table of leap counts used in Study 2. Leap origins are in the left (vertical) column and the destination is in the top (horizontal) row. Numbers represent counts of leaps (defined as a third or greater) throughout a single piece in our corpus. Xs along the diagonal highlight a directional boundary: counts appearing right of the diagonal represent ascending leaps, left of the diagonal represent descending leaps. The farther away—horizontally—an instance is from the diagonal, the larger the leap.

To From	C3	D	E	F	G	A	B	C4	D	E	F	G	A
C3	X		1	1									
D3		X		2	2	2			1				
E3			X		3	3		1					
F3		5		X							1		
G3		1	6		X			3					
A3				3		X		4	4				
B3							X		1	1			
C4						6		X		1			
D4						1	2		X		2	1	2
E4						1		1		X			
F4									3		X		
G4										3		X	
A4									1	1			X
Total		6	7	6	5	13	2	9	10	6	3	1	2

In this second experiment our experts were given this new set of (paired—leaps and outlines) tables for each piece, presented in a different random order, and asked to guess the mode for each one. In this second study, it was unfeasible to train regression models using this new set of tables representing so few data points (44 pieces), especially since they contain far too many cells (most of them zero).

Study 3: Perfect Vertical Intervals

In addition to melodic "structural tones," we also hypothesized a correlation between the pitches forming consonant perfect vertical intervals and a piece's modal label. That is, we hypothesized that pitch classes of greater structural significance to the mode were more likely to be emphasized with perfect unisons, fifths, and octaves. We chose to analyze the predictive power of the vertical intervals separate from the melodic structural tones so as to determine whether the melodic features or the vertical intervals in isolation provide sufficient information (either to a regression model or to our experiment participants) to determine the mode. As in Study 1, tables were created that would be evaluated both by statistical regression as well as behavioural experiment. Accordingly, our final set of tables carried the information as shown in Table 4.

As in Study 1, the regression models independently test each of row of each table (in this case, unisons, fifths, octaves, or totals) to see which set of data best predicts the mode. The parameters (including theory-driven interactions) and overall procedure for running and evaluating the models were otherwise identical to that of Study 1.

Table 4. Perfect interval counts used in Study 3. This table shows the number of perfect intervals between the two voices in a single piece. Since vertical intervals are measured from the bottom to the top voice, here we count the pitch classes forming the bottom note of any perfect interval (e.g., in this piece there is a single occurrence of a perfect fifth built on A).

Vertical Intervals	A	B	C	D	E	F	G
P1	1						1
P5	1		4	8	5	3	12
P8	2	1	3	5	3	2	2
Total	4	1	7	13	8	5	15

Results

Study 1

A multinomial regression and behavioural experiment were carried out to investigate whether simple tallies of pitch classes representing the end points of leaps and outlines could be used to predict the mode of a piece.

Since there are eight possible modes, chance accuracy would be approximately 1 in 8 or 12.5%. Recall the eight modes form four families, where each mode family shares a final. For mode family, chance accuracy would be approximately 25%. Although the experts had difficulties assigning a single mode to each table, they successfully labeled the piece's mode an average of 15.5 out of 44 times (or 35% of the time). If we analyze where their guesses were in the correct mode family, then our experts' averaged accuracy rises from 35% to 65%. Their overall proportion of agreement on mode family was 86%. After identifying the mode family, one expert admits to guessing between the authentic and plagal version. Notice that with a 65% accuracy rate for correctly predicting the mode family, and a subsequent guess of the mode (authentic or plagal) having a 50/50 probability, that an overall accuracy level based on this strategy would be approximately 32.5 (or half of 65), which closely approximates our experts' accuracy at exact mode prediction.

Thus, our experts were able to classify the correct mode family with better-than-chance accuracy using only tallies of leaps and outlines. However, distinguishing between plagal and authentic versions of the mode appears to require additional information.

The best regression models predicted the mode with approximately 36% accuracy, and mode family with approximately 67% accuracy. We were unsure whether to define the minimum distance of a leap or outline as a third or fourth. In the behavioural experiment for this study, our experts were tested with a set of tables based on leaps and outlines defined as a fourth or greater. However, with the regression it was a simple matter to test multiple models with these different parameter settings for the minimum size of a leap or outline. The exact parameters leading to the optimal model performance are not consistent from iteration to iteration, therefore, we do not report the exact parameters for the single best model, but rather point out commonalities between the (consistently) top-performing models.

The best models for predicting mode family all defined leaps and outlines as a fourth or greater (~62-67%). In addition, they all used the rows of "totals" data. Thus, there was no single voice or single musical feature (leaps or outlines) that best predicted mode family. The best models for predicting mode, however, all used the data where leaps and outlines were defined to include thirds, and relied exclusively on the counts of outlines in the upper voice (30-36%). This suggests that in our corpus, outlines in the top voice tend to best highlight the mode. None of the theory-driven interactions led to a consistent advantage in predicting mode or mode family. For both mode and mode family, the simplest models (using only main effects) typically performed with an accuracy level within 6% of the best-performing models, which typically were more complicated. For reference, the baseline model—which simply guesses the mode most represented in the corpus—performs with an accuracy of approximately 18%.

Our results suggest that if, as the theorists imply, composers wrote leaps and outlines in a way that highlights a particular mode, they did not do it in a very consistent manner, or else the distinction between plagal and authentic was not salient in sixteenth-century polyphonic music. Of course, given our small data set, it remains a possibility that the models simply may not have had enough data.

Study 2

As explained in the Methodology section, a regression analysis was not possible for this set of data. As such, we report here only the results from the behavioural experiment.

Contrary to our expectations, the performance of the experts' overall ability to guess the mode was only marginally better than in Study 1, though their predictions for mode family were marginally worse. In this experiment, we constructed two sets of tables for each expert: one set with leaps and outlines defined as being either a third or greater, and the other a fourth or greater. This strategy may explain the decrease in overall agreement between experts, which dropped from 86% in experiment 1, to 70% in experiment 2. Nevertheless, their individual levels of accuracy overall were similar to experiment 1 (within 5%). For the correct mode, the average success rate was 39%; for mode family it was 61%. Overall, this suggests that the additional melodic information (i.e., start and end pitch, direction, and size of every leap or outline) does not help with the identification of mode family and may only marginally aid with distinguishing plagal from authentic (i.e., identifying exact mode).

Study 3

As in study 1, multinomial regression models and a behavioural experiment were carried out, this time to investigate whether tallies of pitch classes forming perfect consonances between the two voices could be used to predict the mode of a piece. (Refer to table 4).

Our experts did marginally better at guessing both mode and mode family using the vertical intervals compared with the melodic data. Their average accuracy rate was 40% for mode and 70% for mode family. Their overall level of agreement was 80%.

The best regression models predicted the mode with approximately 41% accuracy, and the mode family with approximately 70% accuracy. As with Study 1, we do not report the exact parameters of the best performing model, since this was not consistent across training iterations. The best models for predicting the mode family (63-70%) and for predicting the mode (38-41%) all used the rows of "totals" data. Also like Study 1, the simplest models typically performed within 2 to 7% of the best performing models.

Conclusion

As with any empirical investigation, an important question is: against what are we comparing the evidence? In our studies, we first evaluated the success of our experts and our models based on greater-than-chance accuracy. This is a common approach. Since we reduced a full musical work to a brute count of notes in a table, the fact that our experts (and model) were still able to deduce the mode with a rate of accuracy greater than we would expect by chance, might seem to provide evidence in support of our hypothesis.

However, perhaps chance is not an appropriate control. We tested the hypothesis that simple tallies of pitch classes in positions of emphasis (leap ends, outlines, or forming perfect consonances) would lead to distributions where higher counts of pitch classes align with the "priority" notes in a mode. However, by testing the predictive power of these tables against "chance," it inherently assumes that the distributions of all *non*-structural pitch classes will be random. Of course, we *know* that pitch classes in a given piece (or mode) do not appear with equal frequency, so we would not expect a subset of those pitch classes to be randomly distributed (e.g., Temperley 2007, p.86). Consequently, our evidence needs to be evaluated alongside a more appropriate comparison set. Thus, our regression results from Study 1 and Study 3 were each compared against the predictive power of two alternative "comparison" datasets. Melodic comparison sets (i.e., new tables of data) were created as follows:

- ***Comparison set #1:*** A table of all notes approached by step. (Since we previously tallied all "leaps," this effectively tallied the "remainder" of the notes we previously ignored.)
- ***Comparison set #2:*** The complete count of all notes (i.e., the table of pitch class distributions.)

For Study 1, it turns out that *both* of these comparison sets can predict both mode and mode family just as well as, or better than, our leaps and outlines data. In particular, the pitch class distributions produce the best performing *and* simplest model for predicting both mode and mode family. For simplicity, our original results are compared alongside the results using the comparison data in Table 5.

Harmonic (vertical) comparison sets were created as follows:

- ***Comparison set #1:*** The set of pitch classes forming *imperfect*, consonant vertical intervals (i.e., 3rds and 6ths).
- ***Comparison set #2:*** The complete count of all pitch classes in the lower part. (Since vertical intervals were tallied according to the note they were built upon, this is equivalent to counting *all* vertical intervals.)

For Study 3, it appears that our original (test) data performs much better than our two comparison sets at predicting both mode and mode family. This provides evidence in support of our hypothesis that pitch classes of greater structural significance to the mode were more likely to be emphasized with perfect unisons, fifths, and octaves compared with other intervals. However, the best model overall for predicting mode is the one that uses the simple pitch distribution data (45% accuracy).

A final experiment sought to determine a "baseline" performance for mode identification. Given the entire score (with lyrics, titles, and composers' names removed), how accurately could the mode and mode family be identified by an expert? While the mode family was *always* correctly identified (100%), the mode was only successfully identified an average of 67% of the time (and the level of agreement between experts was only 72%). While this is well above chance level performance, it demonstrates the degree of difficulty even when the full set of information contained in a score is available.

Table 5. Summary of results for all studies. Accuracy ratings for predicting mode and mode family are shown for our experts as well as our regression models comparing the predictive power of our test sets against the relevant comparison sets.

Melodic Data: Studies 1 and 2		
Experiment w/ experts	**Mode**	**Mode family**
experiment 1: pc tallies	35%	65%
experiment 2: pitch, interval size & direction	39%	61%
Regression model		
test data: leaps and outlines	36%	67%
comparison set 1: remainder notes	39%	68%
comparison set 2: pc distributions	45%	71%
Vertical Interval Data: Study 3		
Experiment w/ experts	**Mode**	**Mode family**
pitch class tallies by vertical interval	40%	70%
Regression model		
test data: perfect vertical intervals	40%	70%
comparison set 1: imperfect vertical intervals	29%	51%
comparison set 2: all intervals	25%	62%
Full Score Experiment with Experts		
	67.5%	100%

Discussion

Our study has shown that neither human experts nor a computer are very good at telling the plagal and authentic members of the mode family apart. Is this because we are not good at it, because we are looking at the wrong information, or because the distinction is not clearly marked in the music? We suggest that the third is the case.

The range of a melody is the traditional "measure" between plagal and authentic, but the range of each melody in a polyphonic piece will often be distinct. When we looked to other melodic and harmonic features, our models and experts still had difficulties determining the mode. Finally, our last "experiment" using the complete score illustrated the degree of difficulty between selecting the correct mode.

This corpus study has provided some surprising new information about mode in the Renaissance, and about methodologies for examining theories and assumptions. Using both statistical regression and behavioural experiments allowed us to compare the value of both exercises. The fact that the accuracy rates were similar for both approaches, lends validity to both. The converging evidence suggests that composers "writing in the modes" were primarily thinking in terms of mode family. Our reluctant conclusion is that the plagal-authentic distinction was maintained by theorists because of the authority of the system of modes for chant, but that it has relatively little relevance for polyphonic music.

References

Aiguino, I. (1581). *Il tesoro illuminato di tutti i tuoni di canto figurato*. Venice: G. Varisco.

Aron, P. (1525). *Trattato della natura et cognitione di tutti gli tuoni di canto figurato non da altrui piu scritti*, New York: Broude. 1979.

Glareanus, H. (1965). *Dodecachordon*. (C. A. Miller, Trans.) Musicological Studies and Documents 6. American Institute of Musicology. (Original work published 1547).

Judd, C. C. (2000). *Reading Renaissance Music Theory: Hearing with the Eyes*. Cambridge: Cambridge University Press.

Lasso, O. (1995). *The Complete Motets* vol. 11. (P. Bergquist, Ed.). A-R Editions, Madison, WI. (Original work published 1577).

Marchetto da Padova. (1985). *The Lucidarium of Marchetto of Padua*. (J. W. Herlinger, Trans.) Chicago: Chicago University Press. (Original work published ca.1317).

Pontio, P. (1959). *Ragionamento di musica*. Compiled by S.Clercx. Kassel: Bärenreiter. (Original work published 1588).

Powers, H. S., revised by F. Wiering. 2017. Mode (sections 1-3). *Grove Music Online*. Accessed 10 Dec. 2017. http://www.oxfordmusiconline.com.proxy3.library.mcgill.ca/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000043718.

Powers, H. S. (1981). Tonal types and modal categories in Renaissance polyphony. *Journal of the American Musicological Society, 34,* 428-70.

Ramis de Pareia, B. (1993). *Musica Practica*. (C. A. Miller, Trans.) American Institute of Musicology. (Original work published 1482).

Schubert, P. (1993). Mode and Counterpoint. In C. Hatch & D. Bernstein (Eds.), *Music Theory and the Exploration of the Past* (103-36). Chicago: University of Chicago Press.

Schubert, P. (1995). A lesson from Lassus: Form in the duos of 1577. *Music Theory Spectrum, 17*, 1–26.

Temperley, D. (2007). *Music and Probability*. Cambridge: MIT Press.

Tinctoris, J. (1976.) *Concerning the Nature and Propriety of Tones*. 2nd edition. (A. Seay, Trans.) Colorado College Music Press Translations, vol. 2. Colorado Springs: Colorado Music Press. (Original work published 1477).

Wiering, F. (2001). *The language of the modes: studies in the history of polyphonic modality*. New York: Routledge.

Zarlino, G. (1983). *On the Modes: Part Four of Le Istitutioni Harmoniche*. (V. Cohen, Trans.) C. V. Palisca (Ed.) Music Theory Translation Series. New Haven: Yale University Press. (Original work published 1558).

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Modeling Aural Skills Dictation

David John Baker,[1] Elizabeth Monzingo,[2] and Daniel Shanahan[3,4]

[1,2,3] *School of Music, Louisiana State University, United State of America*

[4]*School of Music, Ohio State University, United States of America*

[1]dbake29@lsu.edu, [2]emonzi1@lsu.edu, [3]daniel.shanahan@gmail.com

Abstract

Despite its abundance in curricula in music conservatory settings, research on topics pertaining to aural skills is currently limited at best. While anthologies of materials for sight singing and dictation exist, the ways in which people learn melodies are not well understood. This problem is difficult to tackle given the amount of factors that may contribute to the process, such as the complexity of the melody, the degree of exposure needed to commit a melody to long-term memory, and individual differences in cognitive ability that have been shown to contribute to an individual's performance on musical tasks. Fortunately, literature exists in related areas that serve to inform which parameters might contribute to an individual's performance in a melodic dictation setting. This paper presents findings from an experiment (N=39) modeling performance on melodic dictation tasks using both individual and musical features. Results suggest tools from computational musicology as well as individual difference measures need further exploration in order to assess the degree to which various features contribute to melodic dictation performance and inform pedagogical practices.

Introduction

Despite its near ubiquity in Conservatory and School of Music curricula, research surrounding topics concerning aural skills is not well understood. This is peculiar since almost any individual seeking to earn a degree in music usually must enrol in multiple aural skills classes which cover a wide array of topics from sight-singing melodies, to melodic and harmonic dictation– all of which are presumed to be fundamental to any musician's formal training. Skills acquired in these classes are meant to hone the musician's ear and enable them not only to think about music, but, to borrow Gary Karpinski's phrase, to "think in music" (Karpinski, 2000, p.4). The tacit assumption behind these tasks is that once one learns to think *in* music, these abilities should transfer to other aspects of the musician's playing in a deep and profound way. The skills that make up an individual's aural skills encompass many abilities, though are thought to be reflective of some sort of core skill. This is evident in early attempts to model performance in aural skills classes where C. S. Harrison, Asmus, and Serpe (1994) created a latent variable model to predict an individual's success in aural skills classes based on musical aptitude, musical experience, motivation, and academic ability. While their model was able to predict a large amount of variance (73%), modeling at this high, conceptual of a level does not provide any sort of specific insights into the mental processes that are required for completing aural skills related tasks. This trend can also be seen in more recent research that has explored the relationship between how well entrance exams at the university level are able to predict success later on in the degree program. Wolf and Kopiez (2014) noted a multiple confounds in their study attempting to asses ability level in university musicians such as inflated grading, which led to ceiling effects, as well as a broad lack of consistency in how schools are assessing success within their students. But even if the results at the larger level were to be clearer, again this says nothing about the processes that contribute to tasks like melodic dictation. Rather than taking a bird's eye view of the subject, this paper will primarily focus on factors that might contribute to an individual's ability dictate a melody.

Melodic dictation is one of the central activities in an aural skills class. The activity normally consists of the instructor of the class playing a monophonic melody a limited number of times and the students must use both their ears, as well as their understanding of Western Music theory and notation, in order to transcribe the melody without any sort of external reference. No definitive method is taught across universities, but many schools of thought exist on the topic and a wealth of resources and materials have been suggested that might help students better complete these tasks (Berkowitz, Frontier, & Kraft, 1960; Cleland & Dobrea-Grindahl, 2013; Karpinski, 2007; Ottman, 1996). The lack of consistency could be attributed to the fact that there are so many processes at play during this process. Prior to listening, the student needs to have an understanding of Western music notation at least to the degree of understanding of the melody being played. This understanding needs to be readily accessible, since as new musical information is heard, it is the student's responsibility to, in that moment, encode the melody into either hold a chunk of the melody in short term memory or pattern match to long term memory so that they can identify what they are hearing and transcribe it moments later into Western notation. So no matter what, performing some sort of aural skills task requires both long term memory and knowledge for comprehension, as well as the ability to actively manipulate differing degrees of complex musical information in real time while concurrently writing it down. Given this complexity of the task, as well as the difficulty in quantifying attributes of melodies, it is then not surprising that scant research exists on describing these tasks.

Fortunately, a fair amount of research exists in related literature which can generate theories and hypotheses explaining how individuals dictate melodies. Beginning first with factors that are less malleable from person to person would be individual differences in cognitive ability. While dictating melodies is something that is learned, a growing body of literature suggests that other factors can explain unique amounts of variance in performance via differences in cognitive ability. For example, Meinz and Hambrick (2010) found that measures of working memory capacity (WMC) were able to explain variance in an individual's ability to sight read above and beyond that of sight reading experience and musical training. Colley, Keller, and Halpern (2017) recently suggested an individual's WMC also could help explain differences beyond musical training in tasks related to tasks of tapping along to expressive timing in music. These issues become more confounded when considering other recent work by Swaminathan, Schellenberg, and Khalil (2017) that suggests factors such as musical aptitude, when considered in the modeling process, can better explain individual differences in intelligence between musicians and non-musicians implying that within the musical population. They claim there is a selection bias that "smarter" people tend to gravitate towards studying music, which may explain some of the differences in memory thought to be caused by music study (Talamini, Altoè, Carretti, & Grassi, 2017). Knowing that these cognitive factors can play a role warrants attention from future researchers on controlling for variables that might that might contribute to this process but are not directly intuitive and have not been considered in much of the past research. This is especially important given recent critique of models that purport to measure cognitive ability but are not grounded in an explanatory theoretical model (Kovacs & Conway, 2016).

The ability to understand how individuals encode melodies is at the heart of much of the music perception literature. Largely stemming from the work of Bregman (1994), Deutsch and Feroe (1981), and Dowling (1978; 1971) work on memory for melodies has begun to lay the foundation for how people learn melodies. Initial work by Dowling suggested that both key and contour information play a central role in the perception and memory of novel melodies. Interestingly enough, memory for melodies tends to be much worse than memory for other stimuli such as pictures or faces noting that the average area under the ROC curve tends to be at about .7 in many of the studies they reviewed, with .5 meaning chance and 1 being a perfect performance (Halpern and Bartlett,2010). Halpern and Bartlett also note that much of the literature on memory for melodies primarily used same difference experimental paradigms to investigate individual's melodic perception ability similar to the paradigm used in Halpern and Müllensiefen (2008).

Not nearly as much is known about how an individual learns melodies, especially in dictation setting. The last, and possibly most obvious, variable that would contribute to an individual's ability to learn and dictate a melody would be the amount of exposure to the melody and the complexity of the melody itself. There is not much research on the first of these two points, other than an approximation of how many times the melody should be played in a dictation setting according to (Karpinski, 2007, p.100) that accounts for chunking as well as the idea that more exposure would lead to more complete encoding.

Recently tools have been developed in the field of computational musicology to help with operationalizing how complex melodies are. Both simple and more complex features have been used to model performance in behavioral tasks. For example Eerola, Himberg, Toiviainen, and Louhivuori (2006) found that note density, though not consciously aware to the participants, predicted judgments of human similarity between melodies not familiar to the participants. Note density would be an ideal candidate to investigate as it is both easily measured and the amount of information that can be currently held in memory as measured by bits of information has a long history in cognitive psychology (Cowan, 2015; Miller, 1956)

In terms of more complex features, much of the work largely stems from the work of Müllensiefen and his development of the FANTASTIC Toolbox (2009), a few papers have claimed to be able to predict various behavioral outcomes based on the structural characteristics of melodies. For example, Kopiez and Müllensiefen (2011) claimed to have been able to predict how well songs from The Beatles' album *Revolver* did on popularity charts based on structural characteristic of the melodies using a data driven approach. Expanding on an earlier study, Müllensiefen and Halpern (2014) found that the degree of distinctiveness of a melody when compared to its parent corpus could be used in order to predict how participants in an old/new memory paradigm were able to recognize melodies.

These abstracted features also have been used in various corpus studies (Frieler, Jakubowski, & Müllensiefen, 2015; Jakubowski, Finkel, Stewart, & Müllensiefen, 2017; Janssen, Burgoyne, & Honing, 2017; Rainsford, Palmer and Paine 2017). that again use a machine learning approach in order to explain which of the 38 features that FANTASTIC calculates can predict real-world behavior. In addition to looking at individual features, or sets of features, as predictors, recent work by P. Harrison, Musil, and Müllensiefen (2016), Baker and Müllensiefen (2017) and the aforementioned Müllensiefen and Halpern (2014) study have used data reduction techniques, namely principal component analysis, to take measures that were successful in predicting behavioral outcomes and boil them down into a single measure of complexity that has had predictive power in modeling experimental performance. While helpful and somewhat explanatory, the problem with many of these approaches is that they take a post-hoc data driven approach with the assumption that listeners are even able to abstract and perceive these features. Doing this does not allow for any sort of controlled approach and without experimentally manipulating the parameters, which is then further confounded when using some sort of data reduction technique. This is understandable seeing as it is very difficult to manipulate certain qualities of a melody without disturbing other features. For example, if you wanted to decrease the "tonalness" of a melody by adding in a few more chromatic pitches, you inevitably will increase other measures of pitch and interval entropy. In order to truly understand if these features are driving changes in behaviour, each needs to be altered in some sort of controlled and systematic way while simultaneously considering differences in training and cognitive ability.

Aims

This paper presents findings from two experiments modeling performance on melodic dictation tasks using both individual and musical features. A pilot study was run (N=11) was used in order to assess musical confounds that might be present in modeling melodic dictation. Results of that pilot study are not reported here. Based on the results of this pilot data, a follow up experiment was conducted to better investigate the features in question.

The study sought to answer two main hypotheses:

- H1: Are all experimental melodies used equally difficult to dictate?
- H2: To what extent do the musical features of Note Density and Tonalness play a role in difficulty of dictation?
- H3: Do individual factors at the cognitive level play a role in the melodic dictation process above and beyond musical factors?

Methods

Participants Forty-three students enrolled at Louisiana State University School of Music completed the study. The inclusion criteria in the analysis included reporting no hearing loss, not actively taking medication that would alter cognitive performance, and individuals whose performance on any task performed greater than three standard deviations from the mean score of that task. Using these criteria two participants were dropped for not completing the entire experiment. Thus, 41 participants met the criteria for inclusion. The eligible participants were between the ages of 17 and 26 (M = 19.81, SD = 1.93; 15 women). Participants volunteered, received course credit, or were paid $10.

Materials Four melodies for the dictation were selected from a corpus of n=115 melodies derived from the *A New Approach to Sight Singing* aural skills textbook by Berkowitz et. al (2005). Melodies were chosen based on their musical features as extracted via the FANTASTIC Toolbox (Müllensiefen, 2009). After abstracting the full set of features of the melodies, possible melodies were first narrowed down by limiting the corpus to melodies lasting between 9 and 12 seconds and then indexed to select four melodies were chosen that as part of a 2x2 repeated measures design including a high and low tonalness and note density condition. Melodies, as well as a table of their abstracted features can be seen in Table 1 and Figures 1—4. Melodies and other sounds used were encoded using MuseScore 2 using the standard piano timbre and all set to a tempo of quarter = 120 beats per minute and adjusted accordingly based on time signature to ensure they all sounded the same absolute time duration. The experiment was then coded in jsPsych (de Leeuw, 2015) andaccesed through a browser offline with high quality headphones.

Table 1. Features of Melodies as Computed by FANTASTIC.

Melody	Note Density (ND)	Tonalness	Label
9	1.75	.71	Low ND, Low Tonal
34	1.66	.94	Low ND, High Tonal
95	3.91	.76	High ND, Low Tonal
112	3.73	.98	High ND, High Tonal

Figure 1: Melody 9

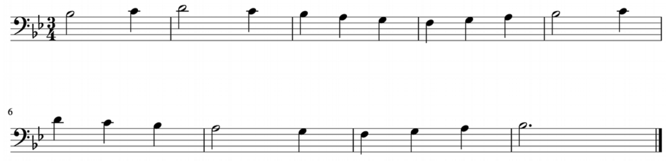


Figure 2: Melody 34

Figure 3: Melody 95

Figure 4: Melody 112

Procedure Upon arriving at the lab, participants sat down in a lab at their own personal computer. Multiple individuals were tested simultaneously although individually. Each participant was given a test packet which contained all information needed for the experiment. After obtaining written consent participants navigated through a series of instructions explaining the nature of the experiment and given an opportunity to adjust the volume to a comfortable level. The first portion of the experiment that participants completed was the melodic dictation. In order to alleviate any anxiety in performance, participants were explicitly told that “unlike dictations performed in class, they were not expected to get perfect scores on their dictations”. Each melody was played five times with 20 seconds between hearings and 120 seconds after the last hearing. After the dictation portion of the experiment, participants completed a small survey on their Aural Skills background, as well as the Bucknell Auditory Imagery Scale C (Halpern, 2015). After completing the Aural Skills portion of the experiment participants completed one block of two different tests of working memory capacity (Unsworth et al., 2005) and Raven’s Advanced Progressive Matrices and a Number Series task as two tests of general fluid intelligence (Gf) (Raven et al., 1998; Thurstone, 1938) resulting in four total scores. After completing the cognitive battery, participants finished the experiment by compiling the self-report version of the Goldsmiths Musical Sophistication

Index (Müllensifen et. al, 2014), the Short Test of Musical Preferences (Rentfrow & Gosling, 2003), as well as questions pertaining to the participants SES, and any other information we needed to control for (Hearing Loss, Medication). Exact materials for the experiment can be found at https://github.com/davidjohnbaker1/modelingMelodicDictation .

Scoring Melodies were scored by counting the amount of notes in the melody and multiplying that number by two. Half the points were attributed to rhythmic accuracy and the other half to pitch accuracy. Points were not deducted for notating the melody in the incorrect octave. Points for pitch could only be given if the participant correctly notated the rhythm. For example, in melody 34 there were 40 points possible (20 notes * 2). If a participant were to have put a quarter note on the second beat of the third measure, and have everything else correct, they would have scored a 19/20. Only if the correct rhythms of the measures were accurate could pitch points be awarded. In cases where there were more serious errors, for example if the second half of the second bar was not notated, points would have been deducted in both the pitch and rhythm sub-scores. Both the first and second author scored all melodies independently and then cross referenced for inter-rater reliability. Using a single score intraclass correlation coefficient calculation $\kappa = .96$ which suggests a high degree of inter-rater reliability (McHugh, 2012).

Results

Data Screening

Before conducting any analyses data was screened for quality. List wise deletion was used to remove any participants that did not have all variables used in modeling. This process resulted in removing four participants: two did not complete any of the survey materials and two did not have any measures of working memory capacity due to computer error. After list-wise deletion, thirty-nine participants remained.

Effects of Melodic Features

In order to investigate H_1, that melodies would differ in their degree of difficulty based on melodic features, we ran a repeated measures ANOVA using the *ez* package in R (Lawrence, 2016). Relevant statistics from the model can be seen in Table 2.

Table 2: Repeated Measures ANOVA

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η_g
(Intercept)	1	38	19.18	3.85	189.21	.000	.73
Tonalness	1	38	0.10	0.83	4.37	.043	.01
NoteDensity	1	38	5.99	1.47	154.90	.000	.46
Tonalness x NoteDensity	1	38	0.15	0.82	6.88	.012	.02

Subsequent models exploring possible exploratory covariance relationships using random slope models that used measures of working memory capacity, general fluid intelligence, and measures of musical training, none of which emerged as significant.

Differences between melodies can be see below in Figure 5.

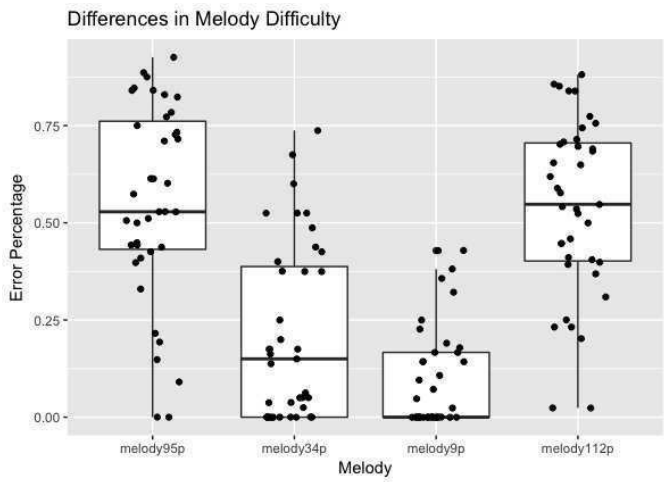


Figure 5: Boxplot of Differences between Melodies

Discussion

Here, we have investigated the extent to which both individual differences and abstracted musical features could be used to model results in melodic dictations. In order to examine H1, we ran a repeated measures ANOVA in order discern any differences in melody difficulty. As noted in Table 2, both a significant main effect of Tonalness and Note Density was found, as well as a small interaction between the two variables suggesting evidence supporting rejecting H2's null hypothesis. The interaction emerged from differences in melody means in the low density conditions with the melody with higher tonalness actually scoring higher in terms of number of errors.

While we expected to find an interaction, this condition (Melody 34) was hypothesized to be the easiest of the four conditions. With Melody 9 there was a clear floor effect, which was also to be expected as when we chose the melodies, we had no previous experimental data explicitly looking at melodic dictation to rely on. For future experiments, we will use abstracted features from Melody 9 as a baseline. The main effect of note density was expected and exhibited a large effect size. ($\eta_g = .46$). While it would be tempting to attribute this finding exactly to the Note Density feature extracted by FANTASTIC, the high and low density conditions could also be operationalized as having compound versus simple meter. Given the large effect of note density, we plan on taking more careful steps in the selection of our next melodies in order to control for any effects of meter and keep the effects limited to one meter if at all possible.

Somewhat surprisingly, the analysis incorporating the cognitive measures of covariance did not yield any significant results. While other researchers have noted the importance of baseline cognitive ability (Schellenberg & Weiss, 2013), the task specificity of doing melodic dictation as we designed the experiment might not be well suited to capture the variability needed for any effects. Hence, this paper would not be able to reject H_3's null hypothesis. Considering that other researchers have founding constructs like working memory capacity and general fluid intelligence to be important factors of tasks of musical perception, a more refined design might be considered in the future to find any sort of effects.

Taken as a whole, these findings suggest that aural skills pedagogues should consider exploring the extent to which computationally extracted features can guide the difficulty expected of melodic dictation exercises.

Conclusion

This paper demonstrates that abstracted musical features such as tonalness and note density can play a role in predicting how well students do in tasks of melodic dictation. While the experiment failed to yield any significant differences in cognitive ability predicting success at the task, our future research plans to continue incorporate measures that others have deemed important. We next plan to replicate this experiment's design with different melodies that use similar features.

Acknowledgements. The authors would like to thank Adam Rosado, Brian Ritter, and Katherine Vukovics for helping run participants on this study. Additionally, the authors would like thank Dr. Emily Elliott for providing hardware and software required to run the tests of working memory capacity and general fluid intelligence.

References

Baker, D. J., & Müllensiefen, D. (2017). Perception of leitmotives in Richard Wagner's der Ring des Nibelungen. Frontiers in Psychology, 8.

Berkowitz, S., Frontrier, G., & Kraft, L. (1960). A new approach to sight singing. WW Norton.

Bregman, A. S. (1994). Auditory scene analysis: The perceptual organization of sound. MIT press.

Cleland, K. D., & Dobrea-Grindahl, M. (2013). Developing musicianship through aural skills: A holistic approach to sight singing and ear training. Routledge.

Colley, I. D., Keller, P. E., & Halpern, A. R. (2017). Working memory and auditory imagery predict sensorimotor synchronization with expressively timed music. The Quarterly Journal of Experimental Psychology(just-accepted), 1–49.

Cowan, N. (2015). George miller's magical number of immediate memory in retrospect: Observations on the faltering progression of science. Psychological review, 122 (3), 536.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. Psychological review, 88 (6), 503.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*(1), 1-12.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. Psychological review, 85 (4), 341.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. The Journal of the Acoustical Society of America, 49 (2B), 524–531.

Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and african folk melodies by western and african listeners. Psychology of Music, 34 (3), 337–371.

Frieler, K., Jakubowski, K., & Müllensiefen, D. (2015). Is it the song and not the singer? Hit song prediction using structural features of melodies. Yearbook of Music Psychology (Jahrbuch Musikpsychologie).

Halpern, A. R. (2015). Differences in auditory imagery self-report predict neural and behavioral outcomes. *Psychomusicology: Music, Mind, and Brain*, *25*(1), 37.

Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies. In Music perception (pp. 233–258). Springer.

Halpern, A. R., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. The Quarterly Journal of Experimental Psychology, 61 (9), 1371–1384.

Harrison, C. S., Asmus, E. P., & Serpe, R. T. (1994). Effects of musical aptitude, academic ability, music experience, and motivation on aural skills. Journal of Research in Music Education, 42 (2), 131–144.

Harrison, P., Musil, J. J., & Müllensiefen, D. (2016). Modelling melodic discrimination tests: descriptive and explanatory approaches. Journal of New Music Research, 45 (3), 265–280.

Huron, D. B. (1994). The humdrum toolkit: Reference manual. Center for Computer Assisted Research in the Humanities.

Jakubowski, K., Finkel, S., Stewart, L., & Müllensiefen, D. (2017). Dissecting an earworm: Melodic features and song popularity predict involuntary musical imagery. Psychology of Aesthetics, Creativity, and the Arts, 11 (2), 122.

Janssen, B., Burgoyne, J. A., & Honing, H. (2017). Predicting variation of folk songs: A corpus analysis study on the memorability of melodies. Frontiers in Psychology, 8 .

Karpinski, G. S. (2000). Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians. Oxford University Press on Demand.

Kopiez, R., & Müllensiefen, D. (2011). Auf der suche nach den popularitätsfaktoren in den song-melodien des beatles-albums revolver eine computergestützte feature-analyse. na.

Kovacs, K., & Conway, A. R. (2016). Process overlap theory: a unified account of the general factor of intelligence. Psychological Inquiry, 27 (3), 151–177.

Michael A. Lawrence (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

McHugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica*. 2012;22(3):276-282.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. Psychological science, 21 (7), 914–919.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological review, 63 (2), 81.

Müllensiefen, D. (2009). Fantastic: Feature analysis technology accessing statistics (in a corpus): Technical report v1. Goldsmiths University of London.

Müllensiefen, D., & Frieler, K. (2006). The simile algorithms documentation 0.3. White Paper.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. PloS one, 9 (2), e89642.

Müllensiefen, D., & Halpern, A. R. (2014). The role of features and context in recognition of novel melodies. Music Perception: An Interdisciplinary Journal, 31 (5), 418–435.

Müllensiefen, D., Wiggins, G., Lewis, D., et al. (2008). High-level feature descriptors and corpus-based musicology: Techniques for modelling music cognition. na.

Ottman, R. W. (1996). Music for sight singing. Prentice Hall.

Rainsford, M., Palmer, M., & Paine, G. (2017). The musos (music software system) toolkit: A computer-based, open source application for testing memory for melodies. Behavior Research Methods, 1–19.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford, England: Oxford Psychologists Press.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of personality and social psychology*, *84*(6), 1236.

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In *The Psychology of Music (Third Edition)* 499-550

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? Intelligence.

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. PloS one, 12 (10), e0186773.

Team, R. C., et al. (2013). R foundation for statistical computing. Vienna, Austria, 3 (0).

Thurstone, L. L. (1938). Primary mental abilities. Chicago, IL: University of Chicago Press.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. Behavior research methods, 37 (3), 498–505.

Wolf, A., & Kopiez, R. (2014). Do grades reflect the development of excellence in music students? the prognostic validity of entrance exams at universities of music. Musicae Scientiae, 18 (2), 232–248.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Explaining Objective and Subjective Aspects of Musical Sophistication: Insights from General Fluid Intelligence and Working Memory

David John Baker,[1] Emily M. Elliott[2], Daniel Shanahan[3], Juan Ventura[4], Elizabeth Monzingo[5], Brian Ritter[6] and Christopher Young[7]

[1,3,5,6] *School of Music, Louisiana State University, United State of America*
[2,4,7] *Psychology Department, Louisiana State University, United States of America*

[1]dbake29@lsu.edu, [2] eelliott@lsu.edu, [3]daniel.shanahan@gmail.com,
[4]jventu4@lsu.edu,[5]emonzi1@lsu.edu,[6]britte1@lsu.edu,[7]cyoun73@lsu.edu

Abstract

Recent work in music psychology has examined the relationship between individual differences and factors that predict various aspects of musical sophistication. Some of the recent research has begun to model how musical sophistication or aptitude relates to various cognitive measures, ranging from executive functions, to measures of general fluid intelligence. Recent research has also investigated how differences in musical training may lead to differences in working, short-term, and long-term memory capacity. While some of the previously mentioned work uses continuous measures of musical sophistication, many only collect data on years of formalized musical training as opposed to a more multi-faceted view of musical sophistication. The aim of this paper is to share findings from a large study investigating how musical sophistication, as measured by the Goldsmiths Musical Sophistication Index (Gold-MSI), relates to measures of working memory and general fluid intelligence. Results using structural equation modeling (SEM) suggest working memory capacity and general fluid intelligence explain more of the variance in perceptual tasks than self-report measures of musical sophistication. In light of these findings, we suggest that further models of music perception should focus on modeling what processes contribute to a task, rather than using large, composite latent variables.

Introduction

Music and Cognitive Ability

Relative to other sub-disciplines within the field of music psychology, the relationship between musical training and cognitive ability is one of the older and more researched topics of the discipline. Dating back to 1904, Charles Spearman in his publication *General Intelligence: Objectively Determined and Measured* used tests of pitch perception as a measure to relate to general intelligence or *g*. While perhaps more affected by training than some of the other measures of interest to Spearman (e.g. Language and Mathematics), a complex relationship between various aspects of musical ability, intelligence, and factors that might mediate a relationship between the two have yet to be fully understood.

The past three decades have shown a marked increase in attention to the relationship between music and cognitive ability. Largely responsible for the initial interest in this relationship were findings published in *Nature* by Rausher et al. (1993) suggesting that listening to Mozart could have temporary beneficial effects on boosting an individual's ability to do spatial reasoning. These findings were picked up by researchers as well as the general public, even leading to government-lead initiatives to provide access to Mozart for newborn children with the hopes of instilling some sort of long-term boost in cognitive ability (Ethridge, 1998). Since then, the "Mozart Effect" has largely been reframed as findings that are better explained by a temporary increase in mood or arousal, rather than any sort of boost in cognitive ability (Schellenberg, 2014).

Despite evidence for anything resembling a "Mozart Effect", researchers have continued to explore the extent to which musical training might have a far-transfer effect to other cognitive abilities. In general, children who engage with musical training score better on tests of cognitive ability than children who do not engage in any sort of music lessons (Bigbson, Folley, & Park, 2009; Hille & Schupp, 2011; Schellenberg, 2011a; Schellenberg & Mankarious, 2012). The effects on cognitive ability tend to increase as a function of exposure (Dege & Schwarzer, 2011, Corrigall & Schellenberg, 2015; Schellenberg, 2006) and these differences remain apparent in undergraduates who are no longer engaged with music lessons (Schellenberg, 2006, 2011b).

In addition to correlational research, experimental designs have also been carried out that report mixed results. For example, Schellenberg (2004) assigned 6-year-old Canadian students to either music, voice, drama, or no lessons, and found that while no specific music group outperformed the others, the combined music groups outperformed the non-musical groups on measures of IQ and more specific abilities like attention, processing speed, verbal ability, and spatial ability. Other studies have attempted to replicate findings that show an increase in cognitive ability, but often fall short when they either do not last long enough to instil any sort of effect (Francois et al., 2013; Moreno et al., 2009) or fail to control for pre-existing differences in the participants (Mehr, Schachner, Katz, & Spelke, 2013). In his summary review on Music Training and General Cognitive Abilities, Schellenberg suggests his theory may explain a lot of these phenomena, stating that "the available evidence indicates that high-functioning individuals are likely to take music lessons, and that music lessons may exaggerate slightly their pre-existing advantages" (Schellenberg, 2016, p. 421).

If it is the case that higher functioning individuals are more likely to engage with musical activity, which then is magnified by training, it would be important to investigate which cognitive abilities are driving these relationships and to what extent. Some recent research has attempted to isolate the effects of certain cognitive abilities in reference to tasks of musical perception.

Investigating executive function Slevc et al., (2016) reported that individuals with five or more years of musical training outperformed their non-musical counterparts in both verbal and tonal working memory performance. Working memory capacity (WMC) has also been shown to contribute significantly in models of musical perception. Using a tapping paradigm and measure of two tasks of WMC (backwards digit span and operation span) Colley, Keller and Halpern (2017) reported that WMC predicted performance above and beyond temporal imagery and an auditory image self report measures of musical training. Additionally, Meinz and Hambrick (2010) found that measures of WMC additionally contributed to an individual's ability to perform sight reading at the piano in a sample of 57 musicians. A recent meta-analysis by Talamini et al., 2017 also reported that musicians tend to outperform people without musical training. Recently Swaminathan et al. (2017) reported being able to predict general fluid intelligence using musical aptitude and failed to find any unique contribution of parent's education-- a variable shown to be related to socioeconomic status (SES) that has been shown by previous research to explain individual differences in musical ability.

Nonetheless, these findings still cannot determine causality. There is a need to establish to what degree cognitive ability plays into any sort of model of musical perception.

Predictions from Process Overlap Theory

In addition to all of this research suggesting a relationship between musical and cognitive ability, additional new theoretical frameworks from cognitive psychology have suggested that in order to better understand cognitive processes, future research should model cognitive activity in terms of processes rather than using composite, latent variables as representations of theoretical constructs. Kovacs and Conway (2016) advocate for this position and suggest that cognitive tests, such as those that generate findings from the positive manifold, tap domain-general executive processes identified in working memory research, as well as other domain-specific processes. They suggest these processes are tapped in an overlapping manner across tests such that the general ones are demanded more than specific ones. If this is true, we would then suspect that measures of WMC/Gf should play a significant role in any sort of task that resembles a high executive load task such as the melodic memory task of the Gold-MSI, which requires an individual to retain information and compare it with new incoming information in a new key context. This task should then be more markedly demanding than being able to detect onset asynchrony in a stimulus, which would be more of a base perceptual ability, such as that required in the beat perception task of the Gold-MSI.

Hypotheses

Given the previous literature and theoretical background, we tested several hypotheses.

- H1: We will replicate results from Müllensiefen et al. (2014) and be able to predict scores on both the beat perception and melodic perception objective tests from the self reported sub-scales of the Gold-MSI.
- H2: We will be able to predict variance above and beyond that of the self report Gold-MSI by adding measures of Working Memory Capacity and General Fluid Intelligence into the model.
- H3: In line with predictions from Process Overlap Theory, we believe that due to the nature of the Melodic Perception Task, that WMC/Gf will better predict scores on the Melodic Perception Task than the Beat Perception Task using structural equation modeling.

In order to test these hypotheses, we used structural equation modeling (SEM) to examine the relative contribution of each of our hypothesised variables using a nested models approach (e.g., Shelton et al., 2010).

Methods

Participants

Two hundred fifty-four students enrolled at Louisiana State University completed the study. We recruited students, mainly in the Department of Psychology and the School of Music. The criteria for inclusion in the analysis were no self-reported hearing loss, not actively taking medication that would alter cognitive performance, and univariate outliers (defined as individuals whose performance on any task was greater than 3 standard deviations from the mean score of that task). Using these criteria, eight participants were not eligible due to self reporting hearing loss, one participant was removed for age, and six participants were eliminated as univariate outliers due to performance on one or more of the tasks of working memory capacity. Thus, 239 participants met the criteria for inclusion. The eligible participants were between the ages of 17 and 43 ($M = 19.72$, $SD = 2.74$; 148 females). Participants volunteered, received course credit, or were paid $20.

Cognitive Measures

All variables used for modeling approximated normal distributions. Processing errors for each task were positively skewed for the complex span tasks similar to Unsworth, Redick, Heitz, Broadway, and Engle (2009). Positive and significant correlations were found between recall scores on the three tasks measuring working memory capacity (WMC) and the two measuring general fluid intelligence (Gf). The WMC recall scores negatively correlated with the reported number of errors in each task, suggesting that rehearsal processes were effectively limited by the processing tasks (Unsworth et al., 2009).

Procedure

Participants in this experiment completed eight different tasks, lasting about 90 minutes in duration. The tasks consisted of the Gold-MSI self-report inventory, coupled with the Short Test of Musical Preferences, and a supplementary demographic questionnaire that included questions about socioeconomic status, aural skills history, hearing loss, and any medication that might affect their ability to perform on cognitive tests. Following the survey they completed three WMC tasks: a novel Tonal Span, Symmetry span, and Operation span task; a battery of perceptual tests from the Gold-MSI (Melodic Memory, Beat Perception, Sound Similarity) and two tests of general fluid intelligence (Gf): Number Series and Raven's Advanced Progressive Matrices.

Each task was administered in the order listed above on a desktop computer. Sounds were presented at a comfortable listening level for the tasks that required headphones. All participants provided informed consent and were debriefed. Only measures used in modeling are reported below.

Measures

Goldsmiths Musical Sophistication Index Self Report (Gold-MSI)

Participants completed a 38-item self-report inventory and questions consisted of free response answers or choosing a selection on a likert scale that ranged from 1-7. (Müllensiefen, et al., 2014). The complete survey with all questions used can be found at goo.gl/dqtSaB.

Tone Span (TSPAN)

Participants completed a two-step math operation and then tried to remember three different tones in an alternating sequence (based upon Unsworth et al., 2005). We modelled the three tones after Li, Cowan, Saults (2005) paper using frequencies outside of the equal tempered system (200Hz, 375Hz, 702Hz). The same math operation procedure as OSPAN was used. The tones was presented aurally for 1000ms after each math operation. During tone recall, participants were presented three different options H M and L (High, Medium, and Low), each with its own check box. Tones were recalled in serial order by clicking on each tone's box in the appropriate order. Tone recall was untimed. Participants were provided practice trials and similar to OSPAN, the test procedure included three trials of each list length (3-7 tones), totalling 75 letters and 75 math operations.

Operation Span (OSPAN)

Participants completed a two-step math operation and then tried to remember a letter (F, H, J, K, L, N, P, Q, R, S, T, or Y) in an alternating sequence (Unsworth et al., 2005). The same math operation procedure as TSPAN was used. The letter was presented visually for 1000ms after each math operation. During letter recall, participants saw a 4 x 3 matrix of all possible letters, each with its own check box. Letters were recalled in serial order by clicking on each letter's box in the appropriate order. Letter recall was untimed. Participants were provided practice trials and similar to TSPAN, the test procedure included three trials of each list length (3-7 letters), totalling 75 letters and 75 math operations.

Symmetry Span (SSPAN)

Participants completed a two-step symmetry judgment and were prompted to recall a visually-presented red square on a 4 X 4 matrix (Unsworth et al., 2005). In the symmetry judgment, participants were shown an 8 x 8 matrix with random squares filled in black. Participants had to decide if the black squares were symmetrical about the matrix's vertical axis and then click the screen. Next, they were shown a "yes" and "no" box and clicked on the appropriate box. Participants then saw a 4 X 4 matrix for 650 ms with one red square after each symmetry judgment. During square recall, participants recalled the location of each red square by clicking on the appropriate cell in serial order. Participants were provided practice trials to become familiar with the procedure. The test procedure included three trials of each list length (2-5 red squares), totalling 42 squares and 42 symmetry judgments.

Gold-MSI Beat Perception

Participants were presented 18 excerpts of instrumental music from rock, jazz, and classical genres (Müllensiefen et al., 2014). Each excerpt was presented for 10 to 16s through headphones and had a tempo ranging from 86 to 165 beats per minute. A metronomic beep was played over each excerpt either on or off the beat. Half of the excerpts had a beep on the beat, and the other half had a beep off the beat. After each excerpt was played, participants answered if the metronomic beep was on or off the beat and provided their confidence: "I am sure", "I am somewhat sure", or "I am guessing". The final score was the proportion of correct responses on the beat judgment.

Gold-MSI Melodic Memory Test

Participants were presented melodies between 10 to 17 notes long through headphones (Müllensiefen et al., 2014). There were 12 trials, half with the same melody and half with different melodies. During each trial, two versions of a melody were presented. The second version was transposed to a different key. In half of the second version melodies, a note was changed a step up or down from its original position in the structure of the melody. After each trial, participants answered if the two melodies had identical pitch interval structures.

Number Series

Participants were presented with a series of numbers with an underlying pattern. After being given two example problems to solve, participants had 4.5 minutes in order to solve 15 different problems. Each trial had 5 different options as possible answers (Thurstone, 1938).

Raven's Advanced Progressive Matrices

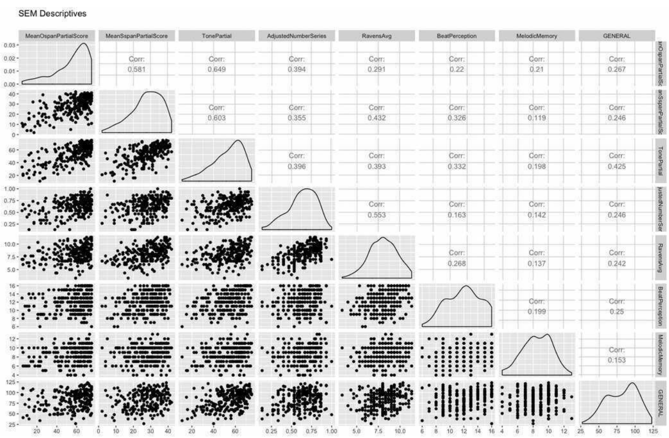
Participants were presented a 3 x 3 matrix of geometric patterns with one pattern missing (Raven et al., 1998). Up to eight pattern choices were given at the bottom of the screen. Participants had to click the choice that correctly fit the pattern above. There were three blocks of 12 problems, totalling 36 problems. The items increased in difficulty across each block. A maximum of 5 min was allotted for each block, totalling 15 min. The final score was the total number of correct responses across the three blocks.

Results

Descriptive Statistics and Data Screening

The goal of the analyses was to examine the relationships among the measures and constructs of WMC, general fluid intelligence, musical sophistication (operationalized as the General score from the Gold-MSI), in relation to the two objective listening tests on the Gold-MSI. Before running any sort of modeling, we inspected our data to ensure in addition to outlier issues as mentioned above, the data exhibited normal distributions. We report both our correlation values, as well as visually displaying our distributions in Figure 1.

Figure 1: Descriptives in Sample



Before running any modeling, we checked our data for assumptions of normality since violations of normality can strongly affect the covariances between items. While some items in Figure 1 displayed a negative skew, many of the individual level items from the self report scale exhibited high levels of Skew and Kurtosis beyond the generally accepted ± 2 (Field, Miles, & Field, 2012), but none of the items with the unsatisfactory measures are used in the general factor.

Measurement Model

We then fit a measurement model to examine the underlying structure of the variables of interest used to assess the latent constructs (general musical sophistication, WMC, general fluid intelligence) by performing a confirmatory factor analysis (CFA) using the *lavaan* package (Rosseel, 2013) using *R* (R Core Team, 2017). Model fits in can be found in Table 3. For each model, latent factors were constrained to have a mean of 0 and variance of 1 in order to allow the latent covariances to be interpreted as correlations. Since the objective measures were on different scales, all variables were converted to z scores before running any modeling.

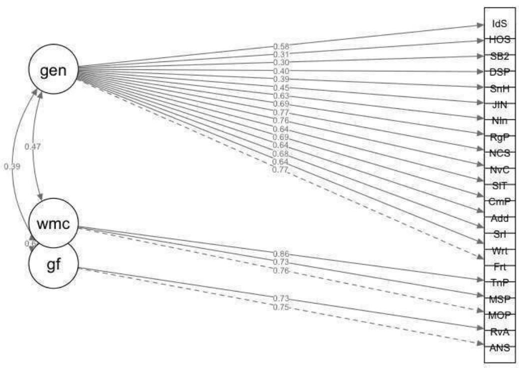


Figure 2: Measurement Model derived from CFA

Variables are defined as follows: gen: general factor latent variable; wmc: working memory capacity; gf: general fluid intelligence; zIS: "Identify What is Special"; zHO: "Hear Once Sing Back"; zSB: "Sing Back After 2-3"; zDS: "Don't Sing In Public"; zSH: "Sing In Harmony"; zJI:"Join In"; zNI: "Number of Instrumetns"; zRP:"Regular Practice"; zNCS: "Not Consider Self Musician"; zNcV: "Never Complimented"; zST: "Self Tonal"; zCP: "Compare Performances"; zAd: "Addiction"; zSI: "Search Internet"; zWr: "Writing About Music"; zFr: "Free Time"; zTP: "Tone Span"; zMS: "Symmetry Span"; zMO: "Operation Span"; zRA: "Ravens"; zAN: "Number Series".

Structural Equation Models

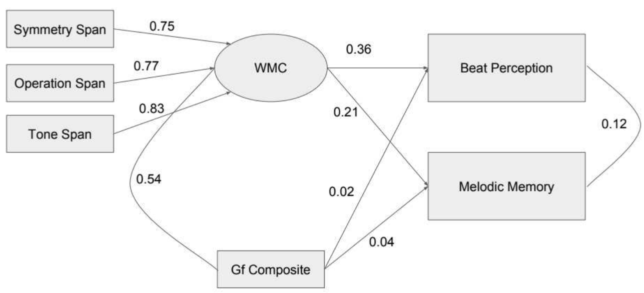
Following the initial measurement model, we then fit a series of SEMs in order to investigate both the degree to which factor loadings changed when variables were removed from the model as well as the model fits. We began with a model incorporating our three latent variables (general musical sophistication, WMC, general fluid intelligence) predicting our two objective measures (beat perception and melodic memory scores) and then detailed steps we took in order to improve model fit. For each model, we calculated four model fits: $\chi 2$, comparative fit index (CFI), root mean square error (RMSEA), and Tucker Lewis Index (TLI). In general, a non-significant $\chi 2$ indicates good model fit, but is overly sensitive to sample size. Comparative Fit Index (CFI) values of .95 or higher are considered to be indicative of good model fits as well as Root Mean Square Error (RMSEA) values of .06 or lower, Tucker Lewis Index (TLI) values closer to 1 indicate a better fit. (Beajean, 2014).

After running the first model (Model 1), we then examined the residuals between the correlation matrix the model expects and our actual correlation matrix looking for residuals above .1. While some variables scored near .1, two items dealing with being able to sing ("I can hear a melody once and sing it back after hearing it 2 – 3 times" and "I can hear a melody once and sing it back") exhibited a high level of correlation amongst the residuals (.41) and were removed for Model 2 and model fit improved significantly (χ^2 (41)=123.39, $p < .001$). After removing the poorly fitting items, we then proceeded to examine if removing the general musical sophistication self-report measures would significantly improve model fit for Model 3. Fit measures for Model 3 can be seen in Table 3 and removing the self-report items resulted in a significantly better model fit (χ^2 (171)=438.8, $p < .001$). Following the rule of thumb that at least 3 variables should be used to define any latent-variable (Beajuean, 2014) we modelled WMC as latent variable and Gf as a composite average of the two tasks administered in order to improve model fit. This model resulted in significant improvement to the model (χ^2 (4)=14.37, $p < .001$). Finally we examined the change in test statistics between Model 2 and a model that removed the cognitive measures-- a model akin to one of the original models reported in Müllensiefen et al., (2014)-- for Model 5. Testing between the two models resulted in a significant improvement in model fit (χ^2 (78)=104.75, $p < .001$). Figure 3 displays Model 4, our nested model with the best fit indices.

Table 3: CFA and SEM Model Fits

Model	*df*	χ^2	*p*	CFI	RMSEA	TLI
CFA	186	533.60	*>.001*	.83	.09	.81
1. Full Model	222	586.30	*<.001*	.83	.08	.80
2. Remove High Residuals	181	462.90	*<.001*	.86	.082	.83
3. No Self Report	10	24.11	*<.05*	.97	.08	.94
4. Gf as Observed	6	9.74	*.14*	.99	.051	.97
5. General Only	103	358.16	*<.001*	.83	.102	.801

Figure 3: Model 4



Discussion

Measurement Model

After running a confirmatory factor analysis on the variables of interest, the model fit was below the threshold of what is considered a "good model fit" as shown in Table 1 with references to above model fits. This finding is to be expected since no clear theoretical model has been put forward that would suggest that the general musical sophistication score, when modelled with two cognitive measures should have a good model fit. This model was run to create a baseline measurement.

Structural Equation Model Fitting

Following a series of nested model fits, we were able to improve model fits on a series of SEMs that incorporated both measures of WMC and measures of general fluid intelligence. Before commenting on new models, it is worth noting that the Model 5 does not seem to align with the findings from the original 2014 paper by Müllensiefen et al. While the correlation between the objective tasks is the same (r = .16), the factor loadings from this paper suggest lower values for both Beat Perception (.37 original, .27 this paper) as well as Melodic Memory (.28 original, .18 this paper). Note that two items were removed dealing with melody for memory for this model; when those items were re-run with the data, the factor loadings did not deviate from these numbers.

The first two models we ran resulted in minor improvements to model fit. While difference in models was significant (χ^2 (41)=123.39, $p < .001$), probably due to the number of parameters that were now not constrained, the relative fit indices of the models did not change drastically. It was not until the self-report measures were removed from the model, and then manipulated according to latent variable modeling recommendations, was there a marked increase in the relative fit indices. Fitting the model with only the cognitive measures, we were able to enter the bounds of acceptable relative fit indices that were noted above. In order to find evidence that the cognitive models (Models 3 and 4) were indeed a better fit than using the General factor, we additionally ran a comparison between our adjusted measurement model and a model with only the self report. While both of our nested models were significantly different, the cognitive models exhibited superior relative fit indices. Lastly, turning to Figure 3, we note that our latent variable of WMC exhibited much larger factor loadings predicting the two objective, perceptual tests than our measure of general fluid intelligence. We also note that the factor loading predicting the Beat Perception task (.36) was higher than that of the Melodic Memory task (.21). These rankings mirror that of the original Müllensiefen et al., (2014) paper and merit further examination in order to disentangle what processes are contributing to both tasks. These results align with predictions made with Process Overlap Theory (Kovacs & Conway, 2016), which predict that higher executive loads are needed for tasks of perception. While we failed to predict which task would load higher --we assumed that the ability to maintain and manipulate information in the Melodic Memory task would be better predicted by WMC than the Beat Perception task-- this might be due to the fact that performing well in a melodic memory task demands a certain amount of musical training that is not captured by either cognitive measure.

In the future, we are interested in exploring more theoretically-driven models that use specific, task oriented predictors in order to explain the relationships between the perceptual tasks and the cognitive measures. Given the results here that suggest that measures of cognitive ability play a significant role in tasks of musical perception, we suggest that future research should consider taking measures of cognitive ability into account, so that other variables of interest are able to be shown to contribute above and beyond baseline cognitive measures.

Conclusions

In this paper we fit a series of structural equation models in order to investigate the degree to which baseline cognitive ability was able to predict performance on a musical perception task. Our findings suggest that measures of WMC are able to account for a large amount of variance beyond that of self report in tasks of musical perception.

Acknowledgements. The authors would like to thanks RAs that are not authors here for help running participants for this study, as well as Elizabeth Baker for suggestions on improving the model fits for the structural equation modeling.

References

Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.

Colley, I. D., Keller, P. E., & Halpern, A. R. (2017). Working Memory and Auditory Imagery Predict Sensorimotor Synchronization with Expressively Timed Music. *The Quarterly Journal of Experimental Psychology*, , 1-16.

Corrigall, K. A., & Schellenberg, E. G. (2015). Predicting who takes music lessons: Parent and child characteristics. Frontiers in psychology, 6, 282.

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in psychology*, *4*, 222.

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. Frontiers in psychology, 2, 124.

Ethridge, P (1998, June, 24). Georgia program bringing classics to newborns. *CNN.com*. Retrieved from http://edition.cnn.com/SHOWBIZ/Music/9806/24/beethoven.for.babies/index.html

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Sage publications.

François, C., Chobert, J., Besson, M., & Schön, D. (2012). Music training for the development of speech segmentation. Cerebral Cortex, 23(9), 2038-2043.

Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, *41*(6), 779-793. doi:10.1177/0305735612452186

Hille, A., & Schupp, J. (2015). How learning a musical instrument affects the development of skills. Economics of Education Review, 44, 56-82.

Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151-177.

Li, D., Cowan, N., & Saults, J. S. (2013). Estimating working memory capacity for lists of nonverbal sounds. *Attention, Perception, & Psychophysics*, *75*(1), 145-160.

Mehr, S. A., Schachner, A., Katz, R. C., & Spelke, E. S. (2013). Two randomized trials provide no consistent evidence for nonmusical cognitive benefits of brief preschool music enrichment. PloS one, 8(12), e82007.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological science*, *21*(7), 914-919.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. Cerebral Cortex, 19(3), 712-723.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS One*, *9*(2), e89642. http://dx.doi.org/10.1371/journal. pone.0089642

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature, 365,* 611. doi: 10.1038/365611a0

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford, England: Oxford Psychologists Press.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). Journal of statistical software, 48(2), 1-36.

Schellenberg, E.G. (2016). Music training and nonmusical abilities. In S. Hallam, I. Cross, & M. Thaut (Eds.), Oxford handbook of music psychology (2nd ed., pp. 415-429). Oxford, UK: Oxford University Press.

Schellenberg, E.G. (2014). Mozart effect. In W.F. Thompson (Ed.), *Music in the social and behavioral sciences: An encyclopedia*(pp. 717-718). Thousand Oaks, CA: Sage.

Schellenberg, E. G., & Mankarious, M. (2012). Music training and emotion comprehension in childhood. Emotion, 12(5), 887.

Schellenberg, E. G. (2011a). Examining the association between music lessons and intelligence. British Journal of Psychology, 102(3), 283-302.

Schellenberg, E. G. (2011b). Music lessons, emotional intelligence, and IQ. Music Perception: An Interdisciplinary Journal, 29(2), 185-194.

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. Journal of Educational Psychology, 98(2), 457.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511-514. doi:10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch, D. Deutsch (Eds.). *The psychology of music* (pp. 499-550). San Diego, CA, US: Elsevier Academic Press. doi:10.1016/B978-0-12-381460-9.00012-2

Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B. D., & Gouvier, W. D. (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 813-820. doi:10.1037/a0019046

Silvia, P. J., Thomas, K. S., Nusbaum, E. C., Beaty, R. E., & Hodges, D. A. (2016). How does music training predict cognitive abilities? A bifactor approach to musical expertise and intelligence. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(2), 184-190. doi:10.1037/aca0000058

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, 152, 199-211. doi:10.1016/j.cognition.2016.03.017

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201-292.

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? *Intelligence*, *62,* 119-124. doi:10.1016/j.intell.2017.03.005

Talamini F., Altoè G., Carretti B., & Grassi M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PLoS ONE*, *12*(10), e0186773. doi:10.1371/journal.pone.0186773

Thurstone, L. L. (1938). Primary mental abilities. Chicago, IL: University of Chicago Press.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505. doi:10.3758/BF03192720

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635-654. doi:10.1080/09658210902998047

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences*, *20*(3), 188-196. doi:10.1016/j.lindif.2010.02.004

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Neurophysiological Effects of Dance Technologies on the Development of Parkinson's Disease

Jotthi Bansal,[1,2] Alicia Chhin,[1] Alex Zaranek,[1] Michael Balas,[1] Mike Noseworthy,[3] Rick Paulseth,[4] Matthew Woolhouse[1,2]

[1]*School of the Arts, McMaster University, Canada*

[2]*Psychology, Neuroscience, and Behavior, McMaster University, Canada*

[3]*Electrical and Computer Engineering, McMaster University, Canada*

[4]*Department of Medicine, McMaster University, Canada*

[1,2]bansalj@mcmaster.ca, [1,2]woolhouse@mcmaster.ca

Abstract

The combination of music and dance is increasingly seen as a remedial modality for neurodegenerative diseases such as Parkinson's (PD). Inspired by this notion, our team has developed a therapeutic dance-based technology for PD patients to use in the comfort of their own homes. The system provides users with an on-screen virtual dance instructor, which is animated using the motion-captured data of a real dancer, and simulates a guided dance session. Utilizing a motion-sensing camera, users can interact with the system using their body movements. In order to test the effects of regular engagement with our dance technology on PD progression we conducted a pilot study examining physiological and psychological symptoms of the disease. We tested the hypothesis that daily engagement with the system will benefit three main areas of PD: (1) neurological aspects of the disease's progression explored using magnetic resonance imaging; (2) physiological measures including coordination and mobility assessed by standard clinical motor tests; and (3) subjective ratings of mental and physical symptoms reported via a survey. Results of the clinical motor assessments showed significantly increased mobility as well as trends of reduced tremors and rigidity during and post-intervention. Subjects reported feeling significantly more energized and motivated and had less difficulty sleeping. Neurological markers are still under analysis. Study results will be considered and incorporated into further developments of the dance system in order to maximize its therapeutic potential. We hope in the near future our system will be considered by health care providers as a means of non-invasive palliative care for PD.

Introduction

As an idiopathic neurodegenerative disease (Hughes et al., 1992) with no known cure (Surmeier, 2007), health care providers are in search of therapeutics for Parkinson's (PD). Spurred on by the realization of the palliative efficacy of music and dance (Westbrook & McKibben, 1989) organizations have sprung up which provide PD-focused dance lessons, including Hamilton City Ballet's Dance for Parkinson's program in Ontario. Choreography from these classes is specifically designed to target problem areas of PD, for example shoulder locking and limb rigidity. Based on the choreography from these classes, we have developed a therapeutic dance technology (Woolhouse et al., 2015; Woolhouse & Zaranek, 2016; see http://digitalmusiclab.org/?page_id=1238 for demonstration video) on the Microsoft Xbox One gaming console—an affordable system that can be easily set up within someone's home. The system simulates a virtual dance lesson; users see an on-screen avatar instructor (see Figure 1) programmed with gestures from a professional motion-captured dancer who carries out a structured routine. Users are then to mimic the movements of the dancer to the best of their ability. As they perform each dance their movements are tracked using the motion-sensing Microsoft Kinect camera. The accuracy of the user's performance is instantaneously considered by the system and the difficulty of each gesture is modified in real time based on how well the user performs. Should a user's movements be restricted due to reduced motor control, the instructor modifies the expressivity of its gestures accordingly. For example, when a user performs well, the range and speed required to complete arm movements may increase.

Figure 1. Screenshot of the dance technology. The virtual dance instructor is seen in the centre. The performance indicator on the left signifies the gesture difficulty based on user ability and accuracy. The bottom-right corner box displays a real-time video feed of the subject captured by the Kinect camera.

As gaming consoles such as the Xbox One are usually utilized mostly by younger populations, certain measures were taken when designing the system to ensure it would be suitable for older demographics and those with motor dysfunction. One such feature is the in-game menu; it is designed as a spinning musical box (see Figure 2) that can be maneuvered using only broad arm gestures. For example, swiping your arm right or left will spin the music box displaying various game options including dances and instructions, while swiping your arm downward will select the current option. The menu was designed this way as individuals with PD generally do not have fine motor control as one would require when using standard gaming controllers. Controllers require pushing small buttons and directional joysticks—tasks that may not be possible for the PD population. Furthermore, all dances in the system are done in a seated position or require users to hold the back of a chair or support bar. This reduces the chances of users falling when trying to complete dances. If a user does happen to fall, the system's motion sensing camera will pause the game and ask the user if they require assistance. The user can answer the system with a verbal "yes" or "no" answer. If the user answers "yes" the system will automatically send a text message to the user's caregiver to alert them to help. By working in these accommodations and safety features we have ensured the application is suitable for the PD population.

Figure 2. Screenshot of the music box menu system. The box can be spun by swiping an arm left or right. Users can select dances by swiping downward.

Benefits of dance and exercise therapies for PD are more impactful if they are completed over a long period of time (Reuter et al., 1999). In order to promote regular use of the system motivating features have been incorporated. As users complete more dances and/or improve their performance they may win trophies at either bronze, silver, or gold levels which are displayed in their own personal virtual trophy case. They can also unlock new virtual studios in which they can complete dances.

To date, the system contains two dance styles: ballet and contemporary (choreographed by Dave Wilson of The Parahumans). The gestures are choreographed to simulate "dancified" versions of everyday movements. For example, some gestures may recreate the arm movements involved in opening and close a fridge, putting on a scarf, or brushing one's teeth. Having individuals complete exercises which involve mindful actions has been shown to be more effective in rehabilitation of PD than just exercises alone (Hirsch, Iyer, & Sanjak, 2016; Kwok, Choi, & Chan, 2016; Pickut et al., 2013). To our knowledge, no commensurate system is available, putting our application at the forefront of research exploring the rehabilitation of PD using music and dance.

In order to assess the efficacy of our system on influencing progression of PD we conducted a pilot study considering physiological and psychological symptoms of the disease. We tested the hypothesis that daily engagement with the system will benefit three main areas of PD: (1) neurological aspects of the disease's progression; (2) physiological measures including coordination, and mobility; and (3) self-perception of mental and physical symptoms.

Methods

In order to test our hypothesis, we developed a longitudinal study consisting of two time periods: a 4-week control and a 4-week intervention. Subjects consisted of 5 individuals over the age of 55 who had been diagnosed with PD recruited through the office of our team Neurologist. To be included in this experiment subjects were required to be within an early-to-mid stage of disease progression; they must have been capable of voluntarily sitting upright and exhibiting enough motor control to be able to attempt dances in the system. All participants reported to be within a 1 or 2 on the Hoehn and Yahr scale of PD stages (4 subjects at stage 2, 1 subject at stage 1). Of the 5 subjects 2 were female and 3 were male. Subjects were not asked to change any of their medications or current treatments targeted for PD or other disorders but were asked to report them at the beginning of the study and update any changes during the study period.

During the 4-week control period subjects maintained regular living conditions. The purpose of the control period was to gather a baseline measure of each participants' disease progression based on our target symptoms. One would expect symptoms to worsen or remain consistent throughout this period. During the 4-week intervention period subjects engaged in daily dance activities with our technology. The system was installed in subjects' homes for the duration of the intervention period and they were allowed to use the system at their own disposal as often as they liked. They were asked try and engage with the system each day throughout the 4-week period if possible.

At the beginning and end of each period an assessment day occurred in which participants would undergo the following procedures: a magnetic resonance imaging (MRI) scan, a clinical motor assessment administered by our team Neurologist, and completion of an online subjective survey regarding self-perception of mental and physical health. An additional 2 surveys were also completed at the half way point of the control and intervention periods giving a total of 5 surveys. See Figure 3 for a full study timeline followed by each subject.

The MRI scans were completed at the Imaging Research Center at St. Joseph's Hospital in Hamilton, Ontario, Canada. Prior to completing the scan all subjects were taken through safety procedures and comfortably fitted into the MRI scanner. Subjects completed scans in the supine position at resting state. The scan targeted neurological markers of PD progression; procedures included brain-iron monitoring with susceptibility-weighed imaging, blood-flow rates using arterial spin labelling,

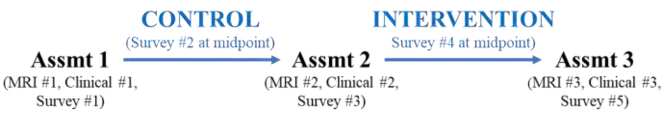


Figure 3. Timeline of the study including a 4-week control and 4-week intervention period between three assessment dates.

resting-state functional magnetic resonance imaging, and cortical-thinning measurement via high-resolution 3-dimensional scanning.

Clinical assessments administered by our team's Neurologist consisted mainly of motor control based measures from the Unified Parkinson's Disease Rating Scale (UPDRS), a widely used scale for measuring PD symptoms (Goetz et al., 2008). We included the following measures from the UPDRS: speech fluency, facial expression, resting tremors (face, hands, and feet), right and left action tremors, neck rigidity, upper and lower extremity rigidity, finger taps, hang grips, prone and supine hand movements, foot agility, ability to rise from a chair, posture, postural stability, gait, and body bradykinesia. Each of these measures was given a rating on a scale from 0 to 4 based on how prominent each motor symptom is (0 = normal, 1 = slight, 3 = mild, 4 = severe). The clinical assessment also included a finger tapping paradigm which utilized a computer keyboard; subjects had to switch between tapping two keys as quickly as possible in both vertical and horizontal movements. This task was meant to gauge reaction time and digit accuracy. All clinical motor scores were analyzed using non-parametric Wilcoxon Signed-Rank tests between time points. Finally, the clinical assessment included the Get-Up and Go Test, a commonly used task in elderly people to gauge mobility (Podsiadlo & Richardson, 1991). In this task a subject is asked to stand from a chair, walk 3 meters, turn around, and walk back to the chair and sit down. They are timed and scored based on how quickly they can complete the move. Get-Up and Go Test scores were analyzed using t-tests between study time points.

The subjective survey included questions about how long subjects used the system each day, which features of the system benefited them most, their mental symptoms including stress, depression, happiness, focus, motivation, sociability, and sleep patterns, as well as physical feelings such as fatigue and restlessness. The survey aimed to get an overview of how each subject perceived their physical and mental well-being before and after using the system. Subjective questions required individuals to rate their feelings on a scale of 1 to 4 (1 = not at all, 2 = sometimes, 3 = often, 4 = always). Subjective data was analyzed using a numerous non-parametric Wilcoxon Signed-Rank tests between timepoints.

Results

Subjects reported via the survey how often they engaged with the dance system. All subjects stated that they used the system almost every day or more than once per day. Sessions would last anywhere from 15 minutes to 60 minutes at a time. Of all survey questions regarding subjective evaluations of physical and emotional symptoms three items showed significant change between the control and intervention conditions. Subjects reported less difficulty sleeping ($p < 0.08$), feeling more energized ($p < 0.08$), and greater motivation to complete tasks ($p < 0.08$). There were also many items across subjects that showed improvement but did not reach significance. Of the five subjects one reported that they felt less restless, one felt more sociable, and one less stressed. Two reported better concentration, less difficulty staying awake, greater focus, and increased happiness during the intervention period compared to control.

One subject was unable to complete their second assessment date due to illness and therefore we do not have clinical or MRI data for this individual. They were able to complete all of their surveys and so are still considered in the subjective results described above, but this subject was eliminated from all further analyses leaving 4 subjects remaining.

Clinical assessment data examining motor symptoms of the UPDRS as well as the times keyboard tap test yielded no significant results. However, there were many trending patterns which exhibited positive effects of the intervention condition on motor control. One participant exhibited greater finger tapping speed, one showed better speech, reduced action tremor, and resting tremor, one showed greater foot agility, and two showed less upper and lower extremity rigidity. Additionally, overall scores on the UPDRS showed quite a bit of improvement even though they did not reach significance. When symptom scores for all participants were added up, the total

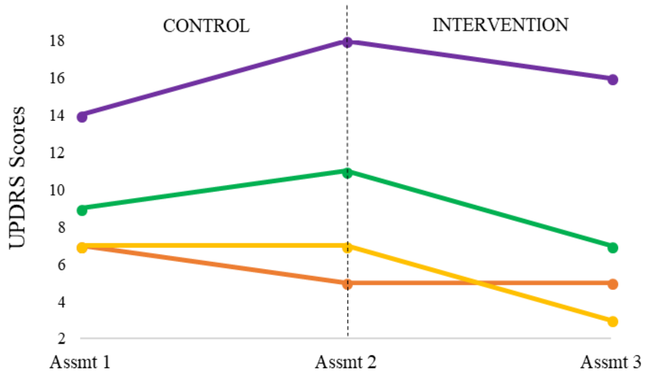


Figure 4. Unified Parkinson's Disease Rating Scale (UPDRS) symptom scores for each subject between control and intervention. Shows progression of symptoms throughout each period. Lower scores demonstrate less motor dysfunction.

UPDRS score was 37 at assessment 1, 41 at assessment 2 (after the control period) and down to 31 at assessment 3 (after intervention) displaying a decrease in overall symptoms. Figure 4 shows total UPDRS symptom scores for each individual subject. For all subjects symptoms progressed or stayed the same throughout the control period. Then, during intervention, either all subjects' scores improved by decreasing, or for one subject, stayed consistent. There was a significant difference

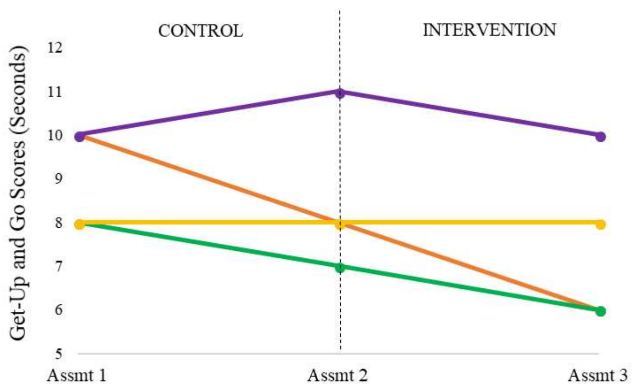


Figure 5. Get-Up and Go test scores for each subject between control and intervention. Shows progression of scored times throughout each period. Lower scores demonstrate quicker times and greater mobility.

between control and intervention for the Get-Up and Go Test ($p < 0.1$) showing increased mobility in the intervention condition. Figure 5 shows Get-Up and Go timing scores (in seconds) for each individual subject.

Analysis of neurological markers of PD via MRI scanning is still underway and result will be released in further reports.

Conclusions

As the aging population in Canada continues to grow (Denton & Spencer, 2010) the need for therapeutics addressing late-onset diseases such as PD escalates. The results of our study displayed the positive effects of dance intervention on PD symptoms. This includes increased limb mobility as seen by results of the Get-Up and Go test and performance on various motor control measures of the UPDRS. Positive effects are also seen on self-reported mental and physical states such as increased energy, greater motivation, and less difficulty sleeping. Analysis of neurological markers of disease progression is currently underway and we expect to release results shortly. Although our sample size was quite small, it is exciting to see the effects that interacting with the dance system on a daily basis had on these subjects.

Certain limitations exist within our study design that may be considered when interpreting results. This includes the length of our control and intervention periods. 4-weeks of control may not have been a long enough interval to gauge a baseline of disease progression. 4-weeks of intervention may not have been sufficient enough to see the effects of engaging with the dance technology on a daily basis. Additionally, as mentioned above, we worked with a small sample size and so it is difficult to generalize our results to the entire PD population. However, as a pilot project, this study was meant to gather an initial assessment of the dance system's effects. Results of this study and feedback given by subjects will be used to further develop the dance system in order to increase its user-friendliness and therapeutic impact. One of the major criticisms of the system was the limited dance styles available to users. We have already begun to incorporate more music and dance styles into the next version to keep users interested and engaged.

This pilot study was just the beginning of our explorations into the use of dance technology for the care of neurodegenerative disease. Future studies will include a larger sample size, a more advanced version of the dance technology, and control and intervention periods of longer lengths.

In the near future we plan to release the dance technology as a downloadable application on the Microsoft online store. Users from around the globe can easily install the system onto their home Xbox One consoles—a relatively low-cost gaming system. By tracking user data, we will be able to conduct large-scale studies on the effects of the dance application with a diverse and widespread population of PD sufferers. We hope this study may inspire researchers and clinicians to incorporate music and dance technologies into palliative healthcare procedures.

Acknowledgements. The authors would like to thank the Hamilton City Ballet and Dave Wilson of The Parahumans for their direction on dance choreography.

References

Denton, F. T., & Spencer, B. G. (2010). Chronic health conditions: changing prevalence in an aging population and some implications for the delivery of health care services. *Canadian Journal on Aging, 29*(1), 11-21.

Goetz, C. G., Tilley, B. C., Shaftman, S. R., Stebbins, G. T., Fahn, S., Martinez-Martin, P., … & Dubois, B. (2008). Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. *Movement Disorders, 23*(15), 2129–2170. https://doi.org/10.1002/mds.22340

Hirsch, M. A., Iyer, S. S., & Sanjak, M. (2016). Exercise-induced neuroplasticity in human Parkinson's disease: What is the evidence telling us? *Parkinsonism and Related Disorders, 22*(1), S78–S81.

Hughes, A. J., Daniel, S. E., Kilford, L., & Lees, A. J. (1992). Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *Journal of Neurology, Neurosurgery & Psychiatry, 55*(3), 181-184.

Kwok, J. J. Y. Y., Choi, K. C., & Chan, H. Y. L. (2016). Effects of mind-body exercises on the physiological and psychosocial well-being of individuals with Parkinson's disease: A systematic review and meta-analysis. *Complementary Therapies in Medicine, 29*, 121–131. https://doi.org/10.1016/j.ctim.2016.09.016

Pickut, B. A., Van Hecke, W., Kerckhofs, E., Crosiers, D., Parizel, P. M., & Cras, P. (2013). Gray matter density increases following a mindfulness based intervention in Parkinson's disease. *In Movement Disorders* (Vol. 28, p. 107).

Podsiadlo, D., & Richardson, S. (1991). The timed "Up & Go": a test of basic functional mobility for frail elderly persons. *Journal of the American Geriatrics Society, 39*(2), 142-148.

Reuter, I., Engelhardt, M., Stecker, K., & Baas, H. (1999). Therapeutic value of exercise training in Parkinson's disease. *Medicine and Science in Sports and Exercise, 31*(11), 1544-1549.

Surmeier, D. J. (2007). Calcium, ageing, and neuronal vulnerability in Parkinson's disease. *The Lancet Neurology, 6*(10), 933-938.

Westbrook, B. K., & McKibben, H. (1989). Dance/movement therapy with groups of outpatients with Parkinson's disease. *American Journal of Dance Therapy, 11*(1), 27-38. https://doi.org/10.1016/j.parkreldis.2015.09.030

Woolhouse, M. H., Williams, S., Zheng, S., & General., A. (2015). Creating dance activities for people with Parkinson's using motion-sensing cameras. *The International Journal of Health, Wellness, and Society, 5*(4), 108-121.

Woolhouse, M. H., & Zaranek, A. (2016). Intuitive navigation in computer applications for people with Parkinson's. *Journal of Biomusical Engineering*, *4*(115).

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Expressivity and emotion: the importance of performer controlled cues.

Aimee Battcock,[1] and Michael Schutz[2,1]

[1] *Psychology, Neuroscience & Behaviour, McMaster University, Canada*
[2] *School of the Arts, McMaster University, Canada*

[1]battcoae@mcmaster.ca, [2]schutz@mcmaster.ca

Abstract

The communication of emotion through music is shaped by a complex collaboration between composers and performers. Although compositional structure alone can reliably convey emotions (Thompson & Robitaille, 1992), they are more clearly conveyed when both expressive and structural cues are available (Quinto, Thompson, Taylor, 2014). Here we begin an exploratory investigation into the relative influence of composer (structure) vs performer (expressive) controlled cues on listener's judgement of perceived emotion. Thirty non-musician participants rated the perceived emotion of expressionless renditions of 48 excerpts from Bach's *Well-Tempered Clavier* (WTC). After each excerpt played, participants evaluated two aspects of perceived emotion, using scales for valence and arousal adapted from Russell's 2D circumplex model of affect (Russell, 1980). Using regression and commonality analyses, we examined how attack rate (timing), modality (major or minor) and pitch height contribute to listener ratings in expressionless MIDI renditions of a renowned musical work. Results indicated all three cues significantly predicted valence ratings, with modality as the strongest predictor. For arousal ratings, our timing cue of attack rate appeared as the only significant predictor. Our three-cue model for predicted ratings of expressionless stimulus, had R^2 values of 0.83 for valence ratings and R^2=0.59 for arousal ratings. This detailed breakdown of cue use and predictability extends on previous work showing how composer and performer controlled cues affect perceived valence and arousal differently. Future work comparing these results with actual performances of the same piece will clarify the relationship between performer and composer controlled cues for emotion.

Introduction

The process of conveying emotion in music involves a complex communicative relationship connecting composers, performers and listeners. Composers' emotional intentions are encoded through the notation of structural cues—pitches, rhythms, notated tempi, etc. Performers then study, interpret, and ultimately perform these notes and rhythms, adding nuance to the compositional structure. For example, performers use expressive variations in timing and intensity in order to create musical phrases, "giving life" to the raw material supplied by composers. Listeners decode the acoustic information presented, recognizing the emotions conveyed jointly by composer and performer. As such, the listener's emotional and musical experience is shaped by crucial contributions from both the performer's interpretation and composer's interpretative decisions (Gabrielsson, 1988).

Contribution of Performer vs Composer Cues

A performer's emotional intention manifests itself through the manipulation of expressive cues. The importance of these manipulations in conveyed emotion can be seen in research modelling listener responses, suggesting performer controlled cues can explain up to 70% of emotion ratings (Juslin, 2000). In that study, professional musicians encoded four emotions (happy, sad, anger and fear) within three short, well-known melodies ("When the Saints", "Nobody Knows", and "Greensleeves'). Regression modelling of five performer controlled cues (tempo, sound level, frequency spectrum, articulation and articulation variability) assessed the relationship between performers' intended expression and listeners' emotional judgements. The analyses indicated a linear combination of cues accounted for a large amount of variance in judgements, leaving only 30% left unexplained. This remained relatively consistent across regression models analyzed for each emotion investigated—emphasizing the importance of performance expression on emotional perception. This suggests that in some circumstances performers can use interpretive cues to convey emotion independent of structural cues.

Expressive cues remain important even in performances varying both structural and interpretive cues. For example, a more recent study from Juslin and Lindström (2010) found 75-80% of variance associated with listener ratings of emotion could be explained from a linear combination of both performer and composer cues. Using synthesized musical stimuli composed for experimental purposes, the authors varied eight features (pitch, mode, melodic progression, rhythm, tempo, sound level, articulation and timbre) according to a factorial design. Musically trained participants rated musical stimuli along five affective adjective scales ('Happy', 'Sad', 'Angry', 'Fearful' and 'Tender'). Regression analyses indicated composer controlled cues contributed to perceived emotion, however these cues may be less influential than performer controlled cues. In other words, the model combining performer and composer cues did not predict much more variance than Juslin (2000)'s model including only performer cues.

Unfortunately, it is difficult to make direct comparisons between these experiments as the 2000 study used recordings of well-known pieces on guitar whereas the latter presented synthesized stimuli with manipulated features. Moreover, Juslin and Lindström (2010) emphasize the importance of investigating the impact of cues using overlapping stimulus materials (in contrast to monophonic stimuli), given excerpts within their study are simpler than most music experienced in the real world. Understanding how performer and composer cues function in more complex, polyphonic (i.e., multi-voiced) musical stimuli as they suggested, provides valuable complementary insight into the complex process of communicating musical emotions.

Comparisons between performer and composer controlled cues are complicated by differential findings regarding the two common dimensions assessed—valence vs. arousal

(Quinto & Thompson, 2013). In a series of studies, listeners rated emotional valence and arousal for musician-composed excerpts either performed with a specific intended emotion or rendered neutral and expressionless through MIDI software. In order to quantify composer and performer controlled cues, musical excerpts were subject to acoustic analysis using Praat (Boersma & Weenink, 2018) and MIRToolbox (Lartillot, Toivianinen, & Eerola, 2008) software. Overall, regression analyses on listener ratings indicated cues predicted 58-59% of valence and arousal variance. The model included compositional cues such as mode, mean fundamental frequency, range (number of semitones between the lowest and highest frequencies), mean interval size, as well as performance cues of articulation, mean intensity level, intensity variability, high-frequency energy and tempo. Intriguingly, compositional cues such as mode had a greater influence on valence ratings whereas performance cues more strongly influenced arousal ratings. In addition, regardless of the intended emotion, expressive interpretation enhanced ratings of emotional valence and arousal. As such, it appears a performer's interpretative decisions strongly influence emotional communication, particularly with respect to arousal.

Structural (composer-controlled) cues for emotion

Listeners are able to decode a wide range of cues to perceive musical emotion. These cues range from low-level psychoacoustic features such as tempo (Balkwill & Thompson, 1999; Gagnon & Peretz, 2003), articulation and dynamics (Eerola, Friberg, & Bresin, 2013) to features specific to music, such as modality (Costa, Fine, Enrico, & Bitti, 2004; Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Hevner, 1937). A large body of experimental evidence shows that tempo, modality and pitch are effective in communicating emotions to an audience (Gabrielsson & Juslin, 1996; Hevner, 1937).

Faster timing information like high tempos is strongly associated with affective terms higher in emotional arousal, whereas slower tempos are related to affective terms lower in emotional arousal (Gagnon & Peretz, 2003; Juslin, 1997). In addition, listeners rate higher pitches and major modalities as expressing more positively valenced emotions than lower pitches and minor modes (Costa et al., 2004; Crowder, 1985; Hevner, 1937). Furthermore, work examining the relative contributions of cues to expressed emotion have indicated timing information and modality cues are important for listener ratings.

Our previous investigation explored the influence of attack rate (note attacks per second and a function of tempo), modality and pitch height cues within the polyphonic work of Bach's *Well-Tempered Clavier (WTC).* Regression analyses determined a linear combination of these three cues predicted for 75-79% of variance associated with valence ratings and 48-51% of variance associated arousal ratings. These results suggest attack rate, modality and pitch height are central cues for listener perception within naturalistic polyphonic music that is reflective of the musical complexity experienced in everyday listening.

Current Experiment

The current study is the first in a series aiming to exploring the influences of composer vs. performer controlled cues for emotion in in Bach's *The Well-Tempered Clavier (WTC)* Book 1. The *WTC* is frequently used as a teaching tool for musicians, and mastery of the *WTC* are required for Royal Conservatory of Music standards. Consequently it is a useful set of pieces for a naturalistic study as it is widely revered—both for its theoretical structure and as a set of teaching pieces (Royal Conservatory of Music, 2015). Furthermore, its "balance" with respect to modality (with equal numbers of pieces in all 24 keys) makes it useful for studying issues of emotion, given the well-known importance of modality in its communication.

Our experiment aims to explore the importance of cues in conveyed emotion in an expressionless context, where composer controlled cues are 'isolated' from performer cues. Further work will compare these findings to responses of the same pieces as performed with expression, in order to compare how composer vs performer cues impact the perception of emotion in music. Specifically, we examined the effect of expressionless musical stimuli, on listeners' use of cues to perceive emotion.

Method

Participants

Thirty participants (22 female, 8 male) ranging in age from 18- 51 years (M = 20.45, SD = 5.81) recruited from the McMaster University undergraduate psychology pools and word of mouth participated in this study. Participants reported normal hearing and normal or corrected-to-normal. Each participant received course credit in return for participation.

Musical Stimuli

Our stimuli included MIDI rendered excerpts of J.S Bach's *WTC's Book 1* (n=48*)*. Excerpts contained the first eight measures of each piece, and included a two-second fade out starting at the ninth measure. In order to use the most normative timing information possible, we set the tempo for each piece according to the median tempo of values of 13 prominent performances analysed by Willard Palmer (Palmer, 1994) and calculated attack rate information from these values. Stimuli durations ranged 6 – 104 seconds (M=26.7 seconds, SD=16.8). We used Amadeus Pro to cut and prep the stimuli for experimental use.

Procedure

Table 1. Regression model for attack rate, modality, pitch height on ratings of valence and arousal for MIDI recordings. Unstandardized beta values indicate strength and direction of relationship between each predictor variable and valence and arousal ratings. Major mode serves as reference for the cue of modality.

	Valence				Arousal			
Predictor Coefficients	*B*	*SE*	*t*	*p*	*B*	*SE*	*t*	*p*
Attack Rate	0.245	0.036	6.817	p<0.001	4.443	0.597	7.444	p<0.001
Modality	-1.315	0.158	-8.322	p<0.001	2.093	2.624	0.798	p=0.429
Pitch Height	0.047	0.018	2.664	p<0.05	0.180	0. 295	0.611	p=0.545

Before each experiment, participants completed a consent form and musical experience survey. The experiments took place in a sound-attenuating booth, where the research assistant gave verbal instructions for the emotion rating task. Four practice trials preceded the experimental trials, using alternate recordings as performed by Rosalyn Tureck (Bach, 1953). After each excerpt, participants rated perceived emotion on scales of valence from 1 (negative) to 7 (positive), and arousal from 1 (low) to 100 (high). Each participant listened to an individually randomized order of the 48 excerpts.

We conducted the experiment using PsychoPy (Peirce, 2007), a Python-based program. Participants heard stimuli at a consistent and comfortable listening level through Sennheiser HAD-300 headphones. They viewed rating scales on a DELL monitor and provided responses via an Apple mouse connected to a 13-inch MacBook Pro located outside the booth.

Results

We first examined the relationship between attack rate, modality and pitch and emotion ratings to provide a useful comparison to previous studies of emotion. This involved assessing the relationship between acoustic cues (attack rate, mode, pitch) and listener responses, as captured by a two-dimensional model of valence and arousal, using a least squares standard multiple linear regression. Next, we used a novel approach not frequently used in music research—commonality analyse—to determine partitioned variance within the regression models.

Regression analysis

A standard multiple linear regression conducted on the ratings of valence revealed all three acoustic cues—attack rate, modality, and pitch height–significantly predicted ratings of valence (Table 1). Our three-cue model for valence had a significant regression equation of (F(3.44)= 74.07, p<0.001), with an R^2 of 0.83. Participants' predicted valence ratings are equal to 0.245(attack rate) + 1.315 (mode) + 0.047 (pitch height), where mode is coded as 0=minor, 1=Major. Valence ratings increased by 0.245 per unit increase in note attacks per second, and 0.047 per unit increase in pitch height. Listeners rated Major modes as 1.315 times higher in valence.

Regression results for arousal ratings indicated only the cue of attack rate significantly predicted for listener responses (Table 1). The model for arousal ratings had a significant regression equation of (F(3,44)= 21.53, p<0.001, with an R^2 of 0.59. Participants' arousal responses are predicted by the relationship with attack rate, where arousal ratings increased by 4.443 per unit increase in note attacks per second.

Commonality Analysis

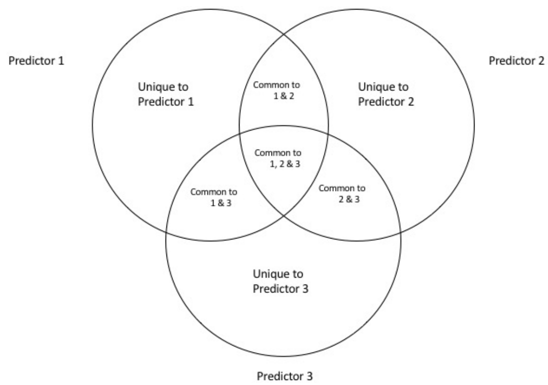


Figure 1. Visual representation of regression predictor relationships using commonality analysis adapted from Capraro & Capraro (2001).

To further explore the relative strengths of each cue, we examined their unique and shared contributions to predictions of participant response (Figure 1) using commonality analysis to decompose the R^2 of each model. Here, 'shared' variance between predictors (overlapping regions in Figure 1) represent the variance those variables have in common with the dependent variable (Ray-Mukherjee et al., 2014). For the interest of our study and in line with Capraro & Capraro (2001) we interpret negative commonalities as zero, given the small negative values calculated in addition to the negative contribution of attack rate, modality and pitch height combined.

Valence. Uniquely, modality accounted for the largest amount of variance for ratings of valence (26%), followed by attack rate (17%) and pitch height (3%). Modality's shared variance with attack rate accounted for the most variance overall (37%), suggesting a relationship between the two cues and predicting more than shared variance predicted by modality and pitch height (6%).

Arousal. Commonality analysis revealed attack rate as the most important predictor for the prediction of arousal rating variance (51%). Modality (0.6%) and pitch height (0.3%) contributed minimally to the model, in accordance with the non-significant regression findings. Again, the relationship between attack rate and modality accounted for a portion of shared variance (7.5%) larger than the portion accounted for by modality and pitch height uniquely or shared between attack rate and pitch height (0.7%).

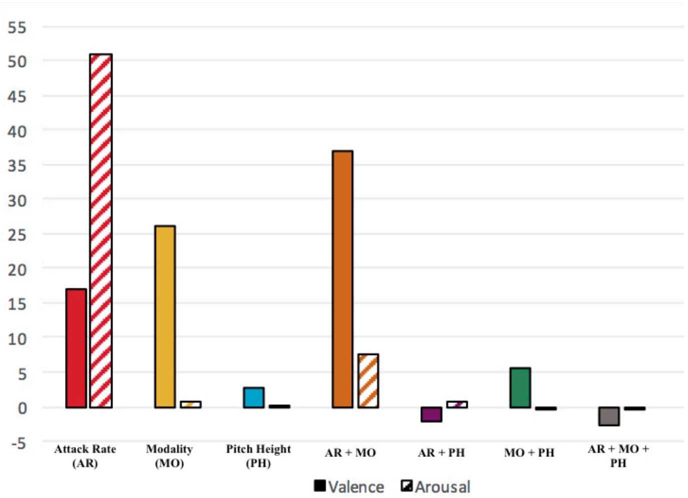


Figure 2. Unique and shared variance of Valence and Arousal ratings by musical cue. Attack rate's unique and shared contribution with modality explained the majority of variance across perceived ratings of arousal. The two bars for each cue depict ratings made of both (1) valence and (2) arousal ratings of emotion.

Discussion

The current exploratory investigation examined the use of composer selected (structural) cues on listener perception of emotion in expressionless music. The experiment is the first in a set that comparing the importance of composer vs performer controlled cues on expressed emotion. Here we build upon previous corpus analysis of Bach's timing and pitch cues (Poon & Schutz, 2015) along with modality, by empirically assessing their perceptual consequences. Using three cues of attack rate (timing), modality and pitch height, we examine the influence of each cue on listener ratings of emotional valence and arousal. Overall, the results suggest the importance of composer controlled cues in conveyed emotion, varying according to the component of emotion measured.

According to our model for valence, attack rate, modality and pitch height significantly predicted listener ratings of emotional valence. Listeners attended to all three cues to perceive the positive or negative nature of the emotion conveyed. The relationships found between structural cues and perceived emotion aligned with previous explorations of communicated emotion in speech and music, where higher pitch height (Bachorowski, 1999) and faster timings (Breitenstein, van Lancker, & Daum, 2001) are linked with positively valenced emotions. In addition, Western listeners often use major and minor modalities to distinguish between emotions with positive or negative valence (Gagnon & Peretz, 2003; Hevner, 1935). In contrast, the regression model for ratings of arousal revealed attack rate as the only significant predictor for arousal (although it contributed significantly to valence). Here, even in music without performance expression, timing information is crucial to the perception of emotional arousal. In addition, the consistent significance of attack rate across both dimensions of emotion is consistent with research suggesting timing as the most salient cue for emotion recognition and perception in music (Gagnon & Peretz, 2003).

Our linear model built using only three cues – attack rate, modality and pitch height- accounted for 83% and 59% of variance in participants' ratings across valence and arousal respectively. These differences are reflective of the limited number of features included in the models, specifically additional features that may be important for ratings of emotional arousal. This may also suggest composer selected cues have greater influence on valence ratings than arousal ratings, in alignment with previous work showing compositional cues such as mode impacted listener rating of valence more than performer cues like articulation (Quinto & Thompson, 2013).

The results also indicate our models could predict for large amounts of variance with only composer selected cues. In addition, these results suggest the selected cues are central to listener perception within naturalistic polyphonic music—reflecting a type of musical complexity experienced by listeners in everyday life. Previous work analyzing models of both performer and composer controlled cues from Juslin and Lindström (2010), demonstrated that 75-80% of variance associated with listener ratings of emotion could be explained from a linear combination of both performer and composer cues. Here we show 59-83% of variance from listener ratings can be predicted with just three composer cues in polyphonic music. Again, the low predictability of our arousal model in contrast to Juslin and Lindström (2010) and similar level of predictability in valence rating variance may reflect the influence of our composer cues on each dimension.

Relative Cue Strengths

The unique component of explained variance represents variance accounted for independently by each predictor. In contrast, shared variance refers to the 'common' variance predicted by two variables in the model (Seibold & Roper, 1979). Commonality analysis affords detailed information breaking down the complex cue relationships associated with listener perception within multi-voiced stimuli. Variance partitioning of attack rate (timing), modality and pitch height ultimately allow us to statistically compare how much each cue contributes, gives us a sense of their musical importance in this experimental context.

The breakdown of cue contribution for valence ratings demonstrated that modality predicted the most variance (26%), followed by attack rate (17%), with pitch height contributing a small amount (3%). This suggests modality is pivotal to conveying emotional valence, over cues of timing and pitch. These findings are further supported from research arguing modality is an important cue for the perception of emotional valence (Hunter, Schellenberg, & Schimmack, 2008; Pallesen et al., 2005). The shared contribution of modality and attack rate predicted the largest amount of shared variance. This points to the compositional relationship between modality and attack rate found in Baroque and Classical Western music, where major modalities are often faster in timing cues and minor modalities are consistently slower (Post & Huron, 2009).

Attack rate accounted for the largest amount of variance of arousal ratings uniquely (51%) and commonly (7%) with modality. Modality and pitch height played a smaller role for ratings of arousal uniquely (Figure 2) with pitch height contributing minimally (<1% uniquely). This is surprising given research indicating modality as the cue of highest importance, followed by features such as tempo and dynamics (Eerola et al., 2013). Our results may vary from these findings however, due to differences in stimuli – our experiment uses performances of baroque era repertoire to explore cue relationships with perceived emotion, in contrast to experimentally composed excerpts with specific intended emotions. Additionally, we used polyphonic (i.e., multi-voice) stimuli, in contrast to the single melodic line stimuli often used in previous research. This suggests that when presented with more ambiguous, complex music participants may rely more on timing cues like attack rate than of the cue modality or pitch.

Furthermore, our results demonstrate a similar finding to the work Quinto & Thompson (2013), where performer controlled (expressive) cues are most important for the dimension of emotional arousal. Although our use of MIDI performances minimized expressive interpretations, listeners used timing information, usually manipulated by the performer for expressive purposes, to decode the conveyed emotion. In this study, the composer alone determined attack rate, as the pieces featured no expressive timing variation. Our valence model found modality the most important cue, with attack rate more important than pitch height. Again, here we made the timing information expressionless, indicating compositional cues influence valence ratings more than cues under the performer's control, as found in previous work (Quinto & Thompson, 2013). However, if the inclusion of expressive timing improves this model, it points to the significance of expressive cues over compositional cues like pitch height for listener ratings of valence.

Future Directions

As the first step in an exploratory investigation on the influence of composer versus performer cues on perceived emotion in music, future work will expand on these findings by exploring the role of performer interpretation through comparison of different performances of this same set of pieces. Specifically focusing on the influence of expressive timing information (in the form of attack rate), we will look to explore how performers' expressive decisions on timing influence how listeners rely on cues to rate conveyed emotion.

Previous work has touched on expressive timing and conveyed emotion, indicating its impact on listener perception of emotion (Bhatara, Tirovolas, Duan, Levy, & Levitin, 2011; Juslin & Madison, 1999). Future research will extend on these findings, to consider timing decisions of multiple notable performers and compare the impact of these decisions in comparison to composer chosen cues. For this, the current experiment serves as a useful baseline of emotion ratings to contrast with future explorations of expressive choices of performers. The MIDI rendered stimuli allows us to vary attack rate in a controlled way, while other cues remain constant. This will allow us to return to more ecological assessment with future research, as global timing information can be set to match with the timing selections of notable performers. Furthermore, given the *WTC's prominence* in musical training, continued investigation using the MIDI rendered excerpts will provide a foundation to compare against performer recorded versions, for an in-depth understanding on how listeners use cues to discern emotion in musical performances.

Acknowledgments. We would like to thank Joyce Chan, Isabella Lamantia, Alisha Song, Andrea Alves, and Nic Ballerin for their assistance with data collection and Brian McNulty for assistance with stimulus preparation. Additionally, we would like to thank Sharmila Sreetharan for her ongoing support. This work was supported by NSERC (RGPIN/386603-2010), Early Researcher Award (ER10-07-195), Canadian Foundation for Innovation (CFI-LOF 30101), and McMaster Arts Research Board grants to Dr. Michael Schutz.

References

Bach, J. S. (1953). *Bach: The Well-Tempered Clavier Book I [Recorded by R. Tureck]*. [DC]. RFA-Neef, Wittingen, Germay : Deutsche Grammonphon(1953).

Bachorowski, J.-A. (1999). Vocal Expression and Perception of Emotion. *Current Directions in Psychological Science*, *8*(2), 53–57. http://doi.org/10.1111/1467-8721.00013

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception: An Interdisciplinary Journal*, *17*(1), 43–64.

Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 921–934. http://doi.org/10.1037/a0021922

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program].

Breitenstein, C., van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion*, *15*(1), 57–79. http://doi.org/10.1080/0269993004200114

Capraro, R. M., & Capraro, M. M. (2001). Commonality analysis: Understanding variance contributions to overall canonical correlation effects of attitude toward mathematics on geometry achievement. *Multiple Linear Regression Viewpoints*, *27*(2), 16–23. Retrieved from http://mlrv.ua.edu/2001/VOL_27_2/vol27_n2_a3.pdf

Costa, M., Fine, P., Enrico, P., & Bitti, P. E. R. (2004). Interval distributions, mode, and tonal strength of melodies as predictors of perceived emotion. *Music Perception: An Interdisciplinary Journal*, *22*(1), 1–14.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, *4*, 1–12. http://doi.org/10.3389/fpsyg.2013.00487

Gabrielsson, A. (1988). Timing in music performance and its relations to music experience. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 27–51). New York, NY: Clarendon Press/Oxford University Press.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 68–91. Retrieved from http://pom.sagepub.com/content/24/1/68.short

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to" happy-sad" judgements in equitone melodies. *Cognition and Emotion*, *17*(1), 25–40.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, *47*, 103–118. Retrieved from http://www.jstor.org/stable/1416710

Hevner, K. (1937). The affective value of pitch and tempo in music. *The American Journal of Psychology*, *49*(4), 621–630.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, 383–418.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(6), 1797–1813. http://doi.org/10.1037//0096-1523.26.6.1797

Juslin, P. N., & Lindström, E. (2010). Musical expression of emotions: Modelling listeners' judgements of composed and performed features. *Music Analysis*, *29*(1–3), 334–364. http://doi.org/10.1111/j.1468-2249.2011.00323.x

Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception: An Interdisciplinary Journal*, *17*(2), 197–221.

Palmer, W. A. (1994). *J.S. Bach: The Well-Tempered Clavier* (Third). USA: Alfred Music Publishing.

Peirce, J. W. (2007). PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. http://doi.org/10.1016/j.jneumeth.2006.11.017

Post, O., & Huron, D. (2009). Western Classical music in the minor mode is slower (Except in the Romantic period). *Empirical Musicology Review*, *4*(1), 2–10.

Quinto, L., & Thompson, W. F. (2013). Composers and Performers Have Different Capacities to Manipulate Arousal and Valence. *Psychomusicology: Music, Mind, and Brain*, *23*(3), 137–150. http://doi.org/10.1037/a0034775

Ray-Mukherjee, J., Nimon, K., Mukherjee, S., Morris, D. W., Slotow, R., & Hamer, M. (2014). Using commonality analysis in multiple regressions: A tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution*, *5*(4), 320–328. http://doi.org/10.1111/2041-210X.12166

Royal Conservatory of Music. (2015). Piano Syllabus. Toronto, ON: Frederick Harris Music Co., Ltd.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Seibold, D. R., & Roper, R. E. (1979). Psychosocial determinants of health care intentions: Test of the Triandis and Fishbein models. *Communication Yearbook*, *3*, 625–643.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Sharing and Enacting Cognitive Metaphors in Musical Distributed Contexts: A Case Study from IRCAM

José L. Besada

GREAM, Université de Strasbourg, France
APM, IRCAM, France

besadajl@gmail.com

Abstract

Cognitive linguistics and cognitive psychology have led to an important shift that has touched many fields of research wherein multimodal givens are scrutinized. For instance, Conceptual Metaphor Theory and Conceptual Blending Theory have reached music theory, helping to analyse a well-established corpus of theoretical texts and musical concepts, but their implementation in current musical practices remains relatively infrequent. This paper aims at understanding the role of cognitive metaphors when shared and implemented within a real situation of musical creativity, in which several ideations may also be physically enacted. For that purpose, it is based on qualitative data through the ethnographic tracking of several meetings that took place primarily at IRCAM, from 2016 to 2017, gathering a composer, an astrophysicist, and a computer music designer. Data provided a set of linguistic, graphical, musical, and gestural information that can be qualitatively analysed through the aforementioned cognitive theories. In addition, it helps to criticise particular extensions of cognitive models from the individual mind towards the field of social cognition. Last, data has also shown how, depending on the degrees of shared metaphoricity, actions could boost or weaken the common task of the group.

Introduction

Although the impact of cognitive linguistics on music theory is a fact that can be particularly traceable since the late 90s, Lawrence M. Zbikowski (2002) marked a crucial milestone in this path. His book largely implemented Conceptual Metaphor Theory (CMT) (Lakoff & Johnson, 1980) and Conceptual Integration Networks (Fauconnier & Turner, 1998)—later baptised as Conceptual Blending Theory (CBT) (Fauconnier & Turner, 2008)—into the field of music theory. This theoretical viewpoint has spread and covers nowadays a large range of musical and musicological topics. Nevertheless, its influence on the study of contemporary music is still quite sparse (Bauer, 2004; Besada, 2017, 2018; Brower 1997-1998; Kendall, 2010; Stefanou, 2018).

The aforementioned cognitive theories, mainly CMT, have clearly reached other nonlinguistic domains beyond music, in particular gesture studies (Cienki & Müller, 2008). As gesture is also an important current topic of musicology, some scholars have started to analyse regular gesticulation that happens during real musical practices in the light of cognitive linguistics, e.g. conducting (Boyes Braem & Bräm, 2000), teaching music (Chuang, 2012), playing piano (Poggi, 2006), or singing (Zbikowski, 2011a).

This paper aims to provide a qualitative analysis of the role of metaphorical thought within a case of distributed musical creativity, based on multimodal data retrieved from an ethnographic tracking. In particular, it will also help to discuss the potential and the limits of some extensions of CBT that have arrived from cognitive anthropology and pragmatics.

Case Study

In 2014, Catalan composer Hèctor Parra received a commission for a new musical piece for large ensemble, orchestra and electronics. The work, entitled *Inscape*, was premiered at Barcelona on 19 May 2018. The composer took advantage of several theories about cosmic black holes as a source of inspiration, and for that purpose he gathered a team of collaborators around his project. The whole team was constituted as follows:

- Hèctor Parra: composer; deep interest in popular scientific books; skills in computer music design; prior experience collaborating with physicists (Parra, 2008)
- Thomas Goepfer: computer music designer at IRCAM; important knowledge of acoustics.
- Jean-Pierre Luminet: astrophysicist; amateur pianist; prior experience collaborating with contemporary music composers (Luminet, 1993).

The team scheduled some meetings in order to develop several pre-compositional—i.e. before Parra started to fix his musical ideas on a score—choices and concepts (see Table 1).

Table 1. Schedule around Parra's compositional project that took place from January 2016 to January 2017. Legend: P = = Parra, G = Goepfer, and L = Luminet.

Venues and dates of the meetings	P	G	L
Luminet's (Marseille): 27-29 January 2016	×		×
IRCAM (Paris): 29 June-1 July 2016	×	×	
IRCAM (Paris): 5-6 October 2016	×	×	
IRCAM (Paris): 9-11 January 2017	×	×	
IRCAM (Paris): 16 January 2017	×	×	×

Ideation of Musical Concepts and Objects

Metaphors and Blends of Motion in Musical Thought

As stated in the Introduction, CMT provides a useful framework for explaining the way we conceptualise music. In particular, a quantitative study has evinced that a linguistic discourse about music incorporates a significantly wider percentage of metaphorical terms than an ordinary one (Pérez Sobrino & Julich, 2014). The study also highlighted the prevalence of metaphors of musical motion. This aspect is in tune with precedent theoretical research that surmised the importance of image schemata, based on forces and paths—many of them potentially embodied—of the physical world, for conceptualising music in Western culture (Brower, 2000; Johnson & Larson, 2003; Larson, 1997-1998; Saslaw, 1996). Empirical evidence from psychological experiments has reinforced this hypothesis (Eitan & Granot, 2006).

Similarly, CBT and some of its extensions—discussed in the next section—have been quite fruitful for explaining the emergence of musical concepts as a blend of sound experience and physical motion, also taking into account cross-cultural variations (Antović, 2018). Again, psychological experiments on multimodal perception have reinforced this theoretical frame (Athanasopoulos & Antović, 2018).

All in all, this evidence seems to point towards an embodied conception of musical experience in which "music materials have the *potential* to serve as analogues of motor movement" (Zbikowski, 2011b, p. 189). This statement is not far from what Arnie Cox (2001, 2016) has surmised as the mimetic hypothesis of embodied musical meaning, in particular the mimetic motor imagery.

Material Anchors for Musical Blends

An important extension of CBT is the concept of *material anchors* (Hutchins, 2005), also incorporated by the pioneers of this theory in the revised version of their canonical work (Fauconnier & Turner, 2008, pp. 195-217). In this model, one of the inputs is of material nature; the blend therefore integrates such materiality that renders conceptual stability to the blending process. Material anchors can be useful for describing several cognitive tasks during compositional practices. Beyond considerations of Western musical notation as a kind of textual material anchor itself (Hayes, 2008, p. 21), the cognitive model may help to understand, for instance, the way composers interpret scientific diagrams as potential creative sources (Besada, 2017, pp. 107, 185-186).

The *Inscape* Project [1]: Sharing Cognitive Metaphors

CMT itself is insufficient to explain the way in which complex concepts of physics were translated into musical ideations during the *Inscape* project. It would be misleading to suggest that our daily use of motion metaphors of music naturally leads to physics-based conceptions of compositional practices. At most, it seems that physics-based conceptions of compositional practices could find a good "cognitive accommodation" against the backdrop of our daily use of motion metaphors of music (Besada, 2018, p. 16).

Conversely, material anchors are optimal for describing some complex ideations that were developed by the *Inscape* team. One of their goals was to transpose acoustically the effect of gravitational waves caused by two merging black holes, by means of the electronic devices. For that purpose, they took advantage of the diagrams of the empirical detection of the physical phenomenon (Abbott *et al*, 2016).

Parra already planned an analogous conception for a previous instrumental composition—*Caressant l'horizon*—in which he yearned for imagining "what we could physically experience if we were traversed by the gravitational waves generated by the collision of two black holes" (cited in Besada, 2018, p. 26). His compositional sketches show the informal elaboration of rhythmical patterns, next split into different registers, which are based on a material anchor (p. 27). Here, the composer projected a double image schema—rooted in the Western musical notation—on a preexistent diagram predicting the behavior of gravitational waves: PITCH RELATIONSHIPS ARE RELATIONSHIPS IN THE VERTICAL SPACE and TIME FLOWS FROM LEFT TO RIGHT.

The emergence of this material anchor deserves two remarks. First, sagittally-oriented linguistic expressions of time are particularly recurrent in Western languages—English among them—but there is also some evidence of laterally-oriented mental timelines, mainly expressed by gesture (Casasanto & Jasmin, 2012). Nevertheless, a left-to-right orientation of this timeline seems to be culturally acquired and related to reading (Fuhrman & Boroditsky, 2010), although it might be reshaped in particular experimental conditions (Casasanto & Bottini, 2014). In that sense, Parra's material anchor basically preserves the implicit left-to-right timeline in the Cartesian model for gravitational waves. Second, both the original diagram and the composer's way of interpreting it rely on static representations of motion, either physical or musical. The analogy, probably giving stability to the anchoring process, would be related to what Leonard Talmy (1996) defined as a *fictive motion*. This cognitive concept has helped to explain the ubiquity of the source-path-goal schema in mathematical thought (Lakoff & Núñez, 2000, pp. 37-79).

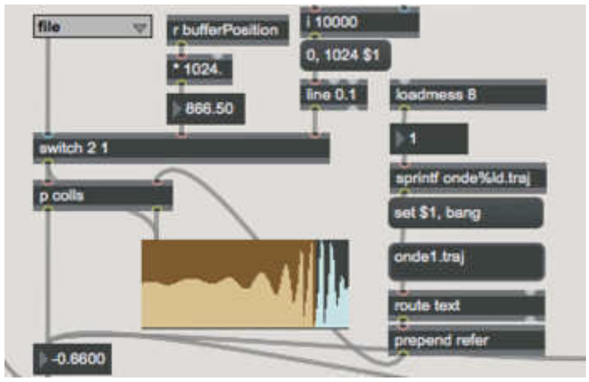


Figure 1. Partial screenshot of Goepfer's patch for implementing the gravitational waves into electronics. The boxed curve is based on the actual physical model (Abbott *et al*, 2016, p. 2).

Back to the *Inscape* project, Parra showed Goepfer his previous sketches for *Caressant l'horizon*, wondering how they could transpose this idea into the electronics of the new work. The existence of shared image schemas allowed Goepfer to understand Parra's starting point. Together, they reshaped the former material anchor: the vertical dimension became a control of live electronics sound processing, implemented in a `Max/MSP` patch (see Figure 1).

Group Interaction with Musical Objects

Extending the Blend: From Individual to Social Cognition

As implicitly remarked in the previous paragraph, sharing conceptual metaphors was a necessary condition that enabled the birth of the `Max/MSP` patch. Nevertheless, the CBT underlying the material anchor is merely a cognitive model for the individual mind. In that sense, unaltered uses of this theory with an eye to explain distributed musical practices (Stefanou, 2015) are rather controversial.

Some scholars have provided theoretical extensions of the CBT aimed at covering its pragmatic function within a communicative context. A quite neutral one is the addition of a grounding box for describing this interaction (Coulson & Oakley, 2005), which has already impacted the psychological study of music (Antović, 2018). A more complex model, based on six different spaces, has been also conceived (Brandt, P. A., 2004; Brandt, L. & Brandt, P. A., 2005). It adds a semiotic or base space featuring the communicative context, a relevance space, and one for meaning. The semiotic space would set the basis in which the relevance one would "filter" the relevant elements of the blend—the virtual space—, then project it into the meaning one. In particular, the relevance space would explain "why we presently see the thing in a particular way" (Brand, P. A., 2005, p. 83). Scholars have taken advantage of this six-spaces model for describing fictive interaction in speech and the arts (Brandt, L., 2008; Oakley & Brandt, P. A., 2008). Furthermore, Per Aage Brandt (2008) has used it for modeling the emergence of musical meaning, but with a strong bias for Western tonal music.

Metaphorical Thought and Enaction

As Shaun Gallagher (2017) stressed, "an enactivist account of perception highlights the integration of a variety of bodily factors into perception" (p. 41), pragmatically speaking. From this viewpoint, enactive approaches to social cognition have proven useful for understanding participatory sense-making (Di Jaegher & Di Paolo, 2007), including gestural matters (Cuffari, 2012). Nowadays, several scholars are combining enactivism with embodied theories of metaphor for a better understanding ***of*** the relationship between speech and gesture (Jensen & Cuffari, 2014), or the adaptive potential of virtual environments based on Human-Computer Interaction (HCI) (Gallagher & Lindgren, 2015, pp. 399-402).

The *Inscape* Project [2]: Enacting Cognitive Metaphors?

During the third meeting between Parra and Goepfer, the latter decided to connect the last version of his `Max/MSP` patch with a tablet. In doing so, the haptic device was allowed to reshape the original gravitational waves model for testing different sound results, based on an orchestral sample. This implementation was aimed at the recognition of new wave-like curves as an input depicted by the user's index finger.

Haptic controllers like touchscreens are spreading as a standard HCI for music (Altinsoy & Merchel, 2018). Although the topic of metaphors for HCI has led to controversial debates in the past, some engineers and computer scientists advocate for CMT-oriented paradigms during the conception and design of these kinds of interfaces (Blackwell 2006; Treglown, 2000), notably when the audio-haptic cross-modality is involved (Amni, 2014).

Considering the description of the `Max/MSP` patch as a material anchor, its extension via HCI admits a theoretical abstraction by means of the six-spaces variant of the CBT (see Figure 2). The following analysis is aimed at testing the pertinence of its relevance space through behavioural data.

Data. Both Parra and Luminet were filmed while using the tablet for interacting with the `Max/MSP` patch. The outcome of their actions—the electronic transformation of an orchestral sample (total timing: ca. 6 seconds)—was simultaneously diffused by the loudspeakers of the studio at IRCAM.

Potential bias. Parra took part in the patch development, while Luminet did not, although Goepfer carefully explained its functionality to him during their meeting. Parra's analysed

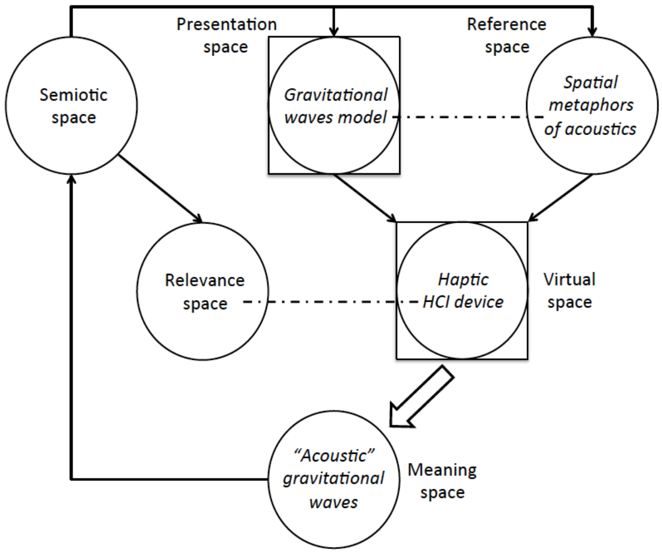


Figure 2. Adaptation of Per *Aage Brandt's* model (2004, p. 84) within the context of the *Inscape* project. Squared spaces highlight the physical support of material anchors, according to the notation by Edwin Hutchins (2005, p. 1557). Expressions in italics denote the content of the spaces.

video (see Figure 3) matches with the fourth time—of six—he interacted with the HCI device; Luminet's (see Figure 4) matches with his second—of two—interactions with it. These differences do not, however, carry a significant bias for a qualitative discussion of the results.

Method. The filmed actions were qualitatively analysed following three main behavioural criteria: the participants' multimodal interaction with aural and visual stimuli, the presence of bodily emphases while performing the action, and their hand gestures while using the touchscreen.

Results. Parra's and Luminet's interactions with the HCI were quite different, almost oppositional (see Table 2). The former was fully playing a metaphorical enaction within the environmental stimuli, even adding at the end an iconic—due to its concrete nature (McNeill, 1992, p. 145)—gesture that imitated an actual technique of musical performance. Conversely, Luminet was visually focused on the screen of the HCI artifact, and mainly worried with a geometric accuracy of the wave-like curve. Therefore, and somehow paraphrasing Susan Goldin-Meadow (2003), Parra truly enacted a "hearing" gesture—with respect to his environment—while Luminet performed, without aural enaction and mathematically speaking, an "analytical" one.

Discussion. The results seem to point towards a lack of agreement among the members of the team about the relevant features of the HCI device with respect to its underlying material anchor. As a matter of fact, the situation led to a short period of misunderstanding and a further debate among them.

Conclusion and Afterthought

Through the case study provided in this paper, different cognitive models have been tested, leading to disparate conclusions. First, CMT is suitable for understanding the basis of a communicative exchange in distributed creative contexts, but insufficient for deeply understanding the conception of complex musical ideations. Second, material anchors have proven relevant for explaining some creative choices in which

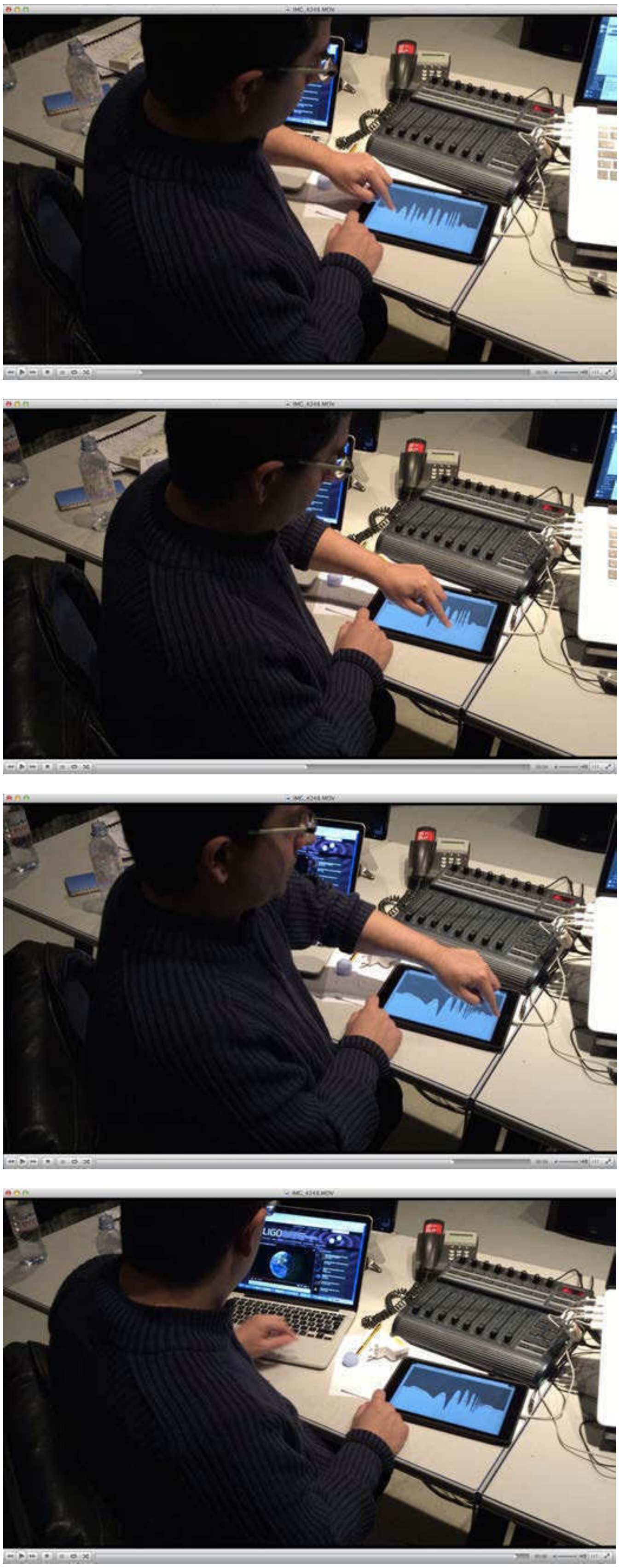


Figure 3(a, b, c & d). From top to bottom, four video stills of Parra playing with Goepfer's haptic HCI device, aimed at musically implementing wave-like curves in real time. Recording venue and date: IRCAM (Studio 6), 9 January 2017. Total timing: 8 seconds. Stills timing: (a) ca. 0:01; (b) ca. 0:04; (c) ca. 0:06; (d) ca. 0:08. © José L. Besada.

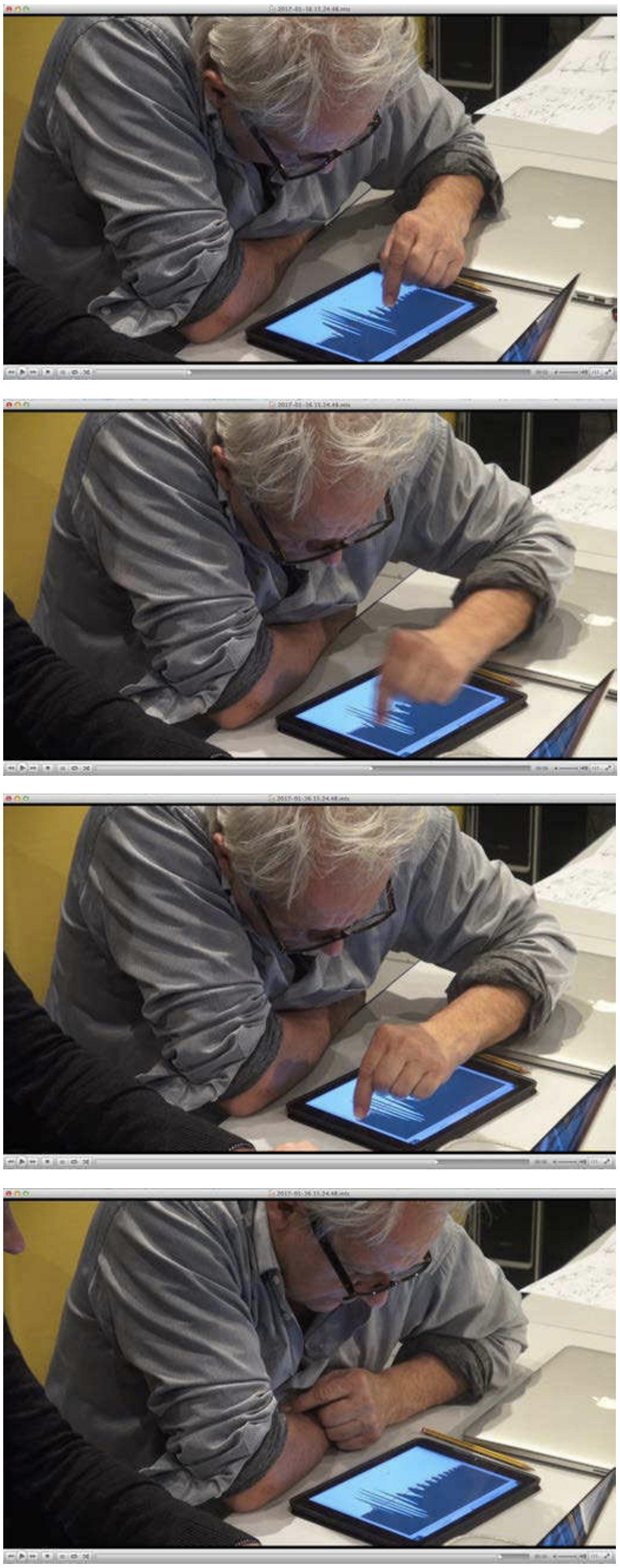


Figure 4(a, b, c & d). From top to bottom, four video stills of Luminet playing with Goepfer's haptic HCI device, aimed at musically implementing wave-like curves in real time. Recording venue and date: IRCAM (Studio 1), 16 January 2017. Total timing: 10 seconds. Stills timing: (a) ca. 0:02; (b) ca. 0:06; (c) ca. 0:08; (d) ca. 0:10. © José L. Besada.

Table 2. Qualitative comparison of the bodily behaviour (in terms of sight, hand gesture, and corporal position) by Parra and Luminet while they performed the wave-like curves on Goepfer's HCI device.

Behaviour	**Parra**	**Luminet**
Multimodal interaction with stimuli	The design of the wave-like curve is fully synchronised with the musical sample.	The design of the wave-like curve is unsynchronised with the musical sample; he starts after the sample sounds and ends during silence.
	His sight sometimes alternates among several visual stimuli beyond the HCI device.	His sight only focuses on the screen of the HCI device, before, during, and after the action.
	He bodily interacts with other devices. For instance, he punctually takes advantage of a laptop displaying a visual simulation of the gravitational-wave effect.	He exclusively interacts with the HCI device, before, during, and after the action.
Bodily emphases	He enhances the wave-like gesture carried by his hand and forearm by means of his shoulder and chest that he freely moves.	His torso remains quite static while performing the action, partially due to a bodily position in which the elbows are bearing his body mass.
Haptic design with the index finger	He performs a relatively free wave-like curve, depending on what he is simultaneously hearing as a consequence of his physical action.	He tries to scrupulously replicate the shape of the gravitational-wave model.
	At the end of the action, his finger rotates on a point of the HCI device without providing further information to the computer: it iconically imitates the *vibrato* on string instruments.	

material objects or representations are involved, in particular when HCI artifacts are present. Indeed, the touchscreen can be explained as a device allowing participants to transform a fictive—graphical—motion into a perceptible—musical—one via enaction. Third, the pertinence of the six-spaces extension of the CBT is put into question because the hypothesised relevant space may not capture a minimal relevance that should be shared by all the members of a creative team. Symbolic interactionism can provide a clue for better understanding this problem: although Herbert Blumer (1969) proposed that meanings arise from social interaction, he also highlighted that "human beings act toward things on the basis of the meanings that the things have for them" (p. 2). This is, ultimately, a task of the individual mind.

Last but not least, canonic research on distributed creativity in music has often emphasised the "positive" features of distributed creative actions (Sawyer, 2003, 2006). The above notwithstanding, the case study has shown how disparate interpretations of the HCI device led to a misunderstanding among several members of the team. Fortunately, current psychological research on musical creativity is taking into account this possibility, even more severe conflicts (Hill, M., Hill, B. & Walsh, 2018). Misunderstandings and conflict may also open the door to unexpected creative paths.

Acknowledgements. This paper has been published within the framework of the IdEx Unistra and benefits from funding from the state managed by the French National Research Agency as part of the « Investments for the future » scheme.

References

Abbott, B. P. *et al.* (2016). Observation of gravitational waves from a binary black hole merger. *Physical Review Letters, 116*. Retrieved from https://physics.aps.org/featured-article-pdf/10.1103/PhysRevLett.116.061102

Altinsoy, M. E., & Merchel, S. (2018). Touchscreens and musical interaction. In S. Papetti & C. Saitis (Eds.), *Musical Haptics* (pp. 239-256). Berlin: Springer.

Ammi, M. (2014). Intermodal audio-haptic metaphor: Improvement of target search in abstract environments. *International Journal of Human-Computer Interaction*, *30*(11), 921-933.

Antović, M. (2018). Schemas, grounds, meaning: On the emergence of musical concepts through conceptual blending. *Musicaœ Scientiœ*, *22*(1), 57-71.

Athanasopoulos, G., & Antović, M. (2018). Conceptual integration of sound and image: A model of perceptual modalities. *Musicaœ Scientiœ*, *22*(1), 72-87.

Bauer, A. (2004). "Tone-color, movement, changing harmonic planes": Cognition, constraints, and conceptual blends in modernist music. In A. Ashby (Ed.), *The Pleasure of Modernist Music: Listening, Meaning, Intention, Ideology* (pp. 121-152). Rochester: University of Rochester Press.

Besada, J. L. (2017). *Metamodels in Compositional Practices: The Case of Alberto Posadas's* Liturgia Fractal. Paris: IRCAM / Delatour France.

Besada, J. L. (2018). Enjeux cognitifs des métaphores physiques dans le processus créatif d'Hèctor Parra. *Cahiers du LabEx CAP*, *5*, 11-44.

Blackwell, A. F. (2006). The reification of metaphor as a design tool. *Transactions on Computer-Human Interaction*, *13*(4), 490-530.

Blumer, H. (1969) *Symbolic Interactionism: Perspective and Method*. Englewood Cliffs: Prentice-Hall.

Boyes Braem, P., & Bräm, T. (2000). A pilot study of the expressive gestures used by classical orchestral conductors. In K. Emmorey & H. Lane (Eds.), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima* (pp. 143-167). New Jersey: Lawrence Erlbaum Associates.

Brandt, L. (2008). A semiotic approach to fictive interaction as a representational strategy in communicative meaning construction. In T. Oakley & A. Hougaard, *Mental Spaces in Discourse and Interaction* (pp. 109-148). Amsterdam, Philadelphia: John Benjamins.

Brandt, L., & Brandt, P. A. (2005), Making sense of a blend: A cognitive-semiotic approach to metaphor. *Annual Review of Cognitive Linguistics*, *3*, 216-249.

Brandt, P. A. (2004). *Spaces, Domains, and Meanings: Essays in Cognitive Semantics*. Bern: Peter Lang.

Brandt, P. A. (2008). Music and the abstract mind, *The Journal of Musical Meaning*, *7*. Retrieved from http://www.musicandmeaning.net/issues/showArticle.php?artID=7.3

Brower, C. (1997-1998). Pathway, blockage and containment in *Density 21.5*. *Theory & Practice*, *22-23*, 35-54.

Brower, C. (2000). A cognitive theory of musical meaning. *Journal of Music Theory*, *44*(2), 323-379.

Casasanto, D., & Jasmin, K. (2012). The hands on time: Temporal gestures in English speakers. *Cognitive Linguistics*, *23*(4), 643-674.

Casasanto, D., & Bottini, R. (2014). Mirror reading can reverse the flow of time. *Journal of Experimental Psychology: General*, *143*(2), 473-479.

Chuang, Y-C. (2012). Gestures, language, and what they reveal about thought: A music teacher's use of metaphor in Taiwan. In F. MacArthur *et al.* (Eds.), *Metaphor in Use: Context, Culture and Communication* (pp. 261-284). Amsterdam, Philadelphia: John Benjamins.

Cienki, A. J., & Müller, C. (Eds.) (2008). *Metaphor and Gesture*. Amsterdam, Philadelphia: John Benjamins.

Coulson, S., & Oakley, T. (2005). Blending and coded meaning: Literal and figurative meaning in cognitive semantics. *Journal of pragmatics*, *37*, 1510-1536.

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicaæ Scientiæ*, *5*(2), 195-212.

Cox, A. (2016). *Music and Embodied Cognition: Listening, Moving, Feeling, and Thinking*. Bloomington: Indiana University Press.

Cuffari, E. (2012). Gestural sense-making: Hand gestures as intersubjective linguistic enactments. *Phenomenology and the Cognitive Sciences*, *11*(4), 599-622.

De Jaegher, H., & Di Paolo, E. (2007). Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, *6*(4), 485-507.

Eitan, Z., & Granot, R. Y. (2006). How music moves: Musical parameters and listeners' images of motion. *Music Perception*, *23*(3), 221–248.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science*, *22*(2), 133-187.

Fauconnier, G., & Turner, M. (2008). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities* (revised edition). New York: Basic Books.

Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive Science*, *34*(8), 1430-1451.

Gallagher, S. (2017). *Enactivist Interventions: Rethinking the Mind.* Oxford: Oxford University Press.

Gallagher, S., & Lindgren R. (2015). Enactive metaphors: Learning through full-body engagement. *Educational Psychology Review*, *27*, 391-404.

Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge MA: Harvard University Press.

Hayes, E. R. (2008). *The Pragmatics of Perception and Cognition in* MT Jeremiah 1:1-6:30: *A Cognitive Linguistics Approach*. Berlin: Walter de Gruyter.

Hill, M., Hill, B., & Walsh, R. (2018). Conflict in collaborative musical composition: A case study. *Psychology of Music*, *46*(2), 192-207.

Hutchins, E. (2005). Material anchors for conceptual blends. *Journal of Pragmatics*, *37*, 1555-1577.

Jensen, T. W., & Cuffari, E. (2014). Doubleness in experience: Towards a distributed enactive approach to metaphoricity. *Metaphor & Symbol*, *29*, 278-297.

Johnson, M., & Larson, S. (2003). "Something in the way she moves": Metaphors of musical motion. *Metaphor & Symbol*, *18*(2), 63-84.

Kendall, G. S. (2010). Meaning in electroacoustic music and the everyday mind. *Organised Sound*, *15*(1), 63-74.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Lakoff, G., & Núñez, R. E. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being.* New York, Basic Books.

Larson, S. (1977-1998). Musical forces and melodic patterns. *Theory & Practice*, *22-23*, 55-71.

Luminet. J.-P. (1993). Musique avec pulsar oblige: A propos du *Noir de l'Etoile* de Gérard Grisey. *Les Cahiers de l'Ircam*, *4*, 133-143.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, London: The Chicago University Press.

Oakley, T., & Brandt, P. A. (2009). Hypotyposis: Metarepresentation, mind-reading, and fictive interaction. In W. Wildgen & B. van Heusden (Eds.), *Metarepresentation, Self-Organization and Art* (pp. 115-136). Bern: Peter Lang.

Parra, H. (2008). *Hypermusic Prologue*: Opera projectiva en set plans. In C. Massip & J. Perelló (Eds.), *En ressonància: Encontre Noves Fronteres de la Ciència, l'Art i el Pensament* (pp. 51-63). Barcelona: Departament de Cultura de la Generalitat de Catalunya.

Pérez Sobrino, P., & Julich, N. (2014). Let's talk music: A corpus based account of musical motion. *Metaphor & Symbol*, *29*(4), 298-315.

Poggi, I. (2006). Body and mind in the pianist's performance. *Proceedings of the 9th International Conference on Music Perception and Cognition* (pp. 1044-1051). Italy: University of Bologna.

Saslaw, J. K. (1996). Forces, containers, and paths: The role of body-derived image schemas in the conceptualization of music. *Journal of Music Theory*, *40*(2). 217-243.

Sawyer, R. K, (2003). *Group Creativity: Music, Theater, Collaboration*. Mahwah: Lawrence Erlbaum.

Sawyer, R. K. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music*, *34*(2), 148–165.

Stefanou, D. (2015). Investigating social creativity and concept invention in collaborative musical situations. *Proceedings of the 1st International Conference on New Music Concepts* (pp. 73-78). Italy: University of Treviso.

Stefanou, D. (2018). The way we blend: Conceptual integration through intermedial and open-form scores. *Musicaæ Scientiæ*, *22*(1), 108-118.

Talmy, L. (1996). Fictive motion in language and "ception". In P. Bloom *et al.* (Eds.), *Language and Space* (pp. 211-276). Cambridge MA, London: MIT Press / Bradford.

Treglown M. (2000) Embodiment and interface metaphors: Comparing computer filing systems. In S. McDonald, Y. Waern, & G. Cockton (Eds.), *People and Computers XIV: Usability or Else!* (pp. 341-355). London: Springer.

Zbikowski, L. M. (2002). *Conceptualizing Music: Cognitive Structure, Theory, and Analysis*. New York: Oxford University Press.

Zbikowski (2011a). Musical gesture and musical grammar: A cognitive approach. In A. Gritten & E. King (Eds.), *New Perspectives on Music and Gesture* (pp. 83-98). Abingdon: Ashgate.

Zbikowski, L. M. (2011b). Music, language and kinds of consciousness. In D. Clarke & E. Clarke (Eds.), *Music and Consciousness: Philosophical, Psychological and Cultural Perspectives* (pp. 179-192). Oxford: Oxford University Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Application of Nonlinear Signal Processing Technique to Analyze the Brain Correlates of Happy and Sad Music Conditions During Listening to Raga Elaboration Phases of Indian Classical Music

Sushrutha Bharadwaj. M[1, 3], Shantala Hegde[2], D. Narayana Dutt[1], Anand Prem Rajan[3]

[1]*Department of Medical Electronics, Dayananda Sagar College of Engineering, India,* [2]*Music Cognition Laboratory, National Institute of Mental Health and Neurosciences, India,* [3]*School of Biosciences and Biotechnology, VIT University, India*

[1]sushrutha.bharadwaj@gmail.com, [2]shantala.hegde@gmail.com, [3]dndutt@gmail.com, [4]janandpremrajan@vit.ac.in

Abstract

Music is one of the major activities that alters the emotional experience of a person. Musical processing in the brain is a complex process involving coordination between various areas of the brain. There are less number of studies that focus on analyzing brain responses due to music using modern signal processing techniques. This research aims to apply a nonlinear signal processing technique i.e. the Recurrence Quantification Analysis (RQA) technique to analyze the brain correlates of happy and sad music conditions while listening to happy and sad ragas of North Indian Classical Music (NICM). EEG signals from 20 different subjects are acquired while listening to excerpts of raga elaboration phases of NICM. Along with behavioural ratings, the signals were analyzed using the Recurrence Quantification Analysis technique. The results showed significant differences in the recurrence plot and recurrence parameters extracted from the frontal and fronto-temporal regions in the right and left hemispheres of the brain. Therefore, from the results, it can be concluded that RQA parameters can detect emotional changes due to happy and sad music conditions.

Introduction

Biomedical signal processing is concerned with the extraction of significant information from biomedical signals. Brain is a complex system made up of many specialized areas working together. Non invasive techniques like EEG/fMRI are used to determine brain responses for various activities. One such activity is music that is known to induce strong emotional experience involving electrophysiological changes and changes in autonomic responses. Processing of music by the brain involves a high level coordination between various areas of the brain. There have been increased attempts by cognitive neuroscientists to study the role of music in the modulation of emotional experience and its neurobiological basis. Internally, the brain indulges itself in an array of cognitive functions like attention, learning, memory and decision making during automatic processing of music. Also, engaging the brain in musical behaviour has been proven to be an excellent indication of mental and emotional fitness and flexibility. Power spectral analysis of EEG has shown that the frontal and fronto-temporal regions are active during music processing and also it has been shown that there is asymmetry between the localizations of two hemispheres based on the variations in valence and intensity of music. Application of signal processing algorithms to investigate neuronal correlates of various features of music like timbre, rhythm etc., have shown better results than the conventional spectral analysis and behavioural studies. As brain is a nonlinear system, it can produce complex behaviours which cannot be analyzed effectively using linear methods. In order to analyze signals produced from nonlinear systems, nonlinear signal processing algorithms can be used in time, frequency or spatio-temporal domains. Therefore, based on the literature, it would be interesting to analyze the effects of music on the brain using advanced nonlinear signal processing algorithms.

The study of literature has shown that while many studies have been carried out to analyze the effects of music on brain and its responses, there are not many studies involving modern signal processing techniques. Moreover, the number of studies based on Indian classical music is significantly less and more so using modern signal processing techniques. This research aims to analyze the brain correlates of happy and sad musical emotions induced during listening to ragas of NICM by using a modern signal processing technique i.e. nonlinear processing of EEG signals.

Methods

Continuous EEG signals were recorded from 20 musically untrained subjects (10:10 - M:F) using a 32 channel Neuroscan system at the Music Cognition Laboratory, National Institute of Mental Health and Neurosciences, Bengaluru, India and were sampled at a rate of 1 KHz. The participants were made to listen to excerpts of raga elaboration phases of six ragas of North Indian/Hindustani classical music (NICM). The ragas were classified as happy and sad emotion eliciting ragas based on Indian music theory. Behavioural ratings from the participants were also recorded. The two sets of raga excerpts were significantly identified as happy and sad emotion excerpts. All ragas were played on the flute by the same artist. The tonic note for all the excerpts was same. Analysis was performed to compare the effects of happy and sad musical stimuli on same region of the brain and also to compare the effects of musical conditions across right and left hemispheres. Signals from frontal and fronto-temporal regions were considered for analysis. The concept of recurrence can be used to understand the dynamics of a nonlinear system. The recurrence plot, a graph representing the times at which a nonlinear system recurs to a former state, was used to investigate the m-dimensional phase space trajectory through a two dimensional representation of its recurrences. Recurrence Quantification Analysis (RQA) - a nonlinear signal processing method that quantifies the number and duration of recurrences - was applied on the signals recorded. RQA parameters such as recurrence rate, divergence, entropy, laminarity and determinism were extracted from the signals.

Results and Discussion

The results showed a significant difference in the recurrence plots and recurrence parameters extracted from right and left frontal regions of the brain while listening to happy and sad musical excerpts (Fig. 1 and Fig. 2). Specifically, the results show an increase in the average entropy in the left frontal region during happy musical conditions. In the right frontal region, there was an increase in average entropy and decrease in average divergence during sad musical conditions. Also, during sad musical stimuli, there was a significant decrease in average divergence and a similar increase in average entropy in the right hemisphere when compared to the left hemisphere in both frontal and fronto-temporal regions (Fig. 2). These results endorse the fact that the emotions elicited by happy musical conditions are processed in the left frontal regions of the brain. Also, the significant changes in divergence and entropy in the frontal and fronto-temporal regions indicate active emotional processing in those regions of the brain.

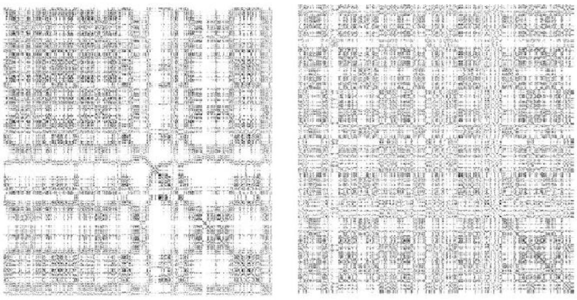


Figure 1: Recurrence plot - Happy and Sad frontal regions

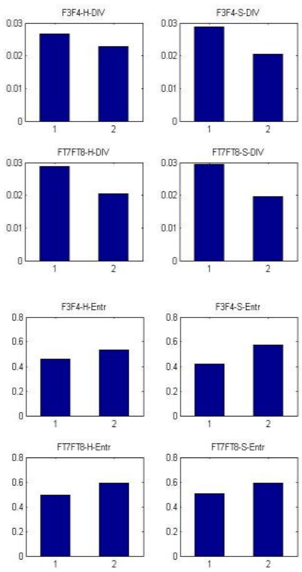


Figure 2: Hemispherical differences: Average Divergence and Average Entropy due to happy and sad music. 1 - Left, 2 - Right regions

Conclusion

This paper is concerned with determining the brain correlates of listening to happy and sad music, particularly listening to the ragas of Indian classical music. Nonlinear signal processing technique has been used instead of the conventional linear techniques. We expect nonlinear processing of EEG to be more appropriate since brain is inherently a nonlinear biological system. The results clearly indicate that the emotions induced by happy and sad music stimuli are processed differently by the brain in different regions. The left frontal region is found to be active in processing happy emotions and right frontal region is active for sad conditions. The results also show that the frontal regions are active in overall emotional processing as compared to other regions of the brain. Therefore, in future, it may be useful to extend the study for analyzing the effects of NICM on brain networks to determine connectivity as music processing involves coordination between different regions of the brain.

References

Scherer. K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? Journal of new music research. 33:1-13

Sluming. V. A, Manning. J. T. (2000). Second to Fourth digit radio in elite musicians: Evidence for musical ability as an honest signal of male fitness. Evol. Hum. Behav. 21:19

Louis. A. Schmidt, Laurel. J. Trainor. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. Cognition and Enotion. 15(4). 487-500.

E. Altenmuller, K. Schurmann. V. K. Lim, D. Parlitz. (2002). Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralization patterns. Neuropsychologia. 40. 2242-2256.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Sight-Reading Strategies and Personality Dimensions

Blanka Bogunović

Faculty of Music, University of Arts in Belgrade, Serbia

bbogunovic@rcub.bg.ac.rs

Abstract

The framework of the self-regulative learning theory was used in order to find out about the metacognitive strategies used in sight-reading while playing piano. We are of the opinion that, besides the confirmed role of cognitive, perceptive and motor abilities in acquiring the sight-reading skills, the non-cognitive factors such as personality and motivation play an important role. Therefore, we intended: a) to identify the metacognitive strategies used at each phase of self-regulated sight reading, and b) to determine the relations of the sight-reading factors with personality dimensions. The sample consisted of 95 music students of the Faculty of Music in Belgrade. The participants filled in the Sight-reading questionnaire and the NEO-PI-R. The results show that there are various factors in the subsequent phases of the sight-reading process: while Preparing (Analytical, Inner hearing, General overview), during Setting goals (to Play through, Technical accuracy, No goals), during Performance (Expertness, Non-perfectionism, Visualization and Continuity control), Problem solving, Monitoring and Self-reflection/evaluation. It seems that personality dimensions are mostly related to the sight-reading factors of Preparation and Performance. Namely, students who tend to have the Analytical approach in preparing for the sight-reading tasks showed higher Emotional stability, Agreeableness and Conscientiousness, especially higher Competence, Dutifulness, Achievement striving, Discipline and Deliberate thinking. The factor of Expertness in performing is significantly related to the facets such as Dutifulness and Achievement striving. These findings point out to the role of motivational factors in reaching high levels of expertness in performance and in the cognitive control of behavior. In conclusion, we could say that in performing sight-reading skills various metacognitive strategies are important for task execution, especially those that are related to the perceptual auditory mental representations and cognitive processes. As to relations between sight-reading and personal factors, the findings suggest that the non-cognitive factors have a certain role in the self-regulated performance of sight-reading skills. Self-efficacy and satisfaction with the sight-reading skill have to do with personality attributes.

Introduction

The framework of the self-regulative learning theory was used in order to find out about the metacognitive strategies used in sight-reading while playing a musical instrument. The existing findings confirm the role of cognitive, perceptive and motor abilities in acquiring the sight-reading skills, as well as the role of practice and education-related factors (Kopiez & Lee, 2006, 2008). There is not much research that explains the role of the non-cognitive factors, especially personality attributes as underlying influential factors of efficiency in sight-reading. Hence, we tried to make a step forward in this field by defining an adapted paradigm, namely, self-regulation in the performance of a sight-reading musical task, and trying to specify the metacognitive strategies which are used in each phase of the process, as well as by exploring their relations with personality attributes.

Sight-Reading

Sight reading is an "online" activity that asks for a quick insight into the whole and/or parts of a music piece, with the task of maintaining the fluency and accuracy of performing without interrupting the music stream. Sight-reading is a complex problem-solving situation, with an intricate interplay of the "bottom-up" mechanisms (driven by the input stimulus of the score and auditory feedback) and the "top-down" processes (driven by expectations and cognitions) (Lehmann & Kopiez, 2008). In the process of sight-reading, memory plays a significant role considering the reconstruction of knowledge and previous experience. Namely, long-term memory is important for holding information on procedural knowledge ("knowing how" to do something), semantic knowledge (facts, e.g. "knowing that") and episodic memory (the details of events or episodes of one's life) (Ginsborg, 2006).

A high level of sight-reading pertains to expert thinking that is characterized by a prompt execution of tasks, by perceiving meaningful units and the underlying structure and using their operative and long-term memory more efficiently. The experts grasp the problem on a deeper level by noticing the relevant aspects of the phenomena; they observe the problem from different angles before they opt for a solution. They have the ability to monitor their activities and, therefore, correct themselves during the process of executing certain operations (Kostić, 2010).

Therefore, an expert in sight-reading, regardless of whether it is playing or singing, needs a developed auditory mental representation of the music from examination of the score only – inner hearing. Performers need to review or scan the whole piece before starting to play, even to briefly mentally rehearse the difficult parts. Competent sight-reading depends on the ability to identify familiar patterns and to spend time evaluating the musical material before beginning to perform (McPherson & Zimmerman, 2002; Radoš, 2010). It is confirmed that the sight-singers with a higher level of self-assessment skills have better insight into the structure of a melodic task and rely on inner hearing (Bogunović & Vujović, 2012).

Self-Regulated Performing

Self-regulation is a cyclical process because feedback obtained from prior performance helps the learners to adjust their performance and future efforts (McPherson & Zimmerman, 2002). This process activates metacognition, motivation and behavior (Zimmerman, 1986). The essence of the self-regulative process is present in the activity of sight-reading in musical performing. An intense intertwining of metacognition, psychomotor behavior, evaluation and adaptation enables an almost simultaneous performance of all cyclical phases of the self-regulation activity. Unlike practicing to play a musical instrument, where the cyclical aspect of the Self-regulated learning happens subsequently, in sight-reading the process takes place only once. Therefore, the strategies to achieve the goal in these musical activities are different, as well as the goal. There are no studies known to us that explore sight-reading as a self-regulated activity.

We used the paradigm of the self-regulated learning cycle phases (McPherson & Zimmerman, 2002) as a basis for the theoretical background of our research in metacognitive strategies in Self-regulated performing (Forethought, Performance/volitional control and Self-reflection phase). The theoretical framework of this research has been adapted to a certain extent and it refers to: Forethought (Preparation and Setting goals) – Performance, Problem solving, Monitoring – Self-reflection (Self-efficacy and Self-satisfaction in sight-reading). Further, we were interested in the metacognitive strategies used in each phase of the Self-regulated performing because they are crucial for expert sight-reading.

What is the role of metacognitive strategies used in each phase of the self-regulated activity? There are data that lead to the conclusion that metacognition is the most important predictor of learning performance, even that it accounts for 40% of variance in learning outcomes. The regulation of cognition refers to the metacognitive skills for control over one's strategy use, that is, to planning, monitoring, and evaluation (Veenman, 2011). Similarly to this, Jorgensen speaks about metastrategies that a musician must have, namely, the thorough knowledge of the repertoire of strategies that any musician must have and must be able to control, regulate and exploit (Jorgensen, 2006). Moreover, music learners are involved in ongoing self-regulatory cognition/metacognition. Certain authors hypothesized that specific goal setting, strategic planning and self-efficacy were the key self-regulatory components that optimize the music students' instrumental practice, progress, and self-regulatory continuity (Hatfield, Halvari & Lemyre, 2017).

Self-efficacy, as well as self-satisfaction, is an important part of self-evaluation while the task is performed or executed. A sense of self-efficacy or belief in one's ability to accomplish a specific task motivates individuals to persist despite setbacks, to become more actively involved in a task, and to work harder and longer toward attainment (Bandura, 1997, 2012). Furthermore, it can positively influence persistence, self-regulation, and subsequent task-based achievement (Hendricks, 2016).

Personality and Performance Skills

There are not many research studies that explore the relationship between personality traits, measured by a personality inventory, and musical skills which concern vocal-instrumental performance. There are findings of Kemp (1996), who spoke about higher anxiety in a group of highly gifted adolescent performers and presumably with higher performance skills. In the Serbian sample of adolescent musicians, it was confirmed that musically and academically more successful students were more sensitive and conscientious and had better control of emotions and behavior, which testifies to stability and perseverance, providing a strong basis for the realization of the artistic potential and the attainment of high-set goals. More successful young musicians are also characterized by a certain dose of desurgency, which, in Kemp's study, appears within the framework of a set of introversion factors (Bogunović, 2010).

There are more research findings that cover the motivational part of personality. Many of them speak about higher inner motivation of musicians, perseverance and mastery orientation. The ties with expertise are supported by the Self-regulated learning theory and it is axiomatic that the self-regulating musicians persist when confronted with challenges. There is a strong association between the intrinsic interest and self-regulation instruction. Persistence was confirmed as a personality/motivational trait that is characteristic for experts and for those who like challenges (Varela, Abrami & Upitis, 2016).

Also, self-efficacy in music, namely the perception of success and failure, has been found to be related to personality attributes (Bogunović & Bodroža, 2015). The research that was conducted at the level of higher music education on students showed that music students with higher neuroticism perceive themselves as less successful in music performing. A closer examination of this result showed that only the facet of vulnerability to stress contributed to the self-perception of music success, implying that perceiving oneself as being less successful is probably due to the lack of self-esteem. The same study gave results indicating clearly that high music achievements in performance significantly correlate with femininity in male students, while any kind of sex-type role does not correlate with music success of female students (Bogunović & Bodroža, 2015).

In our previous research concerning the mindset and its role in the development of performing skills of music students, we obtained the results that support the importance of the self-regulative processes in gaining the skills necessary for musical excellence. The findings set off the role of self-regulation and cognitive and emotional control for high mastery orientation of students. It turned out that those who had better mastery orientation had better self-regulation, namely, better planning, time management and effort investment and a more realistic assessment. They were more efficient in stress reduction and had higher concentration and memory (Bogunović, 2017).

Research Methodology

The aims of the research were the following: a) to identify the metacognitive strategies used by music students at each phase of the self-regulated sight-reading process (preparation, setting goals, performance, problem solving, monitoring, and self-reflection/evaluation), and b) to determine the relations of sight-reading factors with personality dimensions and facets.

The sample consisted of 95 music students, 25 males and 70 females, all playing the piano as their major or minor instrument, who had 12 to 15 years of specialized music education experience. The three-level specialized music education (elementary music school, music high school and the faculty of music) starts at the age of 5 to 7 and offers systematic tuition in instrumental playing, music theory and general subjects for the musically gifted children and youth (Nogaj & Bogunović, 2015). Therefore, participants could be designated as young experts, considering the "10 years rule". The course of sight-reading and playing, as well as sight singing (in the framework of solfeggio lessons), was introduced at the secondary level of education and has been constantly present in the curricula onwards.

The participants filled in the Sight-reading questionnaire (12 questions) constructed for the purposes of the current research. For each subsequent phase of the self-regulated process of sight-reading, the list of strategies was given and the participants rated the frequency of using each strategy on a five-point Likert scale. The strategies in each phase were extracted from psychological and music methodological literature and students' experience. The participants were asked to evaluate the level of their self-efficacy. The Revised NEO Personality Inventory (NEO PI-R) was used to measure the personality dimensions of the musicians. The participants rated 240 items on a scale from 1 to 5. The items load onto the Big-Five personality dimensions (Neuroticism, Extroversion, Openness, Agreeableness and Conscientiousness), each of which contains six facets (Costa & McCrae, 1995).

Six Exploratory factor analysis were performed using the Oblimin rotation method with Kaiser Normalization. The extracted factors were correlated with personality dimensions and facets (Pearson's correlation coefficient).

Results

Sight-reading and Self-regulated Performing

Preparation. The Principal component analysis for the Preparation phase of the sight-reading process extracted three factors and explained 45.55% of variance. The extracted factors were Analysis, Inner hearing and General overview (Table 1). The content of the first factor, Analysis, shows the strategies that refer to procedural knowledge, namely, "knowing how" to do something. The students apply the strategies that are mostly learned through the application of the procedures transferred to them by their teachers. At the same time, these strategies are quite useful when an understanding of the structure and density of a musical piece is a target of the quick insight into new composition. The second factor, Inner hearing, is considered to be rather valuable for the sight-reading performance, as well as for other music skills. It indicates the presence of auditive mental representation, which enhances an immediate understanding of the musical structure. The General overview factor is rather formal and informative, based on "top down" strategies, applied in order to accomplish the formation of the "big picture".

Setting goals. Concerning the Setting goals phase, when the planning of "how to do" goes on, the following factors were extracted, accounting for 73.68% of total variance: To Play through (40.55%), Technical accuracy (19.76%) and No goals (13.37%). We could say that To Play through factor represents the strategies that are in concordance with the core of the sight-reading process. The opposite is represented by the second factor, Technical accuracy, which reduces the sight-reading process to the motoric aspect of performance and highly values precision. The third factor is represented with only one strategy and that is No goals. It is not quite clear what it means. Does it represent a highly automatized skill at the level of expertness when no cognitive or motor control is needed or is it a trial and error attitude?

Table 1. The factor structure matrix of the Preparation phase – Oblimin rotation method.

Preparation phase – Strategies of piano player	Components		
	Analytical	**Inner hearing**	**General overview**
Starting to play immediately, trusting intuition	**-.44**	-.13	.13
Liking to get the instruction from a teacher the most	**.55**	.19	-.06
Running through the text from the beginning to the end	**.46**	.13	.32
Analyzing the form/structure/parts of composition	**.72**	-.04	.33
Determining the measure/meter and tonality	**.66**	.018	.31
Identifying first the tempo	.20	.30	**.66**
Paying attention to rhythmical figures	**.45**	.29	**.42**
Playing "silently" with both hands (using inner hearing)	.22	**.93**	.07
Playing "silently" the most difficult parts	.25	**.93**	.10
Browsing first one, than the second hand	**.58**	.25	-.02
Browsing both hands at the same time	**.50**	.32	-.15
First paying attention to the instructions in a music text (dynamic, articulation, fingering, pedal)	.23	.06	**.66**
At the first glance paying attention to the music style/composer/period of composition	.22	-.20	**.67**
While browsing, hearing music with the inner ear	-.05	**.41**	**.47**
Perceiving the model of harmony	**.43**	.01	.29
Perceiving sequential movements	**.52**	.01	.33
Analyzing the tonality changes	**.65**	-.03	.24
Trying to identify the character of a piece	.15	.01	**.71**
Total variance explained 45.55%	**23.85%**	**11.37%**	**10.33%**

Performance. The factor analysis of the strategies in the Performance phase of sight-reading yielded four factors, which accounted for 51.37% of variance (Table 2). These were: Expertness, Non-perfectionism, Visualization and Continuity control. Expertness is the most saturated factor and contains the strategies that are the essence of sight-reading: Paying attention to expressiveness, taking into account the signs for expressive performance during the first playing, a dynamic, appropriate tempo and "foreseeing" the music flow that follows, as well as "seeing the big picture". Besides this one, the Non-perfectionism factor is the important factor in pursuing the musical flow while a composition is played for the first time. It means that a performer is able to follow the main stream of music, select the most important lines, ignore mistakes and foresee what comes next. The least saturated, but very important as to content, is the fourth factor, Continuity control, which is crucial for any performance, especially for sight-reading, as an "online" activity.

Table 2. The factor structure matrix of the Performance phase – Oblimin rotation method.

	Components			
Performance phase – Strategies of a piano player	**Expertness**	**Non-Perfectionism**	**Visualization**	**Continuity control**
Watching the music text forth all the time	.39	-.27	-.08	**.54**
Trying to watch the music text and not hands	-.13	.17	.05	**.65**
Playing without correcting the mistakes	.19	**.48**	.36	**.53**
Relying on musical intuition	**.47**	-.03	**-.67**	.23
Paying attention to rhythmical figures	**.52**	-.31	-.14	.18
Beware of the expressiveness of playing	**.82**	-.10	-.21	.06
Beware of dynamic	**.83**	-.11	-.26	.05
Beware of playing in an appropriate tempo	**.64**	.14	.20	.11
Taking care not to play the wrong notes	.23	**-.65**	-.15	.13
While playing, singing in my thoughts	.27	.11	-.13	.08
Beware of holding on the same pulse	.17	-.27	-.11	**.56**
Mostly relying on hearing	.12	-.02	-.73	.19
Mostly relying on what I see, not paying much attention to sound	-.27	.35	**.46**	.18
Simplifying the composition by omitting “needless” tones	.39	**.48**	.24	.17
While playing I can “foresee” the music flow that follows	**.60**	-.05	-.22	.19
Taking into account the signs for expressive performance during the first playing	**.83**	-.07	-.08	.07
Trying to play all the tones in chords	.04	**-.82**	.02	.04
Trying to play the given fingering	-.09	**-.65**	.28	.07
When closing my eyes, I can see the board	.05	-.15	**.57**	.39
Total variance explained 51.37%	**21.61%**	**12.64%**	**9.85%**	**7.27%**

Problem solving. While playing, students encounter certain situations which demand quick Problem solving and decision making in order to continue the sight-reading activity. The factor analysis of the problem solving strategies showed two independent factors with 67.24% total variance explained. Namely, students tend to employ either the Corrective or Inefficient strategies. Corrective strategies (47.30%) imply constructive mastery-oriented actions, mainly the “trial-error-correction” pattern, but also the analytical ones (Stopping, correcting the mistake; Stopping and paying attention; Stopping in order to think over where and why the mistake has been made). It seems that problem solving, which asks for consideration and analytical cognitive intervention, has a side effect of stopping of playing, which is in contradiction to the nature of the sight-reading process. This would mean that the efficient action in problem solving is not to analyze (which implies the stopping of an action in our group of participants), but, parallel to playing, to select the important parts of the musical flow, while watching and perceiving what comes next. The factor of Inefficient behavioral strategies (19.934%) is basically typical for beginners (giving up or starting again from the beginning), but it seems that it is not present to such a low extent among students.

Monitoring. The phase of monitoring the process of task execution essentially implies facing the difficulties and ease in the performing phase or during the whole process of sight-reading. It also speaks about the weak and strong sides of students when they are faced with a new musical text. The results explain three factors pointing out to the sources of the difficulties that the students have when sight-reading is in question. These are the Harmonic and structural complexity of musical piece and Hard readability of music notation, as well as the difficulties that stem from Rhythm and tempo (Table 3). We could say that the fluency of sight-reading is influenced clearly by the specifics of the musical composition, starting from the complexity of musical structure and harmony, which pose difficulties for the quick grasping of the musical thought flow and the selection of the main musical lines. Furthermore, when a music text is loaded with information there are problems to take hold of the rhythm and tempo density and with many auxiliary signs written. The surplus (extra given) of information also disturbs the selection of the relevant material, as well as too fast tempo (motoric problems) or too slow (metrical problems). Indirectly, the findings indicate that the problem could be, depending on the performer’s capability, a weaker or slowly psychomotor response.

Table 3. The factor structure matrix of the Monitoring phase – Oblimin rotation method.

	Components		
Monitoring phase – difficulties in playing	**Harmonic and structural complexity**	**Hard readability of music**	**Rhythm and tempo**
Composition has more than two accidentals	-.06	**.71**	.280
Tempo is fast	.28	**.79**	.01
Text is dense (too many small notes)	.32	**.80**	.32
Lot of leaps	**.42**	**.59**	.22
Tempo is slow (problems with counting)	.17	.15	**.85**
Rhythm and meter are complex	.33	**.49**	**.75**
Many ornaments	**.56**	.30	**.41**
Fingering is written already	**.40**	.13	**.40**
Structure of the piece is not clear	**.71**	.18	.31
Many modulations	**.85**	.12	.08
Many alterations	**.76**	**.40**	.09
When parts for other instruments are written too	**.69**	.204	.22
Total variance explained 56.86%	**33.27%**	**13.91%**	**9.67%**

When students were asked what was easy for them, i.e. what kind of music texture made the sight-reading easy, the answers were exactly the opposite. Three factors were extracted (total variance explained 69.155%): Transparent musical structure (39.61%), Chunking of melody and rhythm (17.08%) and Simple harmony (12.47%).

Self-efficacy and self-satisfaction in sight-reading. Students’ long experience in sight-reading enables self-evaluation in skill expertness. Music students estimated their skill on the scale from 1 to 5 (M=3.20; SD=.97), perhaps lower than it would be expected after 12 to 14 years of systematic music tuition. The correlations between self-efficacy and the strategies used during the process of sight-reading have showed that students who have higher self-efficacy in sight-reading significantly more often prepare themselves for playing by getting a General overview of a composition (r=.26, p<.005), and, while performing, they tend to play with Expertness (r=.40, p<.001) and with lower Non-perfectionism (r=-.30, p<.001). Concerning the observed difficulties and ease in sight-reading, those who estimate their skill higher have fewer problems with Hard readability of music notation (r=-.38, p<001); they perceive easily Melodic and rhythmical chunks (r=.33, p<001) and do not take Simple harmony as a challenge (r=-.24, p<.005). We see that those students who perceive their skills in sight-reading higher have a generally "bigger" picture of the composition, and their process of performing is organized "top-down"; namely, it is driven by expectations of musical flow and cognitive functions. They perceive musical patterns better and tend to have control over an activity.

The relationship between self-satisfaction and the sight-reading factors speaks in favor of the theoretical notion about mastery orientation as self-motivational. Namely, the students who perceive their metacognitive strategies as successful are more satisfied and probably more motivated to play and enjoy. Namely, the results show that students who are more satisfied and enjoy their playing more often use To Play through strategy as a goal (r=.22, p<.037). They play with Expertness (r=.23, p<.027), have a good Continuity control (r=.22, p<.029), and, for them, Chunking of melody and rhythm is easy (r=.21, p<.038).

Sight-reading Factors and Personality

The correlations between the sight-reading factors and personality dimensions have showed some significant relatedness, especially with the sight-reading factors of Preparation and Performance. Therefore, we looked a bit closer into the correlations with facets. There are no statistically significant correlations between personality and the factors of Problem solving and Monitoring phases.

Preparation. The most striking is the Analytical factor during the preparation for sight-reading. Those students that have the Analytical approach are more Emotionally stable (r=-.31, p<.05) and have higher scores on Agreeableness (r=.48, p<.05) and Conscientiousness (r=.39, p<.005). They are also less Impulsive (r=-.38, p<.001) and have better control of Excitement (r=-.39, p<.001). The students who have the Analytical approach are significantly higher on almost all facets of Conscientiousness. They have higher striving towards Competence (r=.33, p<.005) and Achievement (r=.28, p<.005); they express Dutifulness (r=.23, p<.005), Discipline (r=.30, p<.005) and Deliberate thinking (r=.39, p<.005).

Setting goals. The results indicate that setting goals at the beginning of sight-reading has to do with striving for Achievement. There are statistically significant correlations with the goal To Play through (r=.31, p<.005) and No goals (r=.31, p<.005). This would mean that music students who want to achieve higher decide to play from the beginning till the end, with dynamic, tempo and expression, as well as with an intention to interpret the composition on the spot. Even if they do not define goals, those who play without much deliberate thinking strive towards Achievement in playing. Those students who have Technical accuracy as a goal do not show Esthetic sensitivity (r=-.28, p<.005).

Performance. Higher Expertness in performing is significantly related to Agreeableness (r=.30, p<005) and Conscientiousness (r=.28, p<.005), especially to its facets such as Dutifulness (r=.33, p<.005) and Achievement Striving (r=.27, p<.005). It is interesting that, in the performance phase, students who have higher Continuity control also have higher Extraversion (r=.30, p<.005), Positive emotions (r=.29, p<.005) and Achievement striving (r=.26, p<.005). The students who have a Non-perfectionist tendency are more oriented towards Order (r=.29, p<.005) and Achievement (r=.26, p<.005). The most prominent facet is Achievement striving, and this finding clearly marks the role of motivational aspects also concerning sight-reading, which is primarily a perceptive and cognitive musical skill.

Self-efficacy and self-satisfaction in sight-reading and personality. There are not many correlations between the personality and evaluation measures. There are only two between self-efficacy in sight-reading and personality traits but those which are there show some inner logic. Namely, music students who perceive themselves as efficient have higher Assertiveness (r=.35, p<.005) and Activity (r=.31, p<.005). Additionally, those who are more Extraverted feel higher Self-satisfaction in sight-reading (r=.26, p<.41) and are less Deliberate in thinking (r=-.27, p<.33).

Conclusion

We could say that the theoretical concept of self-regulated performing is adequate to differentiate and understand the metacognitive strategies employed in the phases of the sight-reading process. Efficient sight-reading skills, integrated with a strong artistic component, are clearly related to the metacognitive strategies that regulate cognition and are important for the task execution at each stage of the process. The main ones are those that are intertwined: perception and awareness of the musical structure elements, inner hearing, setting music expressive and interpretative goals, mastering the harmonic and structural complexity and the adaptive functions in executing task. Auditory mental representations of the musical stream and hierarchical musical structure have a significant role, as well as the expression of musical continuity. We can see that those students who perceive their skills in sight-reading higher have a generally "bigger" picture of the composition, and their process of performing is organized "top-down"; namely, it is driven by expectations of musical flow and cognitive functions. They perceive musical patterns better and tend to have control over activity.

Regarding the relations between sight-reading and personal factors, the findings suggest that an intertwining pattern of personality and motivational attributes has a certain role in the self-regulated performance of sight-reading skills. These results support the assumption that the non-cognitive factors develop the conditions for efficiency in sight-reading. Self-perception and satisfaction with the sight-reading skill have to do with personality attributes. The sight-reading strategies that are most efficient are related to the supportive personality factors. These results support the idea that the non-cognitive factors enable efficiency in sight-reading. There are place for education implications with respect to enhancing the efficient strategies of sight-reading and supporting the development of self-esteem and personality attributes that are linked to expertness. The limitation of the study refers to a relatively small sample on which complex analyses were performed. Therefore, the results should be a subject of verification.

Acknowledgements. The author of this article has been supported with the SEMPRE award for ESCOM 2018 participants. This research was carried out as a part of the scientific project *Identification, measurement and development of the cognitive and emotional competencies important for a Europe-oriented society* (No. 179018), supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia.

References

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, *38*, 9–44.

Bogunović, B. & Bodroža, B. (2015). Gender identity and personality dimensions as correlates of music performance success. In J. Ginsborg, A. Lamont, M. Phillips & S. Bramley (Eds.), *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music* (pp. 220-225). Manchester: Royal Northern College of Music.

Bogunović, B. & Vujović, I. (2012). Metacognitive strategies in learning sight-singing. *Psihološka istraživanja/Psychological research*, *15*, 115-133.

Bogunović, B. (2017). Mentalni sklop i postignuća studenata muzike [Mindset and music students' achievement]. In M. Petrović (Ed.), *Proceedings book of the Pedagogical forum of performing arts/Zbornik radova pedagoškog foruma scenskih umetnosti* (pp.160-173). Belgrade: Faculty of Music/Beograd: Fakultet muzičke umetnosti.

Bogunović, B. (2010). *Muzički talenat i uspešnost* [*Musical talent and successfulness*] (2nd ed.). Belgrade: Faculty of Music and Institute for Educational Research/Beograd: Fakultet muzičke umetnosti i Institut za pedagoška istraživanja.

Costa, P.T.Jr. & McCrae, R.R. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO Personality Inventory. *Journal of Personality Assessment*, *64*, 21-50.

Ginsborg, J. (2006). Strategies for memorizing music. In A. Williamon (Ed.), *Musical excellence. Strategies and techniques to enhance performance* (pp. 123–143). New York: Oxford University Press.

Hatfield, J.L., Halvari, H. & Lemyre, P-N. (2017). Instrumental practice in contemporary music academy: A three-phase cycle of Self-regulated learning in music students. *Musicae Scientiae*, *21*, 316-337.

Hendricks, K.S. (2016). The sources of self-efficacy: Educational research and implications for music. *National Association for Music Education*, *35*, 32-39.

Jorgensen, H. (2004). Strategies for individual practice. In A. Williamon (Ed.), *Musical excellence: strategies and techniques to enhance performance* (pp.85–104). Oxford: Oxford University Press.

Kemp, A. (1996). *The musical temperament: Psychology and personality of musicians*. New York: Oxford University Press.

Kopiez, R. & Lee, J. I. (2008). Towards a general model of skills involved in sight reading in music. *Music Education Research*, *10*, 41-62.

Kopiez, R. & Lee, J.I. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research*, *8*, 97-120.

Kostić, A. (2010). *Kognitivna psihologija* [*Cognitive psychology*]. Belgrade: Institute for Handbook Publishing/Beograd: Zavod za udžbenike.

Lehmann, A.C. & Kopiez, R. (2008). Sight-reading. In S. Hallam, I. Cross & M. Thaut (Eds.), *Oxford handbook of music psychology* (pp. 344-351). Oxford: Oxford University Press.

McPherson, G.E. & Zimmerman, B.J. (2002). Self-regulation of musical learning: A social cognitive perspective. In R. Colwell & C. Richardson (Eds.), *The new handbook of research in research in music teaching and learning* (pp. 327-347). Oxford: Oxford University Press.

Nogaj, A. & Bogunović, B. (2015). The development of giftedness within the three-level system of music education in Poland and Serbia: Outcomes of different stages. *Journal of the Institute of Educational Research/Zbornik Instituta za pedagoška istraživanja*, 47, 153-174.

Radoš, K. (2010). *Psihologija muzike* [*Psychology of music*] (2nd ed.). Belgrade: Institute for Handbook Publishing/Beograd: Zavod za udžbenike.

Varela, W., Abrami, P.C & Upitis, R. (2016). Self-regulation and music learning: A systematic review. *Psychology of Music*, *44*, 55-74.

Veenman, M.V.J. (2011). Learning to self-monitor and self-regulate. In R.E. Mayer & P.A. Alexander (Eds.), *Handbook of research on learning and instruction* (197-218). New York: Routledge.

Zimmerman, B.J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, *81*, 320-339.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Comparing the Cognition of Abstract and Representational Structures in Electronic Music

Anıl Çamcı[1] and Elif Özcan[2]

[1] *Department of Performing Arts Technology, University of Michigan, United States*
[2] *Industrial Design Department, Delft University of Technology, The Netherlands*

[1]acamci@umich.edu, [2]e.ozcan@tudelft.nl

Abstract

In this paper, we compare the cognitive processing of abstract and representational structures in electronic music by discussing the results of a listening experiment conducted to gather both in-the-moment and after-the-fact accounts of a listener's experience with a work of electronic music. We focus on two works that, while sharing various structural similarities, adopt significantly different sonic vocabularies. With this study, we aim to understand the type of conceptual associations that electronic music may evoke, and how these associations shape the listening experience. We first offer an overview of existing listener-based studies on electronic music. We then describe the works used in the current study, offering details about compositional programs, underlying themes, form, material and technique, and compare the works at various structural levels. We outline the aims and the method of the listening experiment, and present its results. Finally, we offer a discussion of these results, focusing on the experiential implications of whether an electronic music piece is primarily abstract or representational; these include where the listener situates themselves in relation to the piece, whether the listener is aware of their "listening self", and how narrative meaning is mediated.

Introduction

As the electronic medium opens music to any and all sounds, the communication between the composer and the listener no longer depends upon a culturally established language of music. Electronic music therefore prompts a unique set of questions pertaining to music perception. The research into what listeners hear in electronic music has a robust history. Composers (e.g., Smalley 1996), musicologists (e.g., Demers 2010), and researchers (e.g., Bridger 1989) have utilized a variety of methods to inquire into the perceptual qualities of experiencing works in this genre. A growing number of researchers today are conducting listening experiments to explore these qualities. In this vein, we have previously presented our findings pertaining to the formation of diegetic affordances (Çamcı and Meelberg 2016), and a gesture-event model of narrativity in electronic music (Çamcı 2016).

In this paper, we focus on two works of electronic music, which were composed in tandem using vastly different sonic vocabularies. Although both works utilize purely synthetic sounds, one of the works stem from the composition of a narrative while the other work relies on a performance-based exploration of perceptual structures such as variations in pitch and amplitude. With this study, we aim to understand the factors that contribute to the cognitive processing of abstract and representational structures in electronic music, and how such structures shape the listening experience.

We describe a study, where 36 subjects were asked to provide both after-the-fact and in-the-moment descriptions of their experiences. We analyse the in-the-moment descriptors on a timeline of the musical track that they correspond to. We then compare descriptors between subjects, and contextualize these within the after-the-fact accounts of the works. We also offer a comparison of these descriptions with the compositional intents underlying these works.

We find that, despite their structural similarities, the difference in vocabulary stemming from compositional intent has a significant impact on how listeners describe their experience of these works. There is a strong preference towards representational descriptors for the piece that emerge from a narrative. We offer an in-depth analysis of how various narratives outlined by participants corroborate in terms of actors, settings and events, and how these relate to the composer's program. Conversely, the work that explores more abstract structures has primarily yielded perceptual and affective descriptors. Comparing the results for the two works, we look at higher-level implications of whether an electronic music piece is primarily abstract or representational; these include how meaning is mediated, whether the listener adopts a first- or third-person perspective towards the piece, whether the listener is aware of their "listening self".

Related Work

Artists and researchers have adopted a variety of methods to explore the perceptual and cognitive processes involved in electronic music listening (Zattra 2005). A majority of these studies rely on expert analyses: for instance, the composer Simon Emmerson describes a discourse grid for electroacoustic music consisting of a *mimetic discourse*, which evokes imagery external to the musical material, and an *aural discourse* based on sound objects which are free of such associations (Emmerson 1986). Similarly, the composer Suk-Jun Kim offers a framework to analyze the emergence of sound-images in electronic music, where *perceiving* and *imagining* function as the two primary processes of listening (Kim 2010). In another study, the composer Gary Kendall adopts an event-schema model to analyze electroacoustic music, and concludes that there is a continuity between the cognitive processes involved in electroacoustic music listening, and those that give rise to meaning in everyday experiences (Kendall 2010).

Besides such studies that draw from expert analyses, there are also those, such as the one discussed in this paper, that derive perspectives from listener feedback. In the Intention-Reception Project, Landy and Weale investigate how revealing details about a work of electroacoustic music to the listener during the course of repeated listening sessions can alter the listener's reception of the work (Weale 2006). The study shows that when inexperienced listeners are given dramaturgical information about a piece, they are able to use it to guide themselves through parts of the music that are problematic in terms of access and appreciation. In another example, the researcher François Delalande conducts a listening experiment with Pierre Henry's *Sommeil* to search for consistencies across listening behaviors (Delalande 1998). Based on testimonial analogies from 8 participants, Delalande outlines three types of listening behaviors: taxonomic listening, where the listener focuses on an overall structure; emphatic listening, where the listener focuses on in-the-moment affective perceptions of the work, and figurativisation, where the listener imagines moving and living things that the sounds suggest. In a study that evaluates the behaviours laid out by Delalande, the composer Elizabeth Lang Anderson finds that the listeners often display a tendency to either identify sound sources or label abstract sounds in a generalized manner (Anderson 2011).

Listening Experiment

As evidenced in these studies, the practically unlimited sonic vocabulary of electronic music opens up a variety of considerations in music cognition that cannot be fully addressed with theories and techniques that apply to more traditional musical practices, therefore necessitating new approaches. Adopting one such approach, we conducted an exploratory listening experiment to evaluate the cognitive processing of abstract and representational structures in electronic music. Electronic music can be broadly defined as music produced using electronic instruments, such as computers, synthesizers, and samplers. There are numerous musical styles that fall under this definition, including Electronic Dance Music genres such as House, Techno and Trance. In this paper, we focus on electronic music compositions that follow an experimental tradition using "analog and digital technologies, concrète and synthetic sound sources, and systematic and intuitive composition strategies" (Roads 2015).

Aim

With this study, we aimed to understand the ways in which the listeners interpret two works of electronic music differently on the basis of representationality. In a preliminary study, we evaluated our procedure and apparatus, and derived broad descriptor categories based on the responses gathered from the participants. In the main study, we explored how a listener's experience of a work is affected by the abstract and the representational elements in the work; to achieve this, we collected both after-the-fact and in-the-moment accounts of this experience.

Stimuli

We focused on two works of electronic music composed concurrently by the author of this paper. Although the two works share many structural similarities, the compositional programs behind these works are significantly different. Furthermore, even though both works are composed using purely synthesized sounds, the processes applied to the sounds reinforce the differences in compositional aims. We chose to focus on the author's work to maintain an intimate dialogue between listener feedback and the author's insights into these works. We attempted to address the conflicts this choice might create by concealing any information about the works from the participants of the study, and subjecting the analysis to peer review.

Birdfish. This piece narrates a story of aquatic organisms gradually evolving first into amphibians, and then into avian creatures. This predetermined narrative informed the composition process from beginning to end. The micro sound structures were created through various combinations of multiple synthesis and signal processing techniques. These processes were applied to achieve representational qualities that were intended to instigate visual references. The resulting structures, which range from few milliseconds to few seconds, were then micromontaged to fulfill the narrative arc. Overall, the composition of *Birdfish* was a non-real-time process that relied on complex sound collages rather than performed sections. The work exhibits a relatively clear motivic development, and a climactic ending signifying the transmutation of the creatures into avian form.

Element Yon. This piece explores the spectral and dynamic extremes of sounds that, in themselves, do not exhibit much textural or temporal complexities. Relying on markedly synthetic sounds, the piece explores how various contrasts in the domains of pitch and amplitude can be highlighted through macro structuring in the time domain. The piece is predominantly a result of performances with a subtractive synthesizer. The formal structure developed from these performances is intended to obfuscate a motivic progression in style of *moment form* (Stockhausen 1963).

Structural Comparisons. The two works were composed concurrently over a 9-month period. Despite the differences in the sound production methods used in these works (i.e. signal processing and micromontaging in *Birdfish* vs. performance with a synthesizer in *Element Yon*), how these materials were sculpted at macro levels show distinct similarities between the two works. Figure 1 shows macro structures of these works, with the spectrogram of each work shown as color variations. The spectrum ranges of both works are similar; however, *Birdfish* utilizes broad-spectrum sounds while *Element Yon* articulates narrow-spectrum sounds from different parts of its spectral range. Both works utilize similar phrase lengths at the gesture level. Furthermore, the macro structures are also of similar length as seen in Figure 1. Another distinct similarity between the pieces is the use and the extent of silence in between structures.

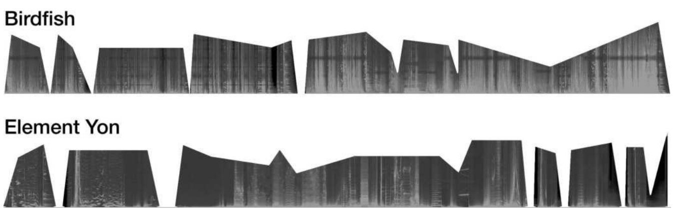


Figure 1. Visual representations of the macro structuring in *Birdfish* (top) and *Element Yon* (bottom) with the X-axis denoting normalized durations. The spectrograms (X: time, Y: frequency) of the pieces are shown with colour variations, where brighter colour indicates higher intensity.

Participants

The preliminary study was conducted with 12 participants (M=10, F=2, average age: 32). The following main study was conducted with 24 participants (M=18, F=6, average age: 28.6). Half of the participants listened to *Birdfish* while the other half listened to *Element Yon*. 4 participants in each group indicated that they had no musical background. The remainder of the participants consisted of students, educators, and professionals in the fields of sound engineering, sound art, and music composition.

Procedure

Both the preliminary and the main studies were based on a two-section, between-subject design. The studies were conducted one listener at a time, and involved two sections. Each instance took approximately 20 minutes, and was administered on a personal computer with closed-back headphones. In the first section, the listener was asked to listen to one of the tracks in its entirety. No information regarding the piece (e.g., title, composer, duration, period) was provided to the listener in advance. Furthermore, the listener was not given a specific task besides listening to the work. Once the listening was completed, the listener was asked to provide their general impressions about the piece in writing; no restrictions were imposed upon the format or the length of this feedback.

Once the first section was completed, the listener was presented with a browser-based user interface to complete the second section of the study. This interface allowed the listener to type and submit descriptors in real time while they listened to the same track from the first section. Each descriptor was saved in a database with a timestamp.

Results

The general impressions gathered from the first section of the study were analyzed both individually and in relation to the corresponding real-time descriptors from the second section. Furthermore, the real-time descriptors were analyzed both within piece and across the participants who listened to the same piece.

Birdfish	**Element Yon**
[water, bubble, splashing, sparkling, fluid, flow, liquid, waves, lake], [living, creatures, animal, amphibian, bird], [slimy, worm, snail, squishing, insect, swarm], [alien, Zerg, StarCraft, sci-fi, Star Wars], [high tech, robots, electronic], [granulating, grinding], [metallic, blades, gong], [sense of space, cave, water park], bass, dialogue	[wide spectrum, spectral, high frequency, low frequency, contrast], [electronic, oscillators, synthetic, abstract], [unpredictable, unstable, unclear, confusing, surprising, exciting], [dangerous, scary, chaotic, argument], [painful, irritating, exhausting, annoying], [relief, relieving, calm], [slow movement, stable, still], silence, [science fiction, Tron]

Table 1. Discourse analysis of the general impressions submitted for *Birdfish* and *Element Yon* with keywords, or keyword groups, that are consistent across 3 or more participants.

A discourse analysis was applied to the general impressions to determine consistent themes and concepts. General impressions expressed in a multiplicity of formats (e.g. prose, list, drawing) are split into "meaningful sections" (Özcan 2012) in the form of keywords. These keywords were grouped across participants by semantic similarity as seen in Table 1. The format of the general impressions, and how frequently a format was used is seen in Table 2. Multiple formats used by a single participant (e.g. a word list and a drawing) were counted separately.

	Word list	Sentence list	Prose	Drawing
Birdfish	5	4	3	1
Element Yon	3	3	6	1

Table 2. The rates of occurrence for the general impression formats for *Birdfish* and *Element Yon*.

The number of real-time descriptors (RTDs) submitted for the two pieces can be seen in Table 3. The descriptors were primarily single words or 2-word noun phrases although a limitation was not imposed on word count. The categories produced from the analysis of the preliminary study were used as an initial guide to categorize these descriptors. Through an iterative process of thematic analysis, these categories were refined and expanded.

	Birdfish	Element Yon
Piece duration	4'40"	3'45"
Total number of RTDs	334	170
Average number of RTDs per participant	27.83	14.16
Average number of RTDs per minute per participant	6.05	3.77

Table 3. Total and average numbers of real-time descriptors (RTDs) for *Birdfish* and *Element Yon* per participant and minute.

The resulting descriptor categories are *source*, *concept*, *location*, *affective*, *perceptual*, *onomatopeia*, and *meta descriptors*. *Source* category includes descriptors that can be prepended by the phrase "the sound of". These descriptors can indicate objects (e.g., "water", "bugs"), actions (e.g., "friction", "explosion"), and musical sources (e.g., "piano"). *Concept descriptors* can be objects or actions; however, they do not refer to sounding objects or phenomena in themselves but might refer to concepts that *imply* such phenomena (e.g., "war", "science fiction"). *Location* descriptors define imagined spaces other than the one inhabited by the listener (e.g., "jungle", "cave"). *Affective descriptors* can indicate emotions experienced by the listener (e.g., "curious", "relief"), binary apparaisal of an experience (e.g., "nice", "cool"), or affective qualities that the listeners attribute to an object (e.g., "weird", "exciting"). *Perceptual descriptors* denote auditory (e.g., "bass", "silence") or non-auditory (e.g., "deep (cave)", "dark (forest)") perceptual qualities. *Onomatopeia descriptors* include words that are formed from sounds (e.g., "boom", "ding"). Finally, *meta descriptors* refer to the track itself rather than the listener's experience of it (e.g., "counterpoint", "granular").

Once the emergent categories were determined, the category membership of each real-time input was assessed through forced-choice categorization. If a descriptor consisted of multiple words and noun phrases, it was split up into its constituents that would fall under a category individually (e.g. "computers underwater" broken into "computers" and "underwater"). The categorical distribution of the descriptors for the two pieces can be seen in Figure 2. According to this figure, more than half of the descriptors submitted for *Birdfish*, are source descriptors (55.96%). On the other hand, the most salient categories for *Element Yon* are *concept* (28.64%) and *perceptual descriptors* (22.39%).

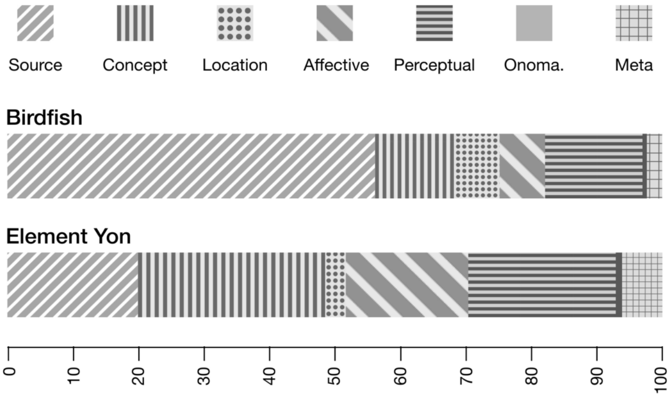


Figure 2. Categorical distribution of real-time descriptors submitted for *Birdfish* (top) and *Element Yon* (bottom).

Discussion

The results conform to our expectation that *Birdfish*, which is composed of complex sound sources based on an underlying narrative, would prompt listeners to describe their experience in terms of sound sources and visual imagery. With *Element Yon*, which is composed of spectrally generic sounds performed with a synthesizer, the listeners seem to have focused more on the perceptual qualities of their experience, and the abstract concepts these qualities have prompted. Furthermore, the prominence of affective descriptors used to describe *Element Yon* points to a tendency to reflect upon the affective experiences these perceptual qualities have instigated. The noticeable difference in the *location descriptors* category for each piece is consistent with the narratively charged feedback received for *Birdfish*: when describing their experiences, the listeners identified both actors and the environments these actors would be situated in.

The general impressions of *Birdfish* were largely in the form of either word and sentence lists enumerating imagined sound sources, or visually-oriented narratives that recounted a story involving sound sources. The general impressions of *Element Yon* were mostly concerned with concepts, such as "flow", "hollowness", "contrast" and "heaviness", and affective appraisals such as "exciting", "curious", "calm" and "annoyed", as well as the physical qualities of sounds, such as "loud" and "bass". Impressions of this piece that relate to objects or environments were also highly conceptual, such as "big magnets", "gravity", "circus-like", "a dark metro station". One participant expressed that "no images came to [their] mind".

This result can be explained by the dominance of representational imagery in *Birdfish*, which brought about impressions that relate more to the descriptions of objects and actions. With *Element Yon*, the participants, besides having *less* to say as evidenced in Table 3, were more inclined to reflect about their affective experience. These results suggest that, despite the structural similarities between these pieces, the sheer difference in how the synthetic sounds were perceived impacted the appraisal of the works on cognitive, perceptual and affective levels.

The collation of the *object* and *action* descriptors under the *source* category was motivated by previous work on sound categorization. In an experiment conducted by Guastavino on the categorization of environmental sounds, the participants often used the source of a sound as a metonym to describe a *sound event* (2007: 55). In another experiment, Dubois et al. similarly found that either source or action descriptors were used by separate participants to classify the same acoustic phenomenon (2006: 867). This behaviour was apparent in the current study as well: for instance, in *Birdfish*, the same gesture was described as "water" and "boiling" by different participants. In another example, the descriptors "fly by" and "bird wings" were used by separate participants to denote the same gesture.

Identifying a Narrative

As shown in Figure 2, *Birdfish* has commonly prompted source descriptors. This was mirrored in the descriptor lists provided in the general impressions for this piece. However, in addition to such lists, the general impressions have also consisted of narratives provided in prose form. Below are two excerpts from such general impressions:

> I heard robotic bugs moving around being commanded by more intelligent robotic beings. There was water, stepping into water, robotic dialogues and also progress made by the robotic bugs in their task. The dialogues were robotic but they had emotion.

> The sounds heard and experienced by a baby in its mother's womb prior to birth, and its eventual coming to earth.

Although the characters and the settings in these two impressions are entirely different, a similar narrative progression can be identified. Furthermore, while the real-time descriptors by these listeners largely adhere to their own general impressions, the similarities in the narrative arcs are evident between listeners in the descriptors that coincide on the timeline of the piece, such as "giving orders" and "talking"; "big cheers" and "volcano"; and "project successful" and "space silence".

Such narrative structures were much less common in the general impressions submitted for *Element Yon*. From a compositional standpoint, this work can be characterized as consisting of "unstructured musical sequences" (Deutsch 1980). The impact of this quality is apparent in the real-time descriptors, such as "exhausting", "chaotic" and "confused". Furthermore, in their general impressions, one participant stated that "sounds without a rhythm made [them] curious but at the same time they were really exhausting". A moment in the piece which this participant marked as being "exhausting" was marked by another participant as "I repeat and repeat but you don't get it". The latter participant had already elaborated on this section in their general impressions when they likened their experience to witnessing a redundant argument between people. The lack of temporal or spectral patterns has therefore created a sense of futility in the listener's mind.

Signs of life. Intonation can be indicative of meaning for both humans (Gussenhoven 2002: 47), and other vocalizing animals (Amador and Margoliash 2013). Even with synthesized sounds, pitch modulations that emulate intonation cues can be expressive of emotion (Scherer and Oshinsky 1978). The rapidly modulated high-resonance filter sweeps used in *Birdfish* were commonly identified by the participants as bird sounds. Similarly, the gestures consisting of rapid frequency modulations of monophonic lines in *Element Yon* were suggestive of an organic origin within the abstract sound-world of the piece. This was evident in descriptors such as "crying", "scream", "I guess he is trying to tell us something", "communication", and "conversation" submitted by 5 different participants.

Situating the narrative. In *Birdfish*, medium to largely sized reverberations and low frequency rumbles are used to establish the sense of a large enclosed environment. These were addressed in the real-time descriptors that identify imaginary environments, which were often in agreement with preceding source descriptors submitted by the listener. A listener who had used such descriptors as "animal" and "water", submitted "big huge cave" at a later point in the piece. Another listener who identified such sources as "factory noises" and "light saber", described that same point in the piece with "spaceship". Two different listeners marked another point in the piece with "dungeon" and "big spaceship" preceded by source descriptors that similarly conform to narratives that transpire in these environments. This indicates that sources and locations identified in the piece tend to reinforce each other's narrative prominence.

Perceptually driven narratives. In *Element Yon*, the frequency and damping characteristics of certain sounds instigated such descriptors as "metal balls getting bigger and smaller", "high tone falls and hits the ground". Here, distinctly perceptual qualities are situated within metaphors while retaining their embodied relationship with the listener. Another similar example is observed in the responses to high frequency gestures in *Birdfish*, which listeners characterized with such descriptors as "ice", "glass", "metal", "blade" and "knife". These metaphorical associations imply a cross-modal similarity between high frequencies and the sense of sharpness in terms of perceptual qualities.

Sense of a "listening self"

Spectral and dynamic extremes, such as very high and very low frequencies that are distinctly loud or quiet, can make the listeners conscious of their listening selves. Almost half of the participants who listened to *Element Yon* expressed a form of annoyance with the high frequencies with such descriptors as "disturbing", "annoyed" or "irritating". Furthermore, three quarters of the participants marked the rests in the piece either by pointing out the pockets of silence themselves, or by describing the relief these induce. While *Birdfish* incorporates high frequency gestures comparable to those in *Element Yon*, only one participant used the descriptor "harsh high" to indicate a similar annoyance. Furthermore, although the silences are structured very similarly between the two pieces, none of these silences were denoted as bringing relief in the context of *Birdfish*. This indicates a variation in engagement with perceptual qualities of a piece based on its representational characteristics. In an abstract work, where the listener is left to focus on the perceptual qualities of the work, same spectral extremes tended to draw an amplified recognition of one's act of listening.

Interestingly, the participants who listened to *Element Yon* wrote their general impressions mainly in first person (i.e. "I felt (…)" while the participants who listened to *Birdfish* commonly used a third-person voice (i.e. "(…) happened"). This, again, indicates a noticeable sense of listening self in response to abstract perceptual qualities in *Element Yon*, where the listeners have reported back on the ways in which such qualities had affected them. Conversely, the representational sounds of *Birdfish* prompted them to describe a series of events that they *witnessed* rather than *experienced*. This variation in voice suggests that the prominence of representational elements in a piece affects where the listeners situate themselves in relation to a piece. While with *Element Yon* the listeners tended to place themselves at the center of the experience, the listeners of *Birdfish* assumed the role of a spectator watching a narrative unfold at the core of the piece.

Conclusion

The expansive sonic vocabulary of electronic music brings about unique considerations in music cognition; it can consist of sounds that are structurally generic, or as complex as everyday sounds. This acoustic variety leads to a wide range of listening experiences. In this paper, we investigated these experiences from a representationality perspective. We proposed a new experimental design for the study of the cognitive processes involved in electronic music listening. By focusing on two works that utilize contrasting sonic vocabularies, we explored the differences in how listeners perceive abstract and representational elements in electronic music. We believe that this form of inquiry bears significant potential to expanding our understanding of how listeners experience electronic music.

We found that the prominence of representational elements in *Birdfish* has prompted a majority of the listeners to recount their experiences in terms of descriptors pertaining to perceived sources, and narratives constructs. These accounts were mainly provided from the perspective of an outside viewer. We also observed that a narrative context can affect the recognition of sources and vice versa. On the other hand, the prominence of abstract structures in *Element Yon* has resulted in a preference towards conceptual descriptors. The listeners focused on perceptual qualities of their experience, and provided affective interpretations of these qualities. Furthermore, these listeners were more inclined to report on their experiences in first-person. When narrative constructs emerged, these were inherently informed by perceptual qualities, and were associated with the "listening self". Despite the structural similarities between the two pieces, both the general impressions and the categorization of real-time descriptors displayed significant variations, indicating the ways in which abstract and representational structures in electronic music can impact a listener's experience.

References

Amador, A., & Margoliash, D. (2013). "A Mechanism for Frequency Modulation in Songbirds Shared with Humans". *The Journal of Neuroscience* 33(27), 11136-11144.

Anderson, E. L. (2011). *Materials, Meaning and Metaphor: Unveiling Spatio-Temporal Pertinences in Acousmatic Music*. PhD Dissertation, City University London, UK.

Bridger, M. (1989). "An approach to the analysis of electro-acoustic music derived from empirical investigation and critical methodologies of other disciplines". *Contemporary Music Review* 3(1), 145-160.

Çamcı, A. (2016). "Imagining through Sound: an experimental analysis of narrativity in electronic music". *Organised Sound*, 21(3), 179-191.

Çamcı, A., & Meelberg, V. (2016). "Diegetic Affordances and Affect in Electronic Music". In *Proceedings of the 2016 International Computer Music Conference* (pp. 104-109). Delft, The Netherlands.

Delalande, F. (1998). "Music Analysis and Reception Behaviours: Sommeil by Pierre Henry". *Journal of New Music Research* 27(1-2), 13-66.

Demers, J. (2010). *Listening Through the Noise: The Aesthetics of Experimental Electronic Music*. New York: Oxford University Press.

Deutsch, D. (1980). "The Processing of Structured and Unstructured Tonal Sequences". *Perception and Psychophysics*, 28(5), 381-389.

Dubois, D., Guastavino, C., & Raimbolt, M. (2006). "A Cognitive Approach to Urban Soundscapes: Using Verbal Data to Access Everyday Life Auditory Categories". *Acta Acustica United with Acustica* 92(6), 865-874.

Emmerson, S. (1986). "The Relation of Language to Materials". In Simon Emmerson (ed.), *The Language of Electroacoustic Music*: 17-40. Houndmills, UK: Macmillan Press.

Guastavino, C. (2007). "Categorization of Environmental Sounds". *Canadian Journal of Experimental Psychology*, 61(1), 54-63.

Gussenhoven, C. (2002). "Intonation and interpretation: Phonetics and Phonology". *Proceedings of the 2012 International Conference on Speech Prosody* (pp. 47-57). Aix-en-Provence, France.

Kendall, G. S. (2010). "Meaning in Electroacoustic Music and the Everyday Mind". *Organised Sound*, 15(1), 63-74.

Kim, S. (2010). "Imaginal Listening: a quaternary framework for listening to electroacoustic music and phenomena of sound-images". *Organised Sound* 15(1), 43-53.

Özcan, E., & van Egmond, R. (2012). "Basic semantics of product sounds". *International Journal of Design*, 6(2), 41-54.

Roads, C. (2015). *Composing Electronic Music: A New Aesthetic*. New York: Oxford University Press.

Scherer, K. R., Oshinsky, J. S. (1977). "Cue Utilization in Emotion Attribution from Auditory Stimuli". *Motivation and Emotion* 1(4), 331-346.

Smalley, D. (1996). "The listening imagination: Listening in the electroacoustic era". *Contemporary Music Review* 13(2), 77-107.

Stockhausen, K. (1963). "Momentform: Neue Beziehungen zwischen Aufführungsdauer, Werkdauer und Moment". In *Texte zur Musik vol. 1: Aufsätze 1952–1962 zur Theorie des Komponierens*, 189-210. DuMont Schauberg: Cologne, Germany.

Weale, R. (2006). "Discovering How Accessible Electroacoustic Music Can Be: The Intention/Reception Project". *Organised Sound*, 11(2), 189–200.

Zattra, L. (2005). Analysis and Analyses of Electroacoustic Music. In *Proceedings of the 2005 Sound and Music Computing Confernece*, Salerno, Italy.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

A Computational Study of the Role of Tonal Tension in Expressive Piano Performance

Carlos Cancino-Chacón[1,2] and Maarten Grachten[2]

[1] *Austrian Research Institute for Artificial Intelligence, Vienna, Austria*
[2] *Institute of Computational Perception, Johannes Kepler University Linz, Austria*
carlos.cancino@ofai.at

Abstract

Expressive variations of tempo and dynamics are an important aspect of music performances, involving a variety of underlying factors. Previous work has showed a relation between such expressive variations (in particular expressive tempo) and perceptual characteristics derived from the musical score, such as musical expectations, and perceived tension. In this work we use a computational approach to study the role of three measures of tonal tension proposed by Herremans and Chew (2016) in the prediction of expressive performances of classical piano music. These features capture tonal relationships of the music represented in Chew's *spiral array model*, a three dimensional representation of pitch classes, chords and keys constructed in such a way that spatial proximity represents close tonal relationships. We use non-linear sequential models (recurrent neural networks) to assess the contribution of these features to the prediction of expressive dynamics and expressive tempo using a dataset of Mozart piano sonatas performed by a professional concert pianist. Experiments of models trained with and without tonal tension features show that tonal tension helps predict *change* of tempo and dynamics more than absolute tempo and dynamics values. Furthermore, the improvement is stronger for dynamics than for tempo.

Introduction

Expressive performance of music constitutes an important part of our enjoyment of several kinds of music, including Western art music and jazz. In these kinds of music, an expressive performance is not expected to be an exact mechanical rendition of what is written in the score, but a combination of the performer's interpretation of both the intentions of the composer and its own expressive intentions that are conveyed to the listener through variations in dimensions such as tempo, dynamics and timbre. Previous work has showed a relation between such expressive variations and perceptual characteristics derived from the musical score, such as musical expectations, and perceived tension (Chew, 2016; Farbood, 2012; Gingras et al., 2016).

The concept of musical tension is highly complex and multidimensional, and thus, difficult to formalize or quantify (Farbood, 2012; Herremans & Chew, 2016). Informally, *"increasing tension can be described as a feeling of rising intensity or impending climax, while decreasing tension can be described as a feeling of relaxation or resolution"* (Farbood, 2012, pp. 387). The music cognition literature has shown that aspects related to musical tension include both psychological factors such as expectation and emotion; and musical factors such as rhythm, timing and dynamics and tonality. For a more thorough description of aspects that contribute to musical tension, we refer the reader to (Farbood, 2012) and references therein.

In this work we use a computational approach to study the role of tonal tension features –as proposed by (Herremans & Chew, 2016)– in the prediction of expressive performances of Classical piano music. Computational models of musical expression can be used to explain the way certain properties of a musical score relate to an expressive rendering of the music (Widmer & Goebl, 2004). The KTH model (Friberg, Bresin, & Sundberg, 2006), one of the most important *rule-based* models of expressive performance includes rules that take into account tonal tension.

The rest of this paper is structured as follows: the Method section is divided into three subsections, the first of which describes the tonal tension and score features, followed by a brief description of the way expressive tempo and dynamics are quantified in this work; and finally, the recurrent neural network model relating the features to the expressive tempo and dynamics is presented. The Experiments section describes the evaluation of the methods using cross-validation experiments. Afterwards we discuss the results of these experiments and, finally, we present conclusions and future research directions.

Method

This section details the computational methodology we use in this study. It describes the features used to represent musical contexts – the inputs of the model, the expressive parameters used to represent tempo and dynamics – the outputs of the model, and the model itself.

Features

Tonal Tension Features (T)

In order to characterize tonal tension, we use a set of three quantities, which are computed using the method proposed by Herremans and Chew (2016). These features capture tonal relationships of the music represented in Chew's (2000) *spiral array model*, a three dimensional representation of pitch classes, chords and keys constructed in such a way that spatial proximity represents close tonal relationships. The tonal tension features are:

1. *cloud diameter* (T_{cd}), which estimates the maximal tonal distance between notes in a segment of music;
2. *cloud momentum* (T_{cm}) quantifies harmonic movement as the tonal distance from a section to the next; and
3. *tensile strain* (T_{ts}), the relative tonal distance between the current segment and the center of effect of the key of the piece, i.e. the point in the spiral array mode that best represents the key of the piece.

Since distances in the spiral array are can be large (in this particular work, an order of magnitude larger than the score features defined below), we scale the tension features described above by dividing them by the distance between enharmonically equivalent notes (e.g. C♯ and D♭). Herremans and Chew (2016) evaluated these features by comparing them to the empirical study by Farbood (2012), showing that these features correlate to human perception of tonal tension.

Score Features

Following Cancino-Chacón, Grachten, Sears, and Widmer (2017), we include two groups of low-level descriptors of a musical score that have been shown to predict characteristics of expressive performance. These features provide a baseline showing to what degree expressive variations can be explained only by the nominal information in the score.

1. **Pitch (P)**
   (a) $(pitch_h, pitch_l, pitch_m)$. Three features representing the chromatic pitch (as MIDI note numbers divided by 127) of the highest note, the lowest note, and the melody note (if given, and zero otherwise) at each score position.
   (b) (vic_1, vic_2, vic_3). Three features describing up to three vertical interval classes above the bass, i.e. the intervals between the notes of a chord and the lowest pitch, excluding pitch class repetition and octaves. For example, a C major triad (C, E, G), starting at C_4 would be represented as $\begin{pmatrix} pitch_l & vic_1 & vic_2 & vic_3 \end{pmatrix} = \begin{pmatrix} \frac{60}{127} & \frac{4}{11} & \frac{7}{11} & 0 \end{pmatrix}$, where 0 denotes the absence of a third interval above C_4, i.e. the absence of a fourth note in the chord.
2. **Metrical (M)**
   (a) $b_{\phi,t}$. The relative location of an onset within the bar, computed as $b_{\phi,t} = \frac{t \mod B}{B}$, where t is the temporal position of the onset measured in beats from the beginning of the score, and B is the length of the bar in beats.
   (b) (b_d, b_s, b_w). Three binary features (taking values in $\{0, 1\}$) encoding the metrical strength of the t-th onset. b_d is nonzero at the downbeat (i.e. whenever $b_{\phi,t} = 0$); b_s is nonzero at the secondary strong beat in duple meters (e.g. quarter-note 3 in $\frac{4}{4}$, and eighth-note 4 in $\frac{6}{8}$), and b_w is nonzero at weak metrical positions (i.e. whenever b_d and b_s are both zero).

Expressive Parameters

We consider an *expressive parameter* to be a numerical descriptor that corresponds to common concepts involved in expressive performance. In this section we briefly describe the parameters used to represent expressive tempo and dynamics.

1. **Tempo**.
   (a) *BPR*. We take the local *beat period ratio* as a proxy for musical tempo. In order to compute this parameter, we average the performed onset times of all notes occurring at the same score position and then compute the *BPR* by taking the slope of the averaged onset times (in seconds) with respect to the score onsets (in beats) and dividing the resulting series by its average beat period.
   (b) d*BPR*. This parameter is computed as the first derivative of *BPR* with respect to the score position, and corresponds to the *relative acceleration*, i.e. the relative changes in musical tempo.
2. **Dynamics**.
   (a) *VEL*. We treat the performed MIDI velocity as a proxy for the loudness of the note. This parameter computed by taking the maximal performed

MIDI velocity per score position, divided by 127. We use the terms loudness and dynamics interchangeably to refer to this parameter.

(b) d*VEL*. This parameter is computed as the first derivative of *VEL* with respect to the score position, and corresponds to the relative changes in loudness from one time step to the next.

Model

We use recurrent neural networks (RNNs), a family of non-linear sequential models, to assess the contribution of the features described above to the prediction of expressive dynamics and tempo. RNNs are a state-of-the-art family of neural architectures for modeling sequential data (Goodfellow, Bengio, & Courville, 2016). These models have been used to model expressive dynamics and tempo (Cancino-Chacón, Gadermaier, Widmer, & Grachten, 2017; Grachten & Cancino-Chacón, 2017). In this work, we use a simple architecture, which we will refer to as bRNN, consisting of a composite bidirectional long short-term memory layer (LSTM) with multiplicative integration (Wu, Zhang, Zhang, Bengio, & Salakhutdinov, 2016) with 10 units (5 units processing information forwards and 5 processing information backwards) and a linear dense layer with a single unit as output.

Experiments

In order to evaluate the contribution of the features described above to the prediction of expressive tempo and dynamics, we perform a cross-validation experiment to test the predictive quality of the model. For this study, we use the Batik/Mozart dataset (Widmer & Tobudic, 2002), which consists of recordings of 13 piano sonatas by W. A. Mozart performed by Austrian concert pianist Roland Batik which have been aligned to their scores. An important characteristic of this dataset is that the melody voices are manually identified. These performances were recorded on a Bösendorfer SE 290, a computer controlled grand piano.

For each expressive parameter, we perform eight 5-fold cross-validation experiments corresponding to models trained on all combinations of feature sets, i.e. all combinations of pitch features (**P**), metrical features (**M**) and tension features (**T**); as well as a feature set consisting of a selection of features using a univariate feature selection method (**FS**). Each 5-fold cross-validation experiment is conducted as follows: each model is trained/tested on 5 different partitions (folds) of the dataset, which is organized into training and test sets, such that each piece in the corpus occurs exactly one in the test set. For each fold, we use 80% of the pieces for training and 20% for testing the model. The parameters of the model are learned by minimizing the mean squared error on the training set using RMSProp, a variant of the stochastic gradient descent algorithm (Tieleman & Hinton, 2012).

The feature selection procedure computes the pairwise *mutual information* between each of the features and each of the expressive parameters. This information-theoretic measure expresses how much knowing the value of one variable reduces uncertainty about the value of the other variable (Ross, 2014), and is a common way of determining the relevance of features in prediction tasks. Formally, the mutual information between two variables x and y is given by

$$\mathrm{MI}(x, y) = \mathbb{E}\left\{\log\left(\frac{p(x, y)}{p(x)p(y)}\right)\right\},$$

where $\mathbb{E}$ is the expectation operator, $p(x, y)$ is the joint probability distribution of x and y, and $p(x), p(y)$ are the marginal probability distributions of x and y, respectively. If x and y are statistically independent (i.e. they do not share information about each other), the mutual information is zero. In the **FS** scenario we select the 10 features with the largest mutual information for each expressive parameter. This procedure was performed on a small subset of the Batik/Mozart dataset (20% of the pieces selected randomly).

Discussion

Figure 1 shows the mutual information between each feature and the expressive parameters, normalized for each expressive parameter. In this plot, the height of a column (the value of the mutual information) signifies how closely related that feature is to the expressive parameter. The results in this plot suggest that the tension features, in particular the cloud diameter might be more related to the prediction of changes in tempo and dynamics, whereas the tensile strain might be more related to absolute tempo and dynamics than their changes. From a musical perspective, these findings seem plausible, since cloud diameter refers to melodic events, whose performance might be dependent on the character of the passage, whereas tensile strain depends on structural harmonic characteristics of the music.

Note that the values in Figure 1 only measure the MI of the features and expressive parameters at isolated time instances, without context. Although this gives a good first impression of the relevance of features, the bRNN model presented above is specifically designed to take advantage of the temporal context to make predictions, implying that feature values at times before and after τ may also help to predict expressive parameters at τ. Therefore Figure 1

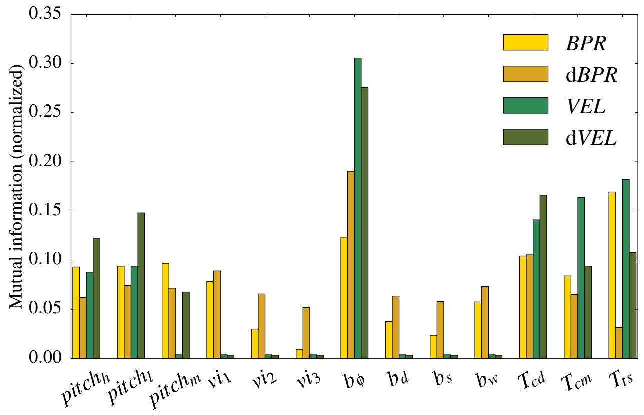


Figure 1: Normalized mutual information between features and expressive parameters. Larger mutual information means that the variables are more related to each other.

does not necessarily reflect the relevance of features when used as input to the bRNN.

Table 1 displays the results of the cross-validation experiments using the bRNN on different feature sets. We evaluate model accuracy with the coefficient of determination (R^2). This measure expresses the proportion of variance of the expressive parameters that can be explained by the models as a function of the feature set. For each expressive parameter, we conducted a paired–samples two-tailed t-test at the $p < 0.01$ level to compare the differences between the R^2 values of features sets with and without tension features (i.e. **P** vs. **P** + **T**, **M** vs. **M** + **T**, and **M** + **P** vs. **M** + **P** + **T**). The effect of including **T** was not significant for the prediction of *BPR* and *VEL* but it *was* significant for the prediction of d*VEL* in all cases. For d*BPR* including **T** was only beneficial in combination with **P**. Based on this result we hypothesize that the concept of tonal tension is relevant for changes in tempo, but the features used to represent tonal tension may not convey enough information by themselves and are therefore only advantageous in combination with more specific pitch information.

In order to visualize the contribution of each feature to the prediction of the changes in tempo and dynamics, we perform a differential sensitivity analysis[1] of the models by computing a local linear approximation of the output of the bRNNs trained on all features (**P** + **M** + **T**). The resulting sensitivities are plotted in Figure 2. This figure can by roughly interpreted as follows: The color in the cell that corresponds to feature f and time step t represents the contribution of the value of f at time t to the prediction of the expressive parameter at time $t = \tau$ (the center column of the plot). Blue tones reflect feature values that negatively contribute to the predicted value (the higher the feature value the lower the predicted value), and red tones reflect feature values that positively contribute to the predicted value.

[1]We use the definition of sensitivity analysis from the applied mathematics literature, not in the sense used in the psychology literature.

Table 1: Proportion of variance explained (R^2) for expressive tempo and dynamics using different feature sets, averaged over all pieces on the Batik/Mozart corpus. Larger R^2 values reflect more accurate predictions. For each combination of target and feature set, the results are listed for that feature set as is (left), and including tonal tension features T (right). For clarity, improvements of R^2 as a result of adding T are marked in green, whereas detriments are marked in red. Bold numbers mark a statistically significant difference ($p <$ 0.01). The effect size (Cohen's d) is reported in parenthesis for those cases with statistically significant differences.

	Tempo				
	BPR		d*BPR*		
Feature Set		+ **T**		+ **T**	d
∅	-	0.010	-	0.011	
P	0.024	0.021	0.068	**0.073**	(0.10)
M	0.051	0.054	0.093	0.092	
P + **M**	0.056	0.054	0.105	**0.110**	(0.06)
	Dynamics				
	VEL		d*VEL*		
Feature Set		+ **T**		+ **T**	d
∅	-	0.018	-	0.026	
P	0.326	0.335	0.236	**0.250**	(0.16)
M	0.048	0.052	0.041	**0.050**	(0.18)
P + **M**	0.351	0.347	0.250	**0.282**	(0.40)

The plots follow a similar trend to the results showed in Figure 1, where the features with higher mutual information also have brighter colors in Figure 2, with the added benefit that the contribution of each feature at different time steps can be visualized.

The plots in Figure 2 suggest a tendency of the performer to emphasize melodic and harmonic events by slowing down. For example, a chromatic melody note in an otherwise tonally stable section of the piece, (like the presentation of the main theme during the exposition of a sonata) is emphasized by slowing down (see the reddish hue in T_{cd} for time-steps τ and $\tau + 1$ in the left plot); while upcoming chromatic notes contribute to speeding up (the bluish hue in T_{cd} for time-steps $> \tau + 1$). On the other hand, chords that are tonally far from the current key are emphasized by slowing and an increase in loudness (the reddish hue for time-step τ in T_{cm} in the left plot and in T_{ts} in the right plot). Furthermore, upcoming sections with modulations to distant keys (like the more unstable parts of the development of a sonata) contribute to speeding up

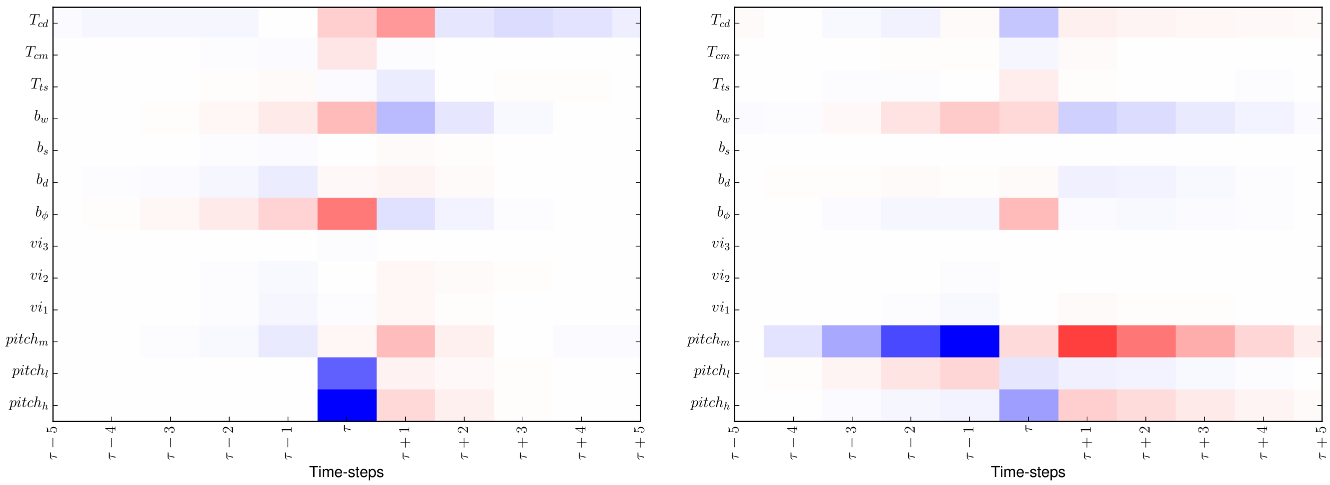


Figure 2: Differential sensitivity plots for d*BPR* (left) and d*VEL* (right). Each row in the plot corresponds to an input feature and each column to the contribution of its value at that time-step to the output of the model at time τ (the center of each plot). Red and blue indicate a positive (i.e. slowing down or increasing loudness) and negative contribution (i.e. speeding up or decreasing loudness), respectively.

(the bluish hue in T_{cd} for time-steps $> \tau+2$ in the left plot) These findings agree with common performance rules (c.f. the *melodic charge* and *harmonic charge* rules in the KTH model (Friberg et al., 2006)).

Conclusions

In this work we have empirically investigated the role of tonal tension in shaping musical expression in classical piano performances. Our experimental results show that using tonal tension information improves predictions of *change* of tempo and dynamics, but not predictions of absolute tempo and dynamics. For predicting changes in tempo, using tonal tension features as defined in (Herremans & Chew, 2016) was only beneficial when low level pitch information was also available. This suggests that tonal tension features are potentially relevant for predicting tempo changes, but by themselves not sufficiently specific for that purpose.

Furthermore, an analysis of the trained models corroborates previously formulated relationships between performance and tension, as defined in the KTH model.

Future work may focus on a more explicit testing of the hypothesis that recurrent neural network models may learn features describing tonal characteristics from low level pitch information as a side effect of learning to predict expressive tempo and dynamics.

Acknowledgments. Funding for this work was provided by the European Research Council (ERC) under the EU's Horizon 2020 Framework Programme (ERC Grant Agreement number 670035, project "Con Espressione").

References

Cancino-Chacón, C. E., Gadermaier, T., Widmer, G., & Grachten, M. (2017). An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonic music. *Machine Learning*, *106*(6), 887–909.

Cancino-Chacón, C. E., Grachten, M., Sears, D. R. W., & Widmer, G. (2017, October). What were you expecting? Using Expectancy Features to Predict Expressive Performances of Classical Piano Music. In *Proceedings of the 10th International Workshop on Machine Learning and Music (MML 2017).* Barcelona, Spain.

Chew, E. (2000). *Towards a Mathematical Model of Tonality* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA, USA.

Chew, E. (2016). Playing with the Edge: Tipping Points and the Role of Tonality. *Music Perception*, *33*(3), 344–366.

Farbood, M. M. (2012, April). A Parametric, Temporal Model of Musical Tension. *Music Perception*, *29*(4), 387–428.

Friberg, A., Bresin, R., & Sundberg, J. (2006). Overview of the KTH rule system for musical performance. *Advances in Cognitive Psychology*, *2*(2-3), 145–161.

Gingras, B., Pearce, M. T., Goodchild, M., Dean, R. T., Wiggins, G., & McAdams, S. (2016). Linking melodic expectation to expressive performance timing and perceived musical tension. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 594–609.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Grachten, M., & Cancino-Chacón, C. E. (2017). Temporal dependencies in the expressive timing of classical piano performances. In M. Lessafre, P.-J. Maes, & M. Leman (Eds.), *The routledge companion to embodied music interaction* (pp. 360–369). Routledge.

Herremans, D., & Chew, E. (2016). Tension ribbons: Quantifying and visualising tonal tension. In *Proceedings* *of the Second International Conference on Technologies for Music Notation and Representation (TENOR 2016).* Cambridge, UK.

Ross, B. C. (2014). Mutual information between discrete and continuous data sets. *PLOS ONE*, *9*(2), e87357.

Tieleman, T., & Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. In *COURSERA: Neural Networks for Machine Learning.*

Widmer, G., & Goebl, W. (2004, September). Computational Models of Expressive Music Performance: The State of the Art. *Journal of New Music Research*, *33*(3), 203–216.

Widmer, G., & Tobudic, A. (2002). Playing Mozart by Analogy: Learning Multi-Level Timing and Dynamics Strategies. *Journal of New Music Research*, *32*, 259–268.

Wu, Y., Zhang, S., Zhang, Y., Bengio, Y., & Salakhutdinov, R. (2016). On Multiplicative Integration with Recurrent Neural Networks. In *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS 2016).* Barcelona, Spain.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

From Play to Gesture: Exploring the Intrinsic Relations between Body and Mind in the Pedagogy of Musical Performance

Isadora Scheer Casari,[1] Mônica Duarte,[2] and Marcos Nogueira[3]

[1,2] *Department of Music, University Federal do Estado do Rio de Janeiro (UNIRIO), Brazil*
[3] *Department of Music, University Federal do Rio de Janeiro (UFRJ), Brazil*

[1]isadorascheer@hotmail.com, [2]monicaduarte33@gmail.com,[3]mvinicionogueira@gmail.com

Abstract

Comprehension is a type of action and as such refers to our corporeality and emotions. In turn, the Western epistemological tradition advocates that the body and emotions are undesirable interferences to our production of knowledge. The notion that the mind is transcendental and the body no more than transient matter leads to a denial of our corporeality, the expectation of absolute control over action, and the hierarchized relations between body and mind, which limits our self-understanding. This paper aims to surpass this dichotomy and propose a theoretical reflection on the construction of the sound production gesture and its association with play and the metaphor. The expression of musical ideas and intentions may only gain materiality through the construction and expansion of a repertoire of actions. Our hypothesis is that the behavior present in play allows for the exchange between different domains of experience, as it happens in metaphorical thinking, and that play configures itself as an experiential field that facilitates the development of creativity and comprehension, enabling the transformation and development of actions and behaviors. Gadamer (2010) uses the concept of play (*spiel*) as the guiding thread for his model since he considers that, as a game, the entire process of constructing meanings is an open act and an interpretive and interactive action.

Introduction

During World War II the German violist and teacher, Bertha Volmer, had the need to work in a factory and her job required the constant use of a heavy hammer. During the first few weeks, she suffered constant pains in her hands and wrists due to her lack of experience in handling the tool. Over time, Volmer realized the importance of letting the weight of the hammer guide the movement of the arm, which in turn became more fluid. When developing a technique for using the hammer, she transformed a disorderly, tense, and inefficient movement into a gesture, that is, an intentionally constructed action. After the war ended, she resumed her profession as a violist, and somehow the experience of having developed a technique for using the hammer changed her "conception" of the bow technique. The understanding that the weight of the hammer should guide the movement of the arm made it clear to her that the same was true of the bow: its weight and the weight of the arm should guide movement. She intuitively selected some particularities of her experience of handling the hammer, such as the control and balance of forces in motion, and projected them unto her experience with the bow. By activating two distinct domains of experience, she designed body schemes learnt from handling the hammer and projected them on her bow technique. When she felt the bow like the hammer, she experienced a metaphorical inference. This insight transformed her understanding of the gesture when conducting the bow.

Movement, Gesture and Performance

I would like to draw attention to two aspects that emerge when we analyze this experience. First, the experience of the movement plays a central role in musical performance. The intention and vitality inflicted upon a sound-producing gesture will noticeably affect the sound result. A direct relationship exists between the concrete physical movement of the body executing the performance and the movement expressed and perceived in the sonic stream. The gesture that conducts the bow, for example, is multifariously complex. The right upper limb, the system comprised of wing bones, shoulders, arms, wrists, hands and fingers, partakes as a whole in conducting the bow. This complex system consisting of muscles, bones, and joints has the task of balancing weight and force distribution in one single point: the contact of the horsehair with the string. The perfection of this gesture, which occurs through the dynamism within this action, depends on perfecting skills common to every moving body (balance, flexibility, distribution of forces, and the imagery of trajectories guiding the movement) walks alongside the refinement of the sonority in bowed string instruments.

Our action on physical objects yields sound and vigor, and the energy and direction of such actions modify the sound result. The same movement can be performed with different vitality forms (Stern, 2010) and leave different traces in the environment. The vitality form in a sound-producing gesture influences its sound result. Understanding the role of the body in musical performance as strictly bound to sound production mechanics follows a partial view of musical technique, one that disconnects intention and musical meaning from the concrete action producing them. By disconnecting means and intention, it ultimately disconnects body and mind (body as means and mind as intention). Gestures are not merely conventions or instrumental for performance, but emerge from spontaneous acts of vitality. "Every gesture is as outward expression of the vitality of the subject's dynamic 'state of mind', the feeling of how motive energy flows in the muscles of the body through a controlled time" (Trevarthen et al, 2011, p.14)

Body and Mind Dichotomy

We must challenge the naturalized dualism in the West by which the experience of the body is an activity with little connection to language, thought, meaningful action, and the construction of meaning. The dualistic conception of mind and body gives rise to the distinction between knowledge and skill. An understanding of mental processes as the computing of information presents limitations when representing the human skill in propositional forms. Especially in debates regarding the construction of knowledge in "performance arts", which implicate the construction of explicit knowledge and tacit knowledge (Gill, 2015), a definition of cognition based solely in the computation of symbolic representations produces a "Non-person Centered Tradition of Knowledge" (Gill, 2015) and neglects the role of corporeality, dialogue, intersubjective experience, and context in the construction of skills and knowledge, since "The idea of representation that lies within the concept of propositional knowledge, requires that the world be defined in terms of components, each of which mirrors the reality."(Gill, 2015, p.43).

In turn, both performance and musical appreciation involve cognitive dimensions produced by the body in action and interaction."Embodied cognition theories suggest that our physical gesturing both expresses and plays an active role in constituting our thought, and also that our use of linguistic metaphor (including that employed in music theories) is based on embodied image schemas" (Clayton & Leante, 2013, p.270). This is not to say that the body rules or prevails on musical performance, or that playing an instrument is merely the fine coordination of complex body movements regardless of a musical meaning. This idea departs from the theory of embodied cognition (Varela, 1991; Lakoff & Johnson, 1999) as well as the Cartesian notion that thought has a purely metaphysical origin. Understanding that body and mind form a cohesive unit implies that our musical experience is not merely the abstract manipulation of symbols, but the outcome of the incorporation of actions experienced in different contexts.

The coordination of actions executed by the body in musical performance entails the creation of narratives that communicate intentions. Trajectories, directions, and intentions change according to the performance of gestures in a stream we learn to coordinate narratively. When we are well coordinated, the narrative becomes clear and easily followed. Conversely, if we are poorly coordinated, whether by stress, anxiety, lack of preparation, or even a motor system dysfunction, our body narrative is impaired, which limits the communication and expression of musical meaning. One of the goals of musical learning is the development of communication through sounds materialized in the gestural experience of sound production. The sound-producing expressive gesturality that mobilizes and communicates affections becomes possible through the perception of the actual body as a cognitive agent and its vitality as an expressive factor.

The Experience of Musical Performance

It is a tacit assumption that musical performance is the expression of an artistic, sonorous, and musical idea. However, this understanding remains insufficient to explain the experience of musical performance from the perspective of its production. The expression of musical ideas and intentions may only gain materiality through the construction and expansion of a repertoire of actions. We consider the experience of performance to be the construction of gesture, understood here as the gestalt between the aural image (which sound do I wish to create? With what intention?), the corporal disposition towards action (how do I do perform the action?), and the consolidation of this gestalt through practice. The construction of gesture in musical performance is the mapping of a physical action in an acoustic consequence. An action enveloped in dynamic vitality forms with the intent of expressing affections and communicating narratives.

Metaphor and Performance: Seeing-as

The second aspect I would like to address, stated at the beginning of the article, is the presence of metaphor as a cognitive device in the construction of meaning. A cognitive process is metaphorical when the duality present in the activation of distinct domains of experience integrates the process of construction of meaning (Müller, 2008). We understand "construction of meaning" as the organization of experiences in coherent units as well as the categorization and allocation of new events and objects within a meaningful network. The use, even if intuitive, of metaphorical procedures in the development and improvement of gestures in musical performance is a common device among musicians and, since they communicate sensations challenging to express otherwise, the metaphor is a pedagogical resource to convey experiences that emerge in musical performance. Metaphor creates co-participation within action and such co-participation is a crucial element in the construction of knowledge (Gill, 2015).

Ikuta (1990) discusses the importance of developing a metaphorical language when teaching and learning skills. The author studies the role of metaphorical language, which he calls "craft language", in the transmission and construction of knowledge in waza – a traditional Japanese artistic performance. This metaphorical language is radically opposite to a descriptive or scientific language. Craft language is effective for inducing and suggesting sensations in the apprentice's body as the metaphor presents an interactive aspect, creating an intersectional space between the teacher and student's experience. The teacher, when aware of the similarity between metaphor and the gesture to be mastered, uses the cognitive activation triggered by the metaphor with the intention of helping the student discover and feel through visual and kinesthetic imagination the intention present in a gesture. The metaphor creates a space for the intersection and negotiation of meanings enabling the communication of the subjective and embodied experience of the subjects. Within the process of teaching and learning skills, a teacher can thus share through metaphors how they perceive their own gestural experience inviting the student to feel that same experience in their bodies.

The creation of metaphors and the establishment of the experience of metaphoricity (Müller, 2008) is intrinsically related to the human capacity to find similarities and map experiences. Such process is always constructive and creative. The metaphor is an entity that emerges from the cognitive activation, and its use allows for the intersection between different cognitive modalities. Metaphors, while also occurring language, albeit not limited to it, are products of momentary cognitive activity and intrinsically related to use and context.

Metaphoricity is a phenomenon of use, an attribute that emerges from the dynamic process of cognitive activation. This process may be inductive, such as the use of metaphors by a teacher in an instrument class, or it may occur intuitively, such as the case described at the beginning of the article. Recognizing metaphor as a momentary cognitive activation allows us to consider metaphoricity as variable, and thus the activation process enabling the construction of metaphorical inferences is selective and dependent on the encyclopedia of the subjects' experiences. Therefore, it is plausible to affirm that metaphoricity is not necessarily activated when a metaphor is employed.

The metaphor as a byproduct of an established metaphoricity emerges from the process that Wittgenstein calls "seeing-as" (Müller, 2008). The philosopher uses this concept to illustrate a process that takes place when the orientation of our perception transforms one object into another. The example given by Wittgenstein is that of an image that may be seeing-as a rabbit or a duck. Interpretation changes depending on the orientation of perception. "Seeing-as" is to notice similarities between distinct objects which enables us to see them differently. Even if the object is the same, our interpretation changes and consequently so does our perception. Wittgenstein (1999) calls this experience "to perceive an aspect." When we notice an aspect, a figure can change before our eyes. The change of an aspect within the experience expresses itself in a new perception of the same object, such as the rabbit and duck; or of the same action, as the case of the bow and hammer. By "noticing an aspect," our perception changes. In addition to noticing shapes and colors, we may identify an organization as an outcome of our pursuit for coherence, as if we had suddenly solved a riddle.

"Seeing-as" is always an interpretive process and, as such, the construction of a meaningful object. "The process of establishing metaphoricity clearly involves a process akin to seeing-as, but a specific form of it, a form that will heuristically be paraphrased as 'seeing one thing in terms of another'." (Müller, 2008, p.24) Seeing-as is a specific mode of seeing and constructing meanings: it is seeing one thing in terms of another. In the construction of meaning in music, Wittgenstein's concept of seeing-as applies to perception and musical interpretation through the understanding that musical meaning is constructed by an interpretive process similar to "seeing-as," which Spitzer (2004) calls "hearing-as."

Play, Gesture and Metaphor

We believe that play, gesture, and metaphor, three central concepts in this article, are found in "seeing-as", or rather, in "feeling-as." Returning to the experience described earlier, we understand that the violist, when noticing, selecting, and projecting aspects of her experience with the hammer to the bow, experienced metaphoricity by feeling the bow as if it were the hammer. Even though the conduction of the bow was a familiar action, the interpretation of this gesture changed through the experience of metaphoricity, which may have prompted her to perceive the vitality of this action differently. Interpretation is not solely a subjective variety of conceptions regarding the same object or action, but it denotes the mode of being of understanding: to understand is to interpret and to interpret is to recreate (Gadamer, 2012). Since it is not limited to the reproduction, interpretation is circumscribed to a context and negotiates with aspects present in the object and with the subject's experiences. In Gadamerian hermeneutics, the other is a crucial element in understanding. The "other" is everything constituted as alterity, that is, as an element not immediately accessible, comprehensible.

Gadamer uses the concept of play (spiel) as the guiding thread for his model of understanding since he considers that, as a game, the entire process of constructing meanings is an open act and an interpretive action. It is circumscribed to a context and dependent upon the encounter and interaction with alterity. Much like play, the experience of "seeing-as" should not be approached as something that happens in the subject, but rather in the intersection with alterity. The understanding and experience of metaphoricity are not mere products of consciousness, but a negotiation process between myself and another. A game.

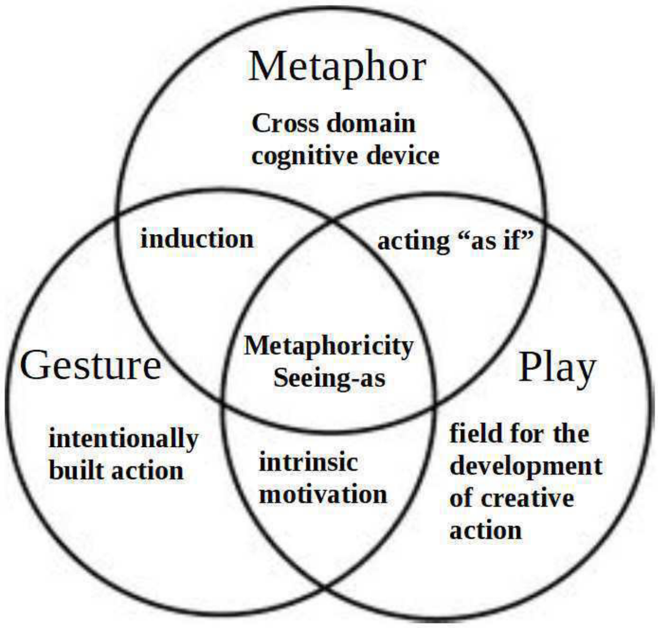


Figure 1. Diagram that shows the relations between play, gesture and metaphor

As a broad category of human action, the game generates a temporary sphere of activity with its own arrangement. By establishing an autonomous reality that entwines and absorbs the player, it creates an atmosphere in which we may experience things as if they were others (Huizinga, 2012, Gadamer, 2012). Gadamer (2010) acknowledges that a key feature of play is to act "as if." The behavior of the game establishes itself within the pursuit of a certain demarcation of free will, which attempts to determine a certain "targeted" behavior. "To act as if" is possible in every action with something in view, that is, any action with intentionality. We always have something in view in a game, which qualifies this activity as intrinsically motivated.

By obeying intrinsic interests, play and seriousness are deeply related and present in the players' behavior. An activity always circumscribed, to a greater or lesser extent, by fixed rules or conducts. This characteristic turns the game into a field in which human configurative action may emerge insofar that creating distinguishes itself from a moment of free contingency (Gadamer, 2010). For the game to happen it requires the indispensable presence of the element of the "other," which responds to or questions the player. This dialectic allows for the entrance "in a dynamic nexus that develops its own dynamics" (Gadamer, 2012(II), p. 180). The human game typically establishes goals for the players, and these typically pursue them by adjusting their movements. Play thus converges with a human trait that is indispensable for musical learning: the capacity to self-regulate one's behavior. To enter a game differs from other forms of behavior. Each game presents a task to the player who may experience a seemingly autonomous fluid alternation "What enters or is at play no longer depends on itself, but is dominated by this relationship called game." (Gadamer, 2012(II), p.154).

Conclusion

The domain of action established through play produces a fertile ground for the formation and development of musical actions. When we are attentive to the performance of an action, rather than its consequences or results, we are in the game's realm of action. Playing refers to any action performed attentively and directed toward itself, free of coercion, performed in a circumscribed space and time, experienced as a dynamic and imponderable process, previously circumscribed by more or less flexible rules, and establishing an autonomous reality. For the game to exist we must also have the presence of an element that responds or questions the player. We hypothesize the behavior present in play allows for the exchange between different domains of experience, as it happens in metaphorical thinking, and that play configures itself as an experiential field that enables the development of creativity and comprehension, allowing the transformation to and development of actions and behaviors.

Play creates a framework for action, establishing a field that allows for interpretive flexibility and engagement – you play when aware of what you do at the moment you do it. It is a key concept for the development of intentional actions and may contribute to the construction of cognitive models to explain our musical experience from a human-centered and embodied perspective. It is my belief that a next step should be taken in understanding how gestural experiences may be suggested by images and metaphors to develop teaching and learning actions motivated by the creation of a favorable pedagogical environment for the development of creativity and engagement in instrumental practice.

References

Clayton, M., & Leante, L. (2013). Embodiment in music performance. In: M. Clayton, B. Dueck & L. Leante (Eds) *Experience and meaning in music performance* (pp. 261–287). Oxford: The Oxford University Press.

Gadamer, H.G. (2010). *A hermenêutica da obra de arte* [The hermeneutics of the art-work]. São Paulo: Martins Fontes.

Gadamer, H.G. (2012). *Verdade e método* [Thuth and method]. (12nd ed). Petrópolis: Editora Vozes.

Gadamer, H.G. (2012). *Verdade e método* II: complementos e índice [Thuth and method II: complements and index]. (12nd ed). Petrópolis: Editora Vozes.

Gill, S.P. (2015). *Tacit engagement: Beyond interaction*. London: Springer

Huizinga, J. (2012). *Homo Ludens: o jogo como elemento da cultura* [Homo ludens: a study of the play-element in culture]. São Paulo: Ed. Perspectiva.

Ikuta, K. (1990). The role of "craft language" in learning "Waza". *AI & Society*, 4 (2), 137–146.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The embodied mind and its challenge to western thought*. New York: Basic Books.

Müller, C. (2008). *Metaphors dead and alive, sleeping and waking: a dynamic view*. Chicago: The University of Chicago Press.

Spitzer, M. (2004). *Metaphor and musical thought*. Chicago: The University of Chicago Press.

Stern, D. (2010). *Forms of vitality: exploring dynamic experience in psychology, the arts, psychotherapy and development*. Oxford: The University of Oxford Press.

Trevarthen, C., Delafield-Butt, J., & Schögler, B. (2011). Psychobiology of Musical Gesture: Innate Rhythm, Harmony an Melody in Movements of Narration. In: A. Gritten & E. King (Eds). *New Perspectives on Music and Gesture* (pp.11–44). Farnham: Ashgate

Varela, F., Thompson, E., & Rosch, E. (1991). *A mente corpórea: ciência cognitiva e experiência humana*. [The embodied mind] Lisboa: Instituto Piaget.

Wittgenstein, L. (1999). *Investigações filosóficas* [Philosophical investigations]. São Paulo: Nova Cultural

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Inhibitory Control in Transposing Musicians, Non-Transposing Musicians, and Non-Musicians

Álvaro M. Chang-Arana[1, 2] and Geoff Luck[2]

[1]*Design Factory, Aalto University, Finland*
[2]*Department of Music, Art and Culture Studies, University of Jyväskylä, Finland*

[1]alvaro.changarana@aalto.fi, [2]geoff.luck@jyu.fi

Abstract

This study examined potential differences in inhibitory control task performance among transposing musicians, non-transposing musicians, and non-musicians. Twenty-nine participants were recruited and categorized into three groups: transposing musicians (n = 9), non-transposing musicians (n = 10) and non-musicians (n = 10) depending on their previous skills and experience. Participants were presented with a music Stroop task and a classic Stroop task. In the music Stroop task, participants were required to select the name of a musical note written inside the (corresponding or different) whole-note on a staff. In the classic Stroop task, participants were presented with colour words written in a corresponding or different colour and had to choose the font colour while ignoring the written name of the colour. Both tasks took into account participants' background, including their previous knowledge of note-naming systems, clef-familiarity and mother tongue. Contrary to predictions, transposing musicians did not perform significantly differently from non-transposing musicians on the music Stroop task. Additionally, transposing musicians, non-transposing musicians, and non-musicians performed similarly on the classic Stroop task. Failure of the results to support the predictions suggests that future research might increase the sample size and quantify transposing skills in a different manner.

Introduction

Inhibitory control (IC) is one of the core executive functions (EF) which involves being able to "control one's attention, behaviour, thoughts, and/or emotions to override a strong internal predisposition or external lure, and instead do what's more appropriate or needed" (Diamond, 2013, p. 137). Previous studies have explored whether IC (and other EF) are susceptible to changes due to experience. In fact, musicians have been frequently targeted as one population of interest for this purpose. It is assumed that the highly cognitive and motor tasks demanded by music making and which musicians are involved with for extended periods of time, should translate into changes in different EF. Indeed, previous studies have explored how music interventions have impacted different EF measurements across different age groups when compared to groups of non-musicians or non-musical interventions (Bialystok & DePape, 2009; Bugos, Perlsetin, McCrae, Brophy, & Bedenbaugh, 2007; Janus, Lee, Moreno, & Bialystok, 2016; Moreno, Wodniecka, Tays, Alain, & Bialystok, 2014; Zuk, Benjamin, Kenyon, & Gaab, 2014).

However, the usual separation of participants as musicians or non-musicians have ignored a criterion which, we believe, might shed new light on EF, and, more specifically, on IC. That is, to the best of our knowledge, no previous EF study has targeted musicians according to whether they play a transposing instrument or not.

Music transposition is a highly complex process in which notated music to be played might not correspond to the sound to be produced (the new Grove dictionary of music and musicians, 1980). For instance, a horn player might have a "C" note printed on her score but produce an "F" when playing it. That is, there is a mismatch between the printed note and the actual note produced. Despite the intuitive conceptual similarities between IC and transposition, no previously reported research has explored whether transposing musicians differ from other musicians or non-musicians in their IC skills.

This raises several pertinent questions. Might it be possible to obtain behavioural differences in inhibitory control within musicians? Will transposing musicians perform better than non-transposing musicians on inhibitory control tasks? Might these differences be observable in a non-musician sample as well? And more generally, are there differences in inhibitory control task performance among transposing musicians, non-transposing musicians, and non-musicians?

Hypotheses

- Transposing musicians will outperform non-transposing musicians on a music Stroop task.
- Transposing musicians will outperform non-transposing musicians on the classic Stroop task.
- Transposing musicians will outperform non-musicians on the classic Stroop task.
- Non-transposing musicians will outperform non-musicians on the classic Stroop task.

Method

For a further description on this section, see Chang-Arana (2018).

Participants

Twenty-nine participants were recruited via social media, face-to-face invitation and email lists. They were further categorized into three groups: transposing musicians (mean age = 26.56, SD = 5.25, n = 9), non-transposing musicians (mean age = 28.50, SD = 8.30, n = 10) and non-musicians (mean age = 27.20, SD = 7.47, n = 10). Regarding their playing experience, transposing musicians (M = 18.22, SD = 7.17) and non-transposing musicians (M = 20.20, SD = 7.18) reported having played their main instrument for a similar number of years. Eight non-musicians reported considerably fewer years of playing experience (M = 4.75, SD = 7.09). However, none of them read music notation, whereas all of the musicians read music notation (n = 29). Transposing instruments most frequently reported were guitar (n = 3), French horn (n = 2) and trumpet (n = 2). Thirteen different languages were identified, the most frequently reported being Finnish (n = 11), Spanish (n = 3), and German (n = 3).

Initial questionnaire

All participants completed a 13-items questionnaire which allowed the identification of their main musical instrument, their knowledge of reading music notation, and their mother tongue. Based on these questions, participants were divided into the three groups described above.

Determining clef and note-naming system

To determine the most familiar clef and note-naming system for each participant, they were presented with a single sheet of paper with a bass, alto and treble clef printed on a pentagram. In every pentagram, a "B" whole-note was also placed. Participants were asked to point at the clef which were more familiar to them and then to say aloud the name of the note printed on the pentagram. Both answers were taken into account when tailoring the music Stroop task for each musician participant. By default, non-musicians were presented with the treble clef, but the name of the notes was determined according to participants' own experience.

Music Stroop task

The music Stroop task was designed primarily after Grégoire, Perruchet and Poulin-Charronnat's (2013) experimental setting, although it also incorporated elements from previous music Stroop tasks (Akiva-Kabiri & Henik, 2012; Stewart, Walsh, & Frith, 2004; Zakay & Glicksohn, 1985).

A whole-note containing a written note name was presented. The whole-note location and the name written inside could either coincide (i.e. congruent condition) or not (i.e. incongruent condition).

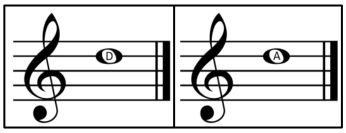


Figure 1. Example of congruent (left) and incongruent (right) conditions.

The music Stroop task was divided into a training session and an experimental session. The training session contained 20 stimuli (10 congruent and 10 incongruent), while the experimental session contained a total of 250 stimuli (125 congruent and 125 incongruent) divided by a self-dictated pause. Participants were asked to choose the note written name while ignoring the note symbol. Stimuli presentation was pseudorandomized and participants answered by pressing a laptop keyboard.

Classic Stroop task

The classic Stroop task was designed primarily after Stroop's (1935) experimental setting, although modified into a motor response format (MacLeod, 1991). A classic Stroop task was designed for every mother tongue of every participant.

Names of colours were shown as written words. These words could either be presented in a colour font which could match the written word (i.e. congruent condition) or not (i.e. incongruent condition). Participants were asked to choose the font colour while ignoring the written word. The classic Stroop task followed the same specifications of the music Stroop task previously described.

Implementation software and statistical analyses

Both Stroop tasks were implemented in PsychoPy2 Experiment Builder version 1.85.2 (Peirce, 2007). Statistical analyses were conducted in SPSS version 24.

Calculation of interference scores

A global interference score was calculated for the music Stroop task and the classic Stroop task scores. This global score took into account the accuracy of participants (i.e. the difference in mistakes between the congruent and the incongruent conditions) as well as their speed (i.e. the difference in speed between the congruent and the incongruent conditions). Therefore, it accounted for situations in which participants prioritized accuracy over speed, as well as situations in which participants prioritized speed over accuracy. The higher the score, the higher the performance in the corresponding Stroop task. The calculation was an adaptation of procedures reported in Scarpina and Tagini (2017).

Results

A Shapiro-Wilk test of normality was conducted on the global interference scores of the classic and music Stroop tasks. Both scores for every group of participants appeared normally distributed.

Table 1. Shapiro-Wilk normality tests for global interference scores of the music and colour Stroop tasks.

	M	*SD*	Statistic	*p*
GIM				
Transposing	.44	.06	.98	.21
Non-transposing	.41	.06	.93	.46
Non-musician	.51	.10	.90	.94
GIC				
Transposing	.47	.10	.97	.20
Non-transposing	.54	.09	.96	.83
Non-musician	.54	.10	.90	.91

Note: GIM = Global interference score of music Stroop task. GIC = Global interference score of classic Stroop task. N = 29 participants (transposing musicians = 9, non-transposing musicians = 10, non-musicians = 10).
$^*p < .05$.

An independent t-test revealed that transposing musicians (M = .44, SD = .06) did not differ significantly from the group of non-transposing musicians (M = .41, SD = .06) in their global interference music Stroop score, $t(17) = 1.10$, $p = .29$. However, a medium effect size was obtained, d = .50, 1 – β = .18. This might suggest that transposing musicians could obtain higher IC scores than non-transposing musicians on larger samples.

A one-way ANOVA revealed that transposing musicians (M = .47, SD = .10), non-transposing musicians (M = .54, SD = .09), and non-musicians (M = .53, SD = .10) did not differ significantly in their global interference score on the classic Stroop task, $F(2, 26) = 1.41$, $p = .26$. However, a small effect size was obtained, ω^2 = .17, 1 – β = .15. Therefore, larger samples could enable the detection of statistically significant manipulation effects.

Discussion

No differences were found in inhibitory control task performance among transposing musicians, non-transposing musicians, and non-musicians. However, several implications can be drawn from our results.

Although transposing musicians' performance was not significantly different from that of non-transposing musicians on the music Stroop task, the medium effect size detected suggests that a bigger sample size might lift the statistical significance of the results. Transposing musicians did score some decimals above non-transposing musicians in the music Stroop task (see Table 1). Thus there are some empirical grounds to support increasing the sample size. In a similar vein, the three remaining hypotheses were not supported either. That is, performance on the classic Stroop task did not differ according to whether participants were transposing musicians, non-transposing musicians or non-musicians. Again, the small sample size might have contributed to these findings.

In addition to the relatively small number of individuals studied, a second methodological limitation should be highlighted. Perhaps the way musicians were classified as transposing or non-transposing musicians was not optimal in determining their transposing skills? For instance, a pianists could be used to transpose as a result of her professional experience, even though the piano is not a transposing instrument. Therefore, an alternative method for establishing transposition skills is required. Future research should aim to establish more objective ways of measuring transposing skills. For example, rather than categorizing musicians as transposing vs. non-transposing, they could be ranked continuously on their transposing skills.

Finally, some reflections on the Stroop tasks. In addition to a composite score such as that utilized in this study, future research might report scores of accuracy and speed separately for each Stroop task. These measures would provide a wider picture of IC. Additionally, it would be relevant to conduct future studies which could test the psychometric properties of the Stroop tasks designed here. In other words, analyse whether scores from both Stroop tasks are actually measuring IC.

Conclusions

- There were no differences in inhibitory control task performance among transposing musicians, non-transposing musicians, and non-musicians.
- Although predictions were not met, data suggest that larger a sample size could enable the detection of statistically significant manipulation effects.
- Categorizing musicians as transposing or non-transposing musicians based solely on their primary instrument might be insufficient.
- Future studies could rate them on a continuous level of transposing skills rather than in a categorical manner.
- Future studies should test whether the Stroop tasks here presented are suitable for measuring IC.

Acknowledgements. We would like to thank the anonymous readers for their valuable reviews and recommendations. Additionally, we wish to extend our gratitude to SEMPRE award for facilitating our participation in this conference.

References

Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical Stroop effect in absolute pitch possessors. *Experimental Psychology, 59*(5), 272-278. doi:10.1027/1618-3169/a000153

Bialystok, E., & DePape, A. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 565-574. doi: 10.1037/a0012735

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging & Mental Health, 11*(4), 464-471. doi: 10.1080/13607860601086504

Chang-Arana, Á. M. (2018). *The impact of transposition skills on inhibitory control performance* (master's thesis). Retrieved from http://urn.fi/URN:NBN:fi:jyu-201806053012

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135-168. doi: 10.1146/annurev-psych-113011-143750

Grégoire, L., Perruchet, P., & Poulin-Charronnat (2013). The musical Stroop effect: Opening a new venue to research on automatism. *Experimental Psychology, 60*(4), 269-278. doi: 10.1027/1618-3169/a000197

Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology, 144*, 84-97. doi: 10.1016/j.jecp.2015.11.009

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203. doi:10.1037//0033-2909.109.2.163

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E., Cepeda, N., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22*(11), 1425-1433. doi: 10.1177/0956797611416999

Moreno, S. & Farzan, F. (2015). Music training and inhibitory control: A multidimensional model. *Annals of the New York Academy of Sciences, 1337*(1), 147-152. https://doi.org/10.1111/nyas.12674

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13.

Scarpina, F., & Tagini, S. (2017). The Stroop color and word test. *Frontiers in Psychology, 8*, 1-8. doi: 10.3389/fpsyg.2017.00557

Stanley, S. (Ed.). (1980). *The new Grove dictionary of music and musicians* (Vol. 19). Hong Kong: Macmillan Publishers.

Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics, 66*(2), 183-195. doi: 10.3758/BF03194871

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. doi: 10.1037/h0054651

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology/Revue Canadienne De Psychologie, 39*(3), 414-423. doi: 10.1037/h0080069

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS One, 9*(6), e99868. doi: 10.1371/journal.pone.0099868

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Film / Music / Narrative: A Multidimensional Mapping Processes

Juan Chattah[1]

[1] *University of Miami, United States*

[1] j.chattah@miami.edu

Abstract

In this study, I advance an analytical framework that (re)frames cross-modal and cross-domain correspondences as metaphorical mappings mediated by image schemas. Because music in film often acts as one agent within a multidimensional mapping that involves the visuals and the narrative, I draw exclusively on the film music repertoire. First, I unpack Lakoff & Johnson's (1980) Conceptual Metaphor theory and Johnson's (1987) Image Schema theory. Second, I apply these theories as investigative frameworks to explore multidimensional mappings within film, and survey related supporting research from cognitive and empirical musicology. Third, I review theoretical and empirical research that seeks to uncover the neural underpinnings of metaphor and image schema theories. Overlaying the metaphor and image schema frameworks onto the investigation of cross-modal and cross-domain correspondences within film provides numerous advantages—such approach reveals an (unobserved) implicit directionality of mappings, uncovers the intrinsic qualities of the domains that enable such mappings, expands the relevant body of research to include investigations on the neural underpinnings of metaphor and image schemas, and ultimately serves to lay an experiential foundation upon which broader testable hypotheses may be constructed.

Introduction

While much research explores cross-modal and cross-domain correspondences through the lens of cognitive and empirical musicology, this study frames these correspondences, particularly those embedded within film music, as *metaphorical mappings* mediated by *image schemas*. Overlaying the metaphor and image schema frameworks onto the study of this phenomenon allows us to:

- extend from correlations between sensory *modes* to correlations between broader *domains* of human experience—e.g., the semantic, affective, aesthetic, valuative domains;
- recognize the analogous gestalts present in the modes or domains entering the correlation;
- discern the implicit 'directionality' embedded in the correlation;
- draw on a wealth of research pertaining to the neural underpinnings of metaphor construction and metaphor comprehension, which may show possible approaches to explore the neural basis of cross-modal and cross-domain correspondences and may lead to a more comprehensive understanding of this phenomena.

In addition, using film music to explore these correspondences allows us to:

- enhance the ecological validity of this construct, by immersing participants in (near-)real-life settings;
- expand the components entering a correlation from spatio-kinetic modes to more abstract domains such as the film's narrative, the characters' affects, and beyond;
- delineate and allocate more precisely the participants' causal attributions, agency, and correspondence inferences—e.g., participants may not feel the 'increasing tension' but may experience a displaced affect that allows them to perceive the increasing tension in the music as increasing tension in the characters.

Conceptual Metaphor Theory

In its most basic form, a metaphor is a literally false comparison or categorization, such as "Her brain is a computer." Understanding metaphors is contingent upon our ability to foreground similarities between the compared elements and project relevant features between them (see Figure 1).

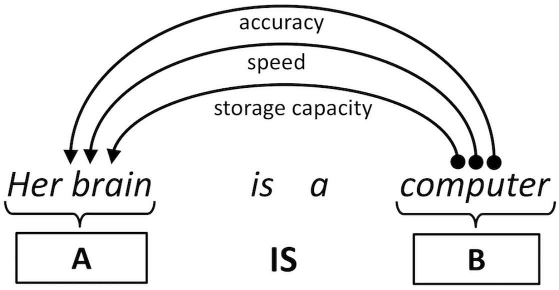


Figure 1. Projection of salient features in a metaphor.

Lakoff and Johnson (1980) argue that metaphorical thinking is the foundation of human thought processes and a "central building block for a more inclusive theory of human representations of reality" (26), and propose clustering single metaphors into larger, more inclusive, *conceptual metaphors*. For example, the metaphors "She has a fertile imagination," "Mathematics has many branches," and "His ideas have finally come to fruition" are bound by the underlying IDEAS ARE PLANTS conceptual metaphor indicating that ideas could be understood in terms of a plant's life cycle, its physical structure, or its ability to reproduce organically. Clustering single examples into conceptual metaphors clarifies the structure of metaphors—as [A] IS [B]—and foregrounds the embedded directionality—*from* [B] *to* [A]. (See Figure 1.)

Image Schema Theory

As a complement to the conceptual metaphor theory, and to reveal the fundamental similarity that enables metaphorical mappings, Johnson (re)formulated the notion of 'image schema.' Johnson (1987) defines an image schema as "a dynamic cognitive construct that functions somewhat like an abstract structure of an image and thereby connects together a vast range of different experiences that manifest this same recurring experience" (2). In other words, schemas are abstract-level mental structures that emerge from our perception of objects, events, or concepts, and from recurrent experiences of embodiment. For instance, the conceptual metaphor SOCIAL HIERARCHIES ARE JOURNEYS, instantiated in metaphorical sentences such as "She's climbing the social ladder" and "He became king of the hill" are primarily structured by the SOURCE, PATH, GOAL, and VERTICALITY image schemas.

Conceptual Metaphors and Image Schemas in Film Music

While the multidimensional mapping processes that include the music within a film are enabled by a variety of image schemas (CONTAINMENT, BALANCE, PATH, etc.), in this study I focus exclusively on the LINEARITY image schema. The LINEARITY schema highlights a *continuous one-dimensional* quality in the two domains that enter into a metaphor. Examples of general concepts structured by the LINEARITY image schema are age (young to old), height (low to high), or speed (slow to fast); analogously, musical concepts structured by the LINEARITY image schema are pitch (low to high), tempo (slow to fast), loudness (soft to loud), density (one layer to multiple layers), consonance (consonant to dissonant), and in some instances timbre (undistorted to distorted). The LINEARITY image schema, is best revealed while attending to correlations of *dynamic* musical parameters—gradual changes in pitch, tempo, or loudness—with *dynamic* spatial, kinetic, and affective modes/domains of human experience—gradual changes in spatial location, speed of movement, degrees of psychological tension. Nevertheless, these insights would benefit from attending to much research within music cognition exploring *static* parameters in both domains, e.g., the correlation of high/low pitch with high/low spatial location rather than raising/falling pitch with raising/falling spatial location.

Film music often elicits the [PITCH FREQUENCY] IS [VERTICAL SPACE] conceptual metaphor, where increasing pitch frequency correlates with upward motion, and decreasing pitch frequency with downward motion. Figure 2 depicts this conceptual metaphor in a scene from a *Tom and Jerry* cartoon. The scene shows Jerry falling from an airplane, stopping temporarily in the middle of his descent (thanks to a brassiere-parachute); correspondingly, the music outlines Jerry's fall through a descending chromatic figure with a brief pause when the parachute opens.

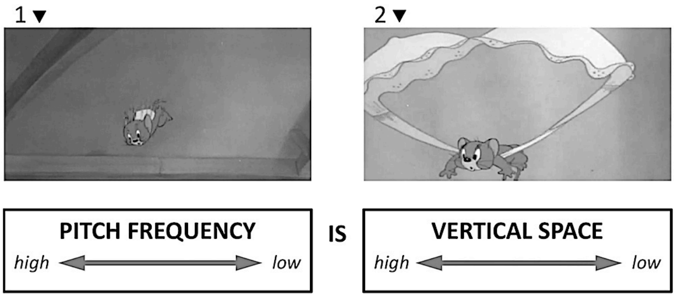


Figure 2. Mapping in *Tom and Jerry* cartoon.

Examples that depict the [PITCH FREQUENCY] IS [VERTICAL SPACE] conceptual metaphor are consistent with empirical findings on cross-modal correspondences between pitch and spatial elevation—high/low pitch and high/low vertical positioning (Cabrera & Morimoto, 2007; Chiou & Rich, 2012; Evans & Treisman, 2009; Rusconi et al., 2006), and raising/falling pitch and ascending/descending motion (Mossbridge et al., 2011; Küssner et al., 2014; Parise et al., 2014). Overlaying the conceptual metaphor framework, however, foregrounds a particular directionality in the correlation—[VERTICAL SPACE] circumscribes the [PITCH FREQUENCY], resulting in musically mapping the visual space. And, identifying the image schema that enables this correlation (i.e., the LINEARITY image schema) reveals the essential quality present in the domains entering in the correlation—a continuous one-dimensional quality.

Action movies frequently use the [TEMPO] IS [SPEED OF PHYSICAL MOVEMENT] conceptual metaphor, in which slow tempo correlates with slow movement in the visuals, and fast tempo with fast movement. This device creates momentum leading to the end of action sequences (car chase, fight scene, escape). Figure 3 shows this conceptual metaphor as it unfolds in a scene from *Crouching Tiger, Hidden Dragon*, where the increasing speed of drumming (an accelerando from 112bpm to 186bpm) pairs with the visuals to create a climactic point toward the end of a fight between the characters.

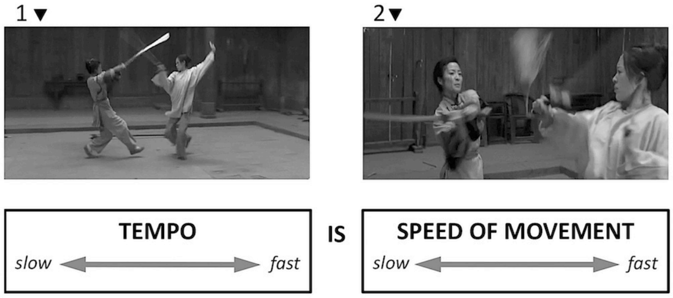


Figure 3. Mapping in *Crouching Tiger, Hidden Dragon.*

Although not as abundant as empirical research on cross-modal correspondences between pitch and spatial elevation, much empirical research shows a consistent correspondence between tempo and speed of movement (Camurri et al. 2005; Caramiaux et al. 2009; Eitan & Granot, 2006; Küssner et al, 2014). In this example, like in all examples in this study, the LINEARITY schema (a continuous one-dimensional quality) mediates the (cross-modal) correspondence between tempo and speed of movement. This example, however, extends beyond musical and spatio-kinetic modal correspondences, capturing the characters' psychophysiological state via the [TEMPO] IS [PSYCHOPHYSIOLOGICAL TENSION] conceptual metaphor. More significantly, this example arguably illustrates the *viewers*' psychophysiological reaction to the scene, warranting a change in the directionality, or direction of projections—[PSYCHOPHYSIOLOGICAL TENSION] IS [TEMPO]—a change that is only revealed by superimposing the conceptual metaphor framework onto the study of cross-domain correlations.

When moving away from mappings between the visuals and the music, and toward mappings between the (characters' or viewers') psychophysiological states and the music, the directionality unequivocally shifts, outlining a trajectory from the music (presented as the [B] component) to the psychophysiological state (presented as the [A] component). For example, in a scene from *Charlie and the Chocolate Factory*, Charlie is hoping to find the golden ticket to visit Willy Wonka's chocolate factory. A crescendo in the string instruments captures (and arguably elicits in the audience) the mounting tension as Charlie opens the chocolate, outlining the [PSYCHOLOGICAL TENSION] IS [LOUDNESS] conceptual metaphor (see Figure 4). Such an unambiguous shift in directionality may also indicate the boundary of what is regarded as Mickey Moussing—i.e., mapping physical movements onto sonic space—a technique of musical underscoring characterized by projections from the visuals onto the music.

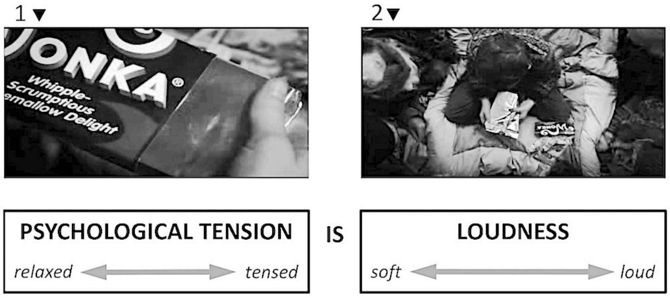


Figure 4. Mapping in *Charlie and the Chocolate Factory*.

Because (real) music is a multi-parametric construct designed to maximize the intended effect, most examples drawn from the film music repertoire illustrate the concomitant relations of musical parameters. A scene from the film *Soundless*, for instance, draws on numerous musical parameters to trigger a feeling of anxiety and anticipation, establishing a concomitant relationship between the [PSYCHOLOGICAL TENSION] IS [LOUDNESS], [PSYCHOLOGICAL TENSION] IS [PITCH FREQUENCY], [PSYCHOLOGICAL TENSION] IS [TIMBRAL DENSITY], and [PSYCHOLOGICAL TENSION] IS [CONSONANCE/ DISSONANCE] conceptual metaphors (see Figure 5).

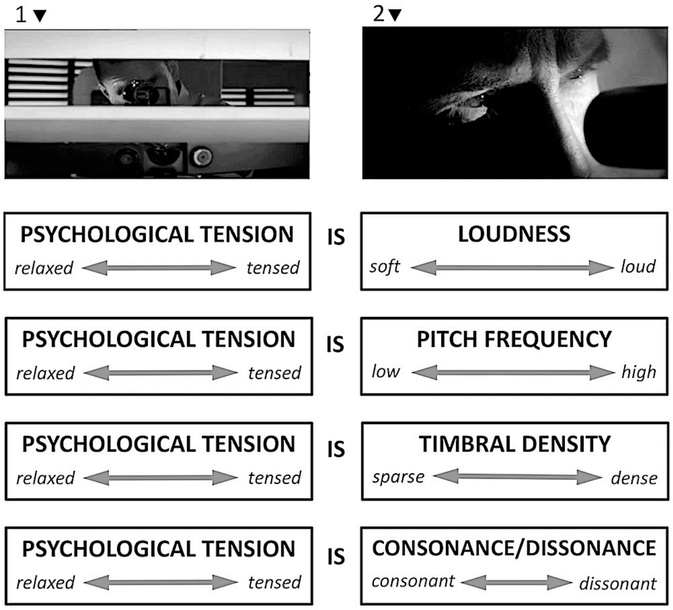


Figure 5. Mappings in *Soundless*.

Eitan and Granot (2006) found substantial evidence that supports cross-modal as well as intra-modal relations in perception; they suggest that "the musical dimensions of loudness, pitch, and tempo seem to interact via concomitant intensity levels or contours . . . [for instance] a pitch rise, a crescendo, and an accelerando are commonly considered intensifying . . . as are speeding up or ascending in the domain of human motion" (225). Additionally, related to cross-domain correspondences, van der Zwaag and colleagues (2011) found that the interplay of music parameters (tempo, mode, and percussiveness) plays a fundamental role in modulating affect and emotion. As a result, film music that attends to these concomitant variations may serve to reinforce the desired effect.

The last example, from *Soundless*, extends beyond the often-explored correlations that involve pitch, tempo, and loudness, to include timbre and texture. These, and many other musical parameters (articulation, consonance/dissonance, formal design, etc.) have not yet had much traction within empirical research. Nonetheless, these (neglected) musical parameters and their potential for establishing cross-modal and cross-domain correlations have been thoroughly explored through the speculative lens of metaphor theory (Bourne, 2015; Cox, 2016; Johnson & Larson, 2003; Saslaw, 1996; Zbikowski, 2002), particularly as manifested within film music (Chattah, 2006, 2015a, 2015b; Clark, 2018). Such interdisciplinary approach may spark future empirical explorations of cross-modal and cross-domain correspondences, while presenting opportunities to enhance the ecological validity of studies, by immersing participants in (near-)real-life settings. Furthermore, because the empathic response to art works may entail *self*-oriented reactions, *other*-oriented reactions, or both (Kesner & Horáček, 2017; Molnar-Szakacs & Uddin, 2013), the use of film music (which is embedded within a narrative dimension), enables researchers to design studies that involve valence, emotions, attitudes, etc., that relate to the participants' perception of the characters' emotions rather than the participants' experience of their own emotions, thus closely tracing a possible displacement of affect.

The Source of Cross-Modal and Cross-Domain Correspondences and The Neuro-Psychology of Metaphors

Audience members and researchers alike would likely agree on the consistency and pervasiveness of the mappings addressed in this study; while audience members draw on introspection to flesh out these mappings, researchers draw on the scientific method to establish consistency of these correspondences. There is wide disagreement, however, on the underlying mechanism that enables these mappings. What is the source of the tendency to establish mappings? Why and how are we predisposed to this tendency? To what extent are mappings innate? Are certain mappings pre-linguistic? Are mappings acquired by means of statistical learning? Does this tendency have a single origin or multiple origins? To what extent is the tendency to establish mappings ecologically advantageous? Do language and semantic processing play roles in establishing consistent mappings? Are consistent mappings the product of cultural practices? Do particular mappings vary across cultures or populations? Can we find the tendency to establish mappings in non-human primates or other species?

To address these and other questions, diverse behavioral and psychological experimental paradigms have been applied within cognitive and empirical musicology, including reaction time, priming, motion imagery, preferential looking in infants, forced-choice matching tasks, goodness-of-fit, real-time motion, and facilitated/inhibited motor responses. Only a handful of studies, however, have used fMRI, MEG, EEG/ERP, or TMS to measure or tinker with brain activity to gain a neural understanding of cross-modal and cross-domain correlations. (For a survey of the research literature on cross-modal and cross-domain correspondences, see Eitan 2013 and 2017.)

Research on the neural underpinnings of metaphor, on the other hand, is abundant. Two potentially complementary (and somewhat overlapping) experimental strands have emerged, both attempting to explain how metaphorical mappings are manifested at the neuronal level. One strand favors the notion that mental models of concrete domains are used as prototypes to build mental models of abstract domains (Boroditsky & Ramsca, 2002; Gentner & Stevens, 1983; Johnson-Laird, 1983). The other, more biologically grounded strand, argues that both domains in a metaphor (the abstract and the concrete) share neural substrates—this sharing process was coined by Anderson as "neural reuse." Anderson's neural reuse theory, which has gained significant traction recently, maintains that "conceptual-linguistic understanding might involve the reactivation of perceptuomotor experiences," which overlaps nicely with "the general idea behind direct neural support for metaphorical mappings, whereby understanding in one domain would involve the reactivation of neural structures used for another" (Anderson, 2010).

Conclusion

By overlaying the metaphor and image schema frameworks onto the investigation of cross-modal and cross-domain correspondences within film music, this study offers a glimpse into three chief benefits. First, exploring these correspondences as metaphorical mappings reveals new phenomenological dimensions, such as the implicit directionality of mappings and the intrinsic qualities of the domains that enable such mappings. Second, music within film generally relies on widely accepted (perhaps encultured) correlations; hence, corpus studies that attend to metaphorical mappings that include film music may serve as starting points to construct broader, testable hypotheses within cognitive and empirical musicology. And last, the investigation of this phenomenon greatly benefits from engaging in a more fluid conversation between research in cognitive/empirical musicology and research on the neural underpinnings of metaphor.

Acknowledgements. This work has been supported by the Center for the Humanities, University of Miami. The author extends his sincere gratitude to April Mann, for her thoughtful suggestions, and to Israel Araiza, for his caring support.

References

Anderson, M. L. (2010) Neural reuse: A fundamental organizational principle of the brain. *Behavioral and brain sciences 33*(4):245-66.

Boroditsky, L. & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological science 13*(2), 185-88.

Bourne, J. (2015). *A Theory of Analogy for Musical Sense-Making and Categorization: Understanding Musical Jabberwocky*. (Doctoral dissertation). Northwestern University, USA.

Cabrera, D., & Morimoto, M. (2007). Influence of fundamental frequency and source elevation on the vertical localization of complex tones and complex tone pairs. *The journal of the Acoustical Society of America, 122*(1), 478-488.

Camurri, A., Volpe, G., De Poli, G., & Leman, M. (2005). Communicating expressiveness and affect in multimodal interactive systems. *Ieee Multimedia, 12*(1), 43-53.

Caramiaux, B., Bevilacqua, F., & Schnell, N. (2009, February). Towards a gesture-sound cross-modal analysis. In *International Gesture Workshop* (pp. 158-170). Springer, Berlin, Heidelberg.

Chattah, J. (2006). *Semiotics, pragmatics, and metaphor in film music analysis*. (Doctoral dissertation). Florida State University, USA.

Chattah, J. (2015a). *David Shire's The Conversation: A film score guide*. Lanham, Maryland: Rowman & Littlefield.

Chattah, J. (2015b). Film music as embodiment. In M. Coëgnarts & Peter Kravanja (Eds.), *Embodied cognition and cinema*, 81-114.

Chiou, R., & Rich, A. N. (2012). Cross-modality correspondence between pitch and spatial location modulates attentional orienting. *Perception, 41*(3), 339-353.

Clark, E. A. (2018). *Harmony, associativity, and metaphor in the film scores of Alexandre Desplat*. (Doctoral dissertation). Victoria University of Wellington, New Zealand.

Cox, A. (2016). *Music and embodied cognition: Listening, moving, feeling, and thinking*. Bloomington: Indiana University Press.

Eitan, Z. (2013). How pitch and loudness shape musical space and motion. In S. L. Tan, A. J. Cohen, S. D. Lipscomb & R. A. Kendall, (Eds.), *The psychology of music in multimedia*, 165-191.

Eitan, Z. (2017). Cross-modal experience of musical pitch as space and motion. In C. Wöllner (Ed.), *Body, sound and space in music and beyond: multimodal explorations*, 49-69. London: Taylor & Francis.

Eitan, Z., & Granot, R. Y. (2006). How music moves: musical parameters and listeners images of motion. *Music perception, 23*(3), 221-248.

Evans, K. K., & Treisman, A. (2009). Natural cross-modal mappings between visual and auditory features. *Journal of vision, 10*(1):6, 1-12.

Gentner, D. & Stevens, A. L. (Eds.). (1983) *Mental models*. Lawrence Erlbaum Associates.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

Johnson, M., & Larson, S. (2003). Something in the Way She Moves. *Metaphor and symbol 18*(2), 63-84.

Kreplin, U., & Fairclough, S. (2015). Effects of self-directed and other-directed introspection and emotional valence on activation of the rostral prefrontal cortex during aesthetic experience. *Neuropsychologia, 71*(C), 38-45.

Küssner, M. B., Tidhar, D., Prior, H. M., & Leech-Wilkinson, D. (2014). Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in psychology, 5*, 789.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Molnar-Szakacs, I., & Uddin, L. Q. (2013). Self-processing and the default mode network: interactions with the mirror neuron system. *Frontiers in human neuroscience, 7*, 571.

Mossbridge, J. A., Grabowecky, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual–spatial attention. *Cognition, 121*(1), 133-139.

Parise, C. V., Knorre, K., & Ernst, M. O. (2014). Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences, 111*(16), 6104-6108.

Rusconi, E., Kwan, B., Giordano, B. L., Umilta, C., & Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition, 99*(2), 113-129.

Saslaw, J. (1996). Forces, Containers, and Paths: The role of body-derived image schemas in the conceptualization of music. *Journal of Music Theory, 40*(2), 217-243.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: The role of cortical midline structures and mirror neurons. *Trends in cognitive sciences, 11*(4), 153–157.

van der Zwaag, M. D., Westerink, J. H., & van den Broek, E. L. (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae, 15*(2), 250-269.

Zbikowski, L. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. New York: Oxford University Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Music, Space and Body: the evolutionary history of vocal and instrumental music

Yong Jeon(YJ) Cheong,[1] and Udo Will[2]

[1,2] *Cognitive Ethnomusicology, The Ohio State University, USA*

[1]cheong.23@osu.edu, [2]will.51@osu.edu

Abstract

Through musical activities, our everyday experience of time and space turns into something special as many music researchers have noted. In ethnomusicology, this transformative power has been considered as one of the most important universals in human music-making. We can distinguish two different modes of music making, namely, singing and instrument-playing, and it is of interest to know whether these two modes transform our experience of the world differently.

As music has been claimed to be an art of time, space has been largely ignored in music research, despite its importance in music performance. Here, we review several aspects of both body space and peripersonal space that relate to the singing and instrument-playing bodies. Body space consists of postural and surface-related (i.e. superficial) schemata that are associated with proprioception and touch respectively. Peripersonal space is characterized by 1) multisensory integration, 2) body-part centered specificity, 3) sensorimotor coupling and 4) plasticity. The singing body is associated mainly with postural schema and lacks some features of peripersonal space (e.g. plasticity). In contrast, the instrument-playing body clearly demonstrates the involvement not only of the two schemata of body space but also all features of peripersonal space. We propose that, therefore, in the evolution of music, the use of musical instruments transforms our spatial experience by integrating both body and peripersonal spaces.

Introduction

Music transforms our everyday experience of the world. People have reported either a loss of sense of time and space or being in other world while listening to music (see Gabrielsson, 2011). Patel (2010) argued that music as a transformative technology of the mind transforms human experience of the world. Ethnomusicoloigsts have considered the transformative power of music as one of the most important universals in human music-making (McAllester, 1971). In terms of music performance, there exist two different modes of music-making: singing and instrument-playing. This raises the question of whether there is any difference in the transformation of human experience of the world between singing vs. instrument-playing.

Differential auditory processing of vocal and non-vocal sounds has been found in neurophysiological (e.g. voice-sensitive response in Levy et al., 2001, 2003) and neuroimaging studies (Belin et al., 2000, 2002; Belin & Zatorre, 2003). Exploring differential processing of vocal and instrumental sounds into the temporal domain, Hung (2011) and Klyn et al. (2015) demonstrated that vocal vs. instrumental rhythms are processed and memorized differently. With regard to the origins of music, Fitch (2005, 2006), in his comparative studies, has argued that co-evolution of vocal and instrumental music is a human-specific phenomenon. This is because we are the only species using both vocal and non-vocal sounds simultaneously and interchangeably (e.g. whistle language, talking drum, & musical onomatopoeia) whereas other animals communicate either vocally (e.g. bird or whale song) or non-vocally (e.g. chimpanzee bimanual drumming).

In sum, the different modes of music-making that potentially explain some aspects of the origin of music have been explored primarily in terms of time. To the best of our knowledge, there is no research investigating whether there is differential processing of space between singing and instrument-playing. As Schäfer et al. (2013) noted, space has been largely ignored in music research despite its importance in music performance. In this report, therefore, we review the relationship of space and music-making by examining different spaces pertaining to the musicking bodies. This will give us a better understanding not only of how music transforms our experience but also of the different evolutionary trajectories of music-making and the development of human cognition.

Body and spaces

We experience the world through our body. We cognitively (re)construct time and space in a meaningful way with sensory inputs. It is mysterious how we experience the world as integrated mental representation although each sensory input contributes information that is specific to its own modality. Therefore, it is of interest to know underlying mechanisms of how information retrieved from different sensory organs is processed, integrated in our mind, and how we build our own world.

When it comes to immediate perception of space, the body serves as an egocentric reference frame and defines different spaces. Among several distinctions of psychological space pertaining to body, the body space and the peripersonal space seem to play important roles with regard to two types of musicking bodies, namely, singing and instrument-playing bodies.

Body space

Body space consists of postural and superficial schemata that primarily rely on proprioception and touch respectively. After examining patients with sensory disturbances, Head & Holmes (1911-2) first applied the term 'schema' with regards to perception of the body. They argued that the schema is characterized by un- or sub-consciousness of body posture and movements, which was called postural schema. This schema primarily relates to proprioception.

> For this combined standard, against which all subsequent changes of posture are measured before they enter consciousness, we proposed the word "schema." By means of perpetual alternations in position we are always building up a postural model of ourselves which constantly changes. Every new posture

> or movement is recorded on this plastic schema, and the activity of the cortex brings every fresh group of sensations evoked by altered posture into relation with it. (Head & Holmes, 1911-2 p. 187)

Following the observation of a patient who could not report his hand position but was able to localize stimulated spots on his body successfully, Head & Holmes (1911-2) presumed the existence of another schema in addition to postural schema. For this surface-related one, Paillard (1999) proposed the notion of superficial schema that is associated with a central, somatotopic mapping formed by tactile information from the body surface.

Peripersonal space

External space outside the body space is not homogenous. Among a division of external spaces, peripersonal space that immediately surrounds the body not only serves a defensive function to protect the body but also plays an important role in goal-directed actions and haptic perception. Heini Hediger (1968), a biologist and director of the Zurich zoo, noted an existence of this special zone around body from his observation that an animal escapes when its predator or enemy approaches within a certain distance that he called escape or flight distance. In contrary, human music-making is associated with the other function of peripersonal space. Peripersonal space is characterized by four features as follows:

1) Multisensory integration. The term 'peripersonal' originates from a series of eletrophysiological studies by Rizzolatti et al. (1981) who discovered the existence of multimodal neurons responding to both tactile and visual stimuli in the arcuate sulcus of macaque monkey's brain. They found that these neurons are activated by stimuli in the space within the animal's reaching distance and called them peripersonal neurons. Since Rizzolatti et al. (1981)'s seminal work, studies from different disciplines have presented a converging evidence that demonstrates representation of both body space and peripersonal space in connection with multisensory integration (for reviews Maravita, et al., 2003; Brozzoli et al., 2012)

2) Body-part centered specificity. The early neuro-physiology studies (e.g. Hyvarinen & Pranen, 1974; Rizzolatti et al., 1981) indicated that tactile receptive fields (RFs) in the multimodal neurons are centered on specific body parts and the RFs overlap spatially with a visual and/or an auditory RFs. Although there are disagreements about how many body-part centered reference frames exist and which frame functions as a common one, it has been accepted that the peripersonal representation consists of multiple body-part specific reference frames.

Cohen & Anderson (2002) noted that there are body-centered, eye-centered, head-centered and limb-centered reference frames and proposed that these frames transform into one common reference system in the posterior parietal cortex. di Pellegrino & Làdavas (2015) pointed out that the brain constructs multiple and modifiable representations of space that are centered on different body parts including head, hand, and trunk which are also known as perihead, perihand and peritrunk spaces, respectively.

3) Sensorimotor coupling. This refers to that perceived objects or events in the peripersonal space can be represented in terms of possible actions. In other words, objects or events in peripersonal space guide movements of the body.

In connection with multisensory integration, Rizzolatti et al. (1981) pointed out multimodal neurons are involved in organizing sequences of movements. di Pellegrino & Làdavas (2015) specifically discussed that multimodal neurons in the putamen, ventral section of the intraparietal sulcus (VIP) of the parietal area and inferior area 6 have both motor functions and multisensory functions.

In connection with body-part centered specificity, Cohen & Anderson (2002) discussed that sensorimotor coupling involves many computations that include a transformation of different body-part centered reference frames into a common one.

With a meta-analysis approach, Clèry et al, (2015) constructed two distinct cortical functional networks of the peripersonal space. The first network, subserving the defensive function, contains area 7b and the anterior part of the intraparietal sulcus (AIP) of parietal regions that are coupled with F5 of the premotor area. The second network is associated with goal-directed action and consists of VIP and the F4 sub region of the premotor area.

4) Plasticity. The boundaries of the peripersonal space are plastic, flexible and dynamic. di Vignemont & Iannetti (2015) argued that two factors influence the peripersonal space representation. The first factor refers to the effect of emotion. For instance, patients with claustrophobia, a persistent and irrational fear of enclosed places or of being confined (*APA dictionary of psychology*, 2015), have a distorted representation of peripersonal space. Exploring the relation between anxiety and defensive space, Lourenco et al. (2011) and Coello et al., (2012) demonstrated that a perception of dangerous objects correlates to a shrinkage of peripersonal space.

Secondly, tool use is also a well-studied factor contributing to plasticity of peripersonal space. Iriki et al. (1996) found that a monkey's manipulation of a long rack extends animal's reaching distance, which further implicates stretched peripersonal space. In connection with sensorimotor coupling, motor knowledge of tool use allows its user to predict the spatial location of the tools as it is moved by linking limbs with tool position. Brown & Goodale (2013) argued that tool-related knowledge induces spatial adaptation near the tool.

Musicking bodies and space

A contemplation of the above discussion about body and the two spaces leads us to inquire whether there is any difference in body space and/or peripersonal space between singing and instrument-playing bodies. The singing body is mainly associated with postural schema of the body space. We un- or subconsciously control our vocal organs to produce vocal sounds. In contrast, playing an instrument involves both the body and peripersonal spaces. Additionally, it requires both postural and superficial schemata of the body space due to the important role of direct tactile contact with, and haptic exploration of a musical instrument. For the instrument-playing body, multiple sensory information, predominantly from auditory and tactile inputs, is constantly integrated. In many cases, hands play an important role and the perihand space can be expected to occupy a prominent position in the body-part centered specificity of the peripersonal space. In terms of sensorimotor coupling, an instrument-playing body interacts with a musical instrument in an action-perception feedback loop. Lastly, musical instruments, the most special tools that humans have ever invented, seem to alter space around our limbs and music instruments. As Brown & Goodale (2013) pointed out, instrument playing involves motor knowledge that is combined with specific spatial information near and on an instrument. Furthermore, some instruments (e.g. drum stick, string bow) extend perihand space.

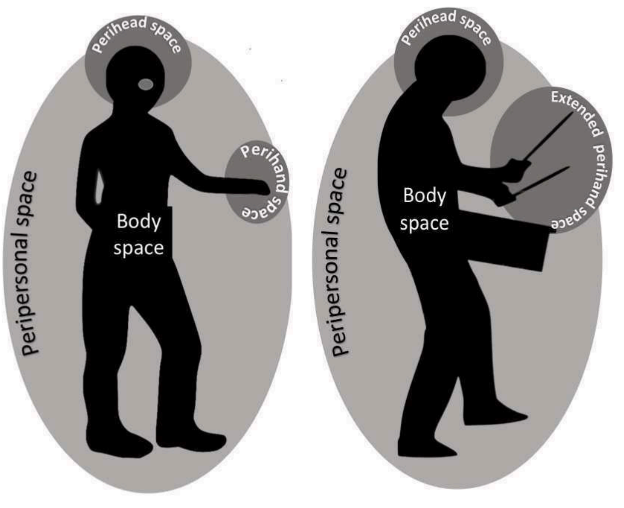


Figure 1. Spaces pertaining to singing vs. instrument-playing bodies

Different origins of musicking

Historically, a distinction of vocal and instrumental music has been made in various societies since the antiquity. The *Nāṭyaśāstra*, a Sanskrit treatise on the performing arts around the 4th century B.C. attributed to the sage Bharata, treats singing and instrument separately. Although there is no written document that clearly classifies vocal vs. instrumental music in East Asia, the names of various musical genres are indicative of whether it may be instrumental (e.g. Korean *Sanjo*, Chinese *Jiangnan sizhu*), vocal (e.g. Korean *pansori*, Chinese *Shuochang*), or combination of them (e.g. Japanese *sankyoku*). In *De institutione musica*, a foundation of the Western music theory and philosophy, Boethius (480–524) classified music into three categories: *musica mundane* (music of heaven), *musica humana* (music of human) and *musica instrumentalis* (music of instrument). The fact that ancient societies made such clear distinctions between vocal and instrumental music can be seen as an acknowledgment of their different relationship to the human body and, potentially, their different origins.

Then, is there any evidence that allows us to speculate about these two different ways of music-making in pre-historic time? One way is to look at archeological evidence and the other way is to compare different animal species. Morley (2013) reconstructed vocal and instrumental music-making in pre-historic time and discussed how different capabilities underlying the vocal and instrumental music-making might have contributed to the development of human cognition. Speculating about consequences of the anatomical changes observed in archeological data, Morley (2013) hypothesized that the physiological development of human vocal organs is associated with better pitch regulation and correlates with the expansion of the size of human brain. With regard to instrumental music-making in the prehistory, Cross (2002) proposed flint blades, as portable lithophone, could be one of the earliest sound tools. Studies on non-human species have shown that the loud/long calls of many primates are, at the climax, associated with increased motor activities in the limbs and body (Geissmann, 2000; Fitch 2011). In humans, playing an instrument, performed in parallel to singing, might have its origin in a dissociation of vocal and locomotor activities and the development of novel use of the latter.

Conclusion

Music transforms our experience of world. We make music in two ways. This review suggests that singing and instrument-playing may transform our experience of space differently. The primary difference between singing and instrument-playing is the recruitment of touch and haptic perception that are associated with different components of body space and peripersonal space. Signing does not make use of the superficial schema of the body space and it also does not require the use of tools that would modify peripesonal space. In contrast, playing an instrument involves both body space and peripersonal space. The singing body involves the use of both postural and superficial schemata of body space. The latter schema is established through touch. In contrast, the instrument-playing body relies on an ensemble of multisensory integration, body-centered specificity, sensorimotor coupling and plasticity of peripersonal space. Compared to singing that is related to part of these components (e.g. perihead space, audio-motor coupling), playing an instrument is more multisensorial due to audio-tactile integration and also requires an interaction between perihead and perihand spaces. Additionally, audio-tactile-motor coupling shapes the instrument-playing body and musical instruments can alter peripersonal space, which is associated with tool specific motor knowledge. Therefore we suggest that the transformation of spatial experience through instrument-playing is a characterizing feature of the evolution of music making.

References

VandenBos, G.R. (2015). *APA Dictionary of Psychology.* Washington, DC: American Psychological Association.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309-12.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*(1), 17-26. doi:10.1016/s0926-6410(01)00084-2

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14*(16), 2105-9. doi:10.1097/00001756-200311140-00019

Brown, L. E., & Goodale, M. A. (2013). A brief review of the role of training in near-tool effects. *Frontiers in Psychology, 4*, 1-5. doi: 10.3389/fpsyg.2013.00576

Brozzoli, C., Makin, T.R., Cardinali, L., Holmes, N.P., & Farne, A. (2012). Peripersonal space: a multisensory interface for body–

object interactions. In Murray, M.M., Wallace, M.T. (Eds.), *The Neural Bases of Multisensory Processes*, Boca Raton, FL: CRC press/Taylor & Francis.

Cléry, J., Guipponi, O., Wardak, C., & Ben, H. S. (2015). Neuronal bases of peripersonal and extrapersonal spaces, their plasticity and their dynamics: Knowns and unknowns. *Neuropsychologia, 70*(5), 313-26. doi: 10.1016/j.neuropsychologia.2014.10.022

Coello, Y., Bourgeois, J., & Iachini, T. (2012). Embodied perception of reachable space: how do we manage threatening objects?. *Cognitive Processing: International Quarterly of Cognitive Science, 13*(1), 131-5. doi: 10.1007/s10339-012-0470-z

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience, 3*(7), 553-62. doi: 10.1038/nrn873

Cross, I., Zubrow, E., & Cowan, F. (2002) Musical behaviours and the archaeological record: a preliminary study. In J. Mathieu (Ed.), *Experimental Archaeology. British Archaeological Reports International Series 1035*, 25-34.

Fitch, W. T. (2005). The evolution of music in comparative perspective. *Annals of the New York Academy of Sciences*, 1060(1) 29-49. doi: 10.1196/annals.1360.004

Fitch, W. T. (2006). On the biology and evolution of music. *Music Perception*, 24(1), 85-88. doi:10.1525/mp.2006.24.1.85

Fitch, W. T. (2011). The biology and evolution of rhythm: unravelling a paradox. In Rebuschat, P. Rohmeier, M., Hawkins, J. A., & Cross, I. (Eds.), *Language and music as cognitive systems* pp.73-95). Oxford: Oxford University Press. doi: 10.1093/ acprof:oso/9780199553426.001.0001

Gabrielsson, A. (2011). *Strong experiences with music: Music is much more than just music.* Oxford: Oxford University Press.

Geissmann, T. (2000). Gibbon Songs and Human Music from an Evolutionary Perspective. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music.* (pp. 103-24). Cambridge, Mass: The MIT Press.

Head, H., & Holmes, G. (1911 -1912). Sensory disturbances from cerebral lesions. *Brain, 34*, 102-254. doi: 10.1093/brain/34.2-3.102

Hediger, H. (1968). *The psychology and behaviour of animals in zoos and circuses*. New York: Dover Publications.

Hung, T. H. (2011). *One music? Two musics? How many musics? Cognitive Ethnomusicological, Behavioral, and fMRI Study on Vocal and Instrumental Rhythm Processing* (Doctoral dissertation). The Ohio State University, Columbus OH.

Hyvärinen, J., & Poranen, A. (1974). Function of the parietal associative area 7 as revealed from cellular discharges in alert monkeys. *Brain, 97*(4), 673-92.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurons. *Neuroreport, 7*(14), 2325-30.

Klyn, N.A.M, Will, U., Cheong, Y.J., & Allen, E.T. (2015) Differential short-term memorization for vocal and instrumental rhythms. *Memory, 24*(6), 766-91. doi:10.1080/09658211.2015. 1050400

Levy, D. A., Granot R., & Bentin S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport 12*, 2653–7 10.1097/00001756-200108280-00013

Levy, D. A., Granot, R., & Bentin, S. (2003). *Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology, 40*(2), 291-305. doi:10.1111/1469-886.00031

Lourenco, S. F., Longo, M. R., & Pathman, T. (2011). Near space and its relation to claustrophobic fear. *Cognition, 119*(3), 448-53. doi: 10.1016/j.cognition.2011.02.009

Maravita, A., Spence, C., & Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Current Biology, 13*(13), 531-9. doi: 10.1016/S0960-9822(03) 00449-4

McAllester, D. P. (1971). Some Thoughts on "Universals" in World Music, *Ethnomusicology, 15*(3), 379-80. doi:10.2307/850637

Morley, I. (2013). *The prehistory of music: Human evolution, archaeology, and the origins of musicality*. Oxford: Oxford University Press.

Paillard, J. (1999), Body schema and body image: A double dissociation in deafferented patients, in G. N. Gantchev, S. Mori, & J. Massion (Eds.), *Motor control, today and tomorrow,* (pp.197–214). Sofia: Academic Publishing House.

Patel, A. D. (2010). "Music, biological evolution, and the brain," In C. Levander & C. Henry (Eds.), *Emerging disciplines: shaping new fields of scholarly inquiry in and beyond the humanities* (pp. 91–144.) Houston, TX: Rice University Press.

di Pellegrino, G., & Làdavas, E. (2015). Peripersonal space in the brain. *Neuropsychologia, 66*, 126-33. doi: 10.1016/j.neuro psychologia.2014.11.011

Rizzolatti, G., Scandolara, C., Matelli, M., & Gentilucci, M. (1981). Afferent properties of periarcuate neurons in macaque monkeys. II. Visual responses. *Behavioural Brain Research, 2*(2), 147-63. doi: 10.1016/0166-4328(81)90053-X

Schäfer, T., Fachner, J., & Smukalla, M. (2013). Changes in the representation of space and time while listening to music. *Frontiers in Psychology*, 4. doi: 10.3389/fpsyg.2013.00508

de Vignemont, F., & Iannetti, G. D. (2015). How many peripersonal spaces?. *Neuropsychologia, 70*(1), 327-34. doi: 10.1016/j.neuro psychologia.2014.11.018

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Comparison of Expressed on Chinese Pipa Performance Motion

TIAN CONG[1], YUKI MITO[1], YUKITAKA SHINODA[2], MASANOBU MIURA[3], HIROSHI KAWAKAMI[1]

[1] *College of Art, Nihon University, Japan*

[2] *College of Science and Technology, Nihon University, Japan*

[3] *Department of System & Information Engineering, Hachinohe Institute of Technology, Japan*

[1] tanegi0321@gmail.com

Abstract

Emotional expression in music performance is one of the important points in performance. However, this emotional expression method is a "sensation" behavior that a performer gained from experience. If you are a musician with a certain level of skill, you can understand this "sense". Therefore, we want to present indices for everyone to aim for better performance by examining the relationship between "emotion", "performance motion" and "sound". In this study, we examined emotional expression from performance motion of Chinese pipa using motion capture system. From the results, we investigated the difference in performance motion between Chinese instruments and Western musical instruments.

Introduction

We examined the motion of emotional expression of performing some instrument (Miura, 2012a,2012b, Mito, 2017a). This study examined the relationship between emotions and motions during the performance of Chinese pipa. The purpose of this study is to investigate the relationship between emotions and actions during music performance of non-Western musical instruments. In the previous study, it is suggested that emotional transmission is done in Chinese music as well as in Western music (Tian, 2015). This study examined whether the relationship between emotion and motion is the same as Western musical instruments.

Mito (2017b) studied Chinese pipa. As the results, performance motion of "emotionless" was small and slow. Also, frequencies and S.D. of each motion of "sadness", “happiness", and “tender" were almost the same. the performance motion of "Anger" showed that is fast and large. We found that the S.D. of active and negative emotion was high from the results of "anger" and "fear". On the other hand, the same active “happiness" did not have a high S. D. value. However, the value of "emotionless" seems to be able to distinguish emotions from S.D. of distance and frequency. From this, we suggested that Chinese pipa is also related to emotion and motion.

However, in this previous study, we experimented with a different melody. Therefore, in this study, we will examine with a similar melody.

Methods

Performer

One female performer played Chinese pipa. Performance motion was measured by using the motion capture system.

Experiment summary

We chose a melody with moderate tempo and difficult expression. The title of the music is “yi zu wu qu (tribal dance)”, which is shown only the melody in Figure 1. The Chinese pipa performer performed this melody by expressing five emotions (happiness, tenderness, anger, sadness, fear) used by Juslin and emotionless for the task.

Figure 1 Performance task of this study

Motion measurement

We explain the optical motion capture system. This system can calculate three-dimensional position data of a marker by photographing a reflection marker attached to the body with a plurality of infrared cameras (having an irradiation function of infrared rays). Two-dimensional data photographed by the camera is transferred to a PC for system control, and the data is analyzed by software (Cortex) to convert two-dimensional data into three-dimensional data in real time. As a result, the three-dimensional position data of the marker can be measured in real time. Three-dimensional position coordinates of markers attached to each part of the performer's body were acquired using an optical motion capture system (MAC 3D System, Motion Analysis) composed of twelve infrared cameras. The system set the capture frame rate to 1/100 [sec] and the shutter speed to 1/2000 [sec].

We used 30 markers on the body, 7 markers on the left hand, 7 markers on the right hand of the performer and 4 markers on the Chinese pipa, 48 markers in total. The marker position is indicated by a stick picture (Figure 2).

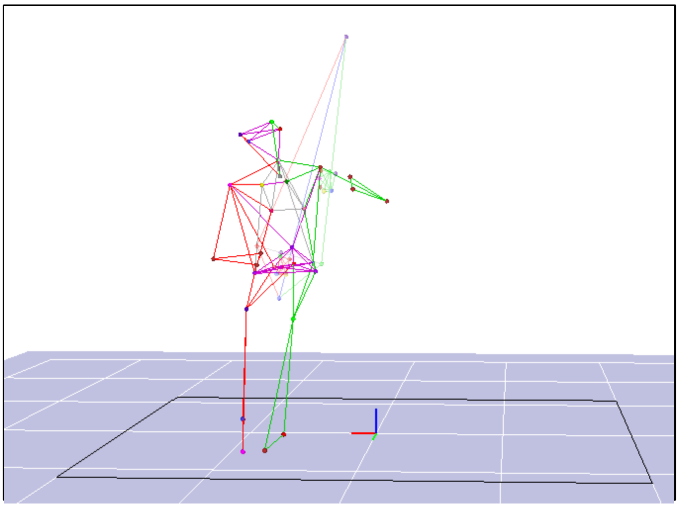


Figure 2. Markers adhesion position

Analytical approach

The analysis section was from the beginning of the first sound to the last sound. In this study, we tried to compare the motion fluctuation width of the upper body, for which we calculated the center of gravity of the upper body. This was calculated by modeling the upper body as a collection of 8 parts (head, torso, upper arms, forearms, hands), using the center of gravity position of each part Pgi(xg(i), yg(i), zg(i)) (i = 1, 2, ..., 8), the mass center ratio m(i) and the position data of each part of the body obtained from the motion capture data. The center of gravity position Pgi(xg(i), yg(i), zg(i)) of each body part was calculated using Eq. (1). Here, the positions Psi(xs(i), ys(i), zs(i)) are the start positions of each body part, and the positions Pei(xe(i), ye(i), ze(i)) are the end positions of each body part.

$$\begin{bmatrix} x_g(i) \\ y_g(i) \\ z_g(i) \end{bmatrix} = (1-m(i))\begin{bmatrix} x_s(i) \\ y_s(i) \\ z_s(i) \end{bmatrix} + m(i)\begin{bmatrix} x_e(i) \\ y_e(i) \\ z_e(i) \end{bmatrix} \qquad (1)$$

Results

Trajectory length of each body part

We examined the trajectory length of each part of the body (left hand, left forearm, left upper arm, head, right upper arm, right forearm and right hand) by the performance motion of each emotion. We showed the trajectory length of each part of six emotions in Figure 3- Figure 6.

In the graph, the horizontal axis is the body part and the vertical axis is the trajectory length.

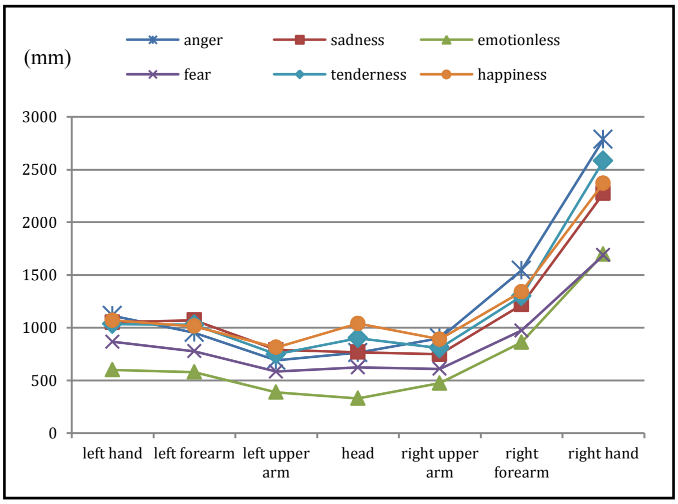


Figure.3 The left and right (X) direction of the trajectory length of each part of the body

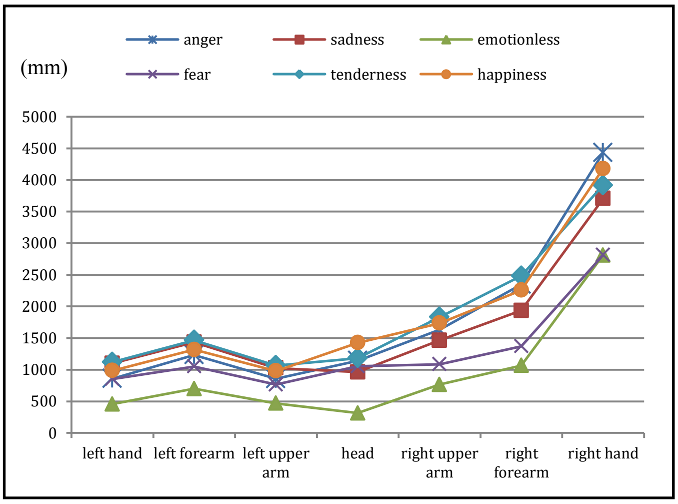


Figure 4. The front and back (Y) direction of the trajectory length of each part of the body

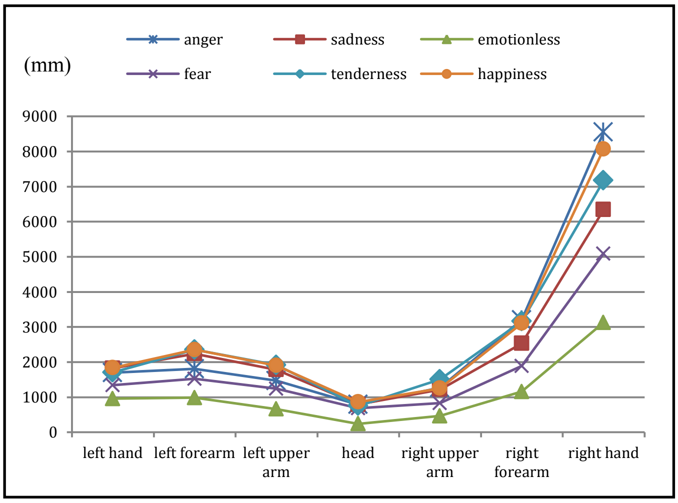


Figure 5. The vertical (Z) direction of the trajectory length of each part of the body

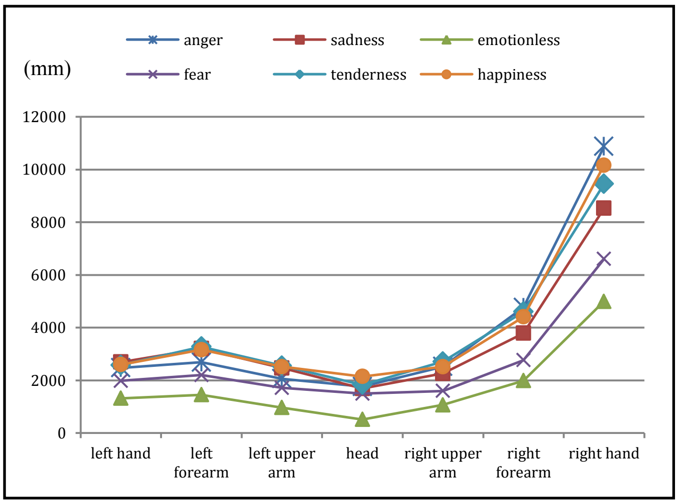


Figure 6. Three-dimensional (R) direction of the trajectory length of each part of the body

The performance of the Chinese pipa is the same as the guitar, holding the string with the left hand and flicking the string with the right hand. Therefore, the trajectory length of the body does not become symmetrical. The trajectory length was longer on the right hand than on the left hand. Especially the motion of the right forearm and the right hand is getting bigger.

We found that the performance of emotionless, fear and sadness is less motion in all graphs. On the other hand, we found that the motions of anger, tenderness, happiness emotions were great. These results were similar to the performance motion of the piano. Also, this result was similar to the previous study which examined in a different melody of Chinese pipa. From this, we thought that there is no influence on the relationship between emotion and performance motion depending on whether the melody is the same or different. Also, we found that emotion and performance motion were not influenced by the difference of instruments.

CONCLUSION

In this study, we examined the relationship between performance motion and emotion of Chinese pipa in the same melody. We analyzed the difference in emotional value from the trajectory length of the performance motion of the part of the body. As a result, we found that the positive emotional value has a long trajectory length of motion and negative emotions are short. The result was the same as the performance motion of the piano etc. In other words, in this study, we thought that Chinese instruments also express emotions similar to Western instruments.

REFERENCES

Juslin, P. (2001). Communicating emotion in music performance: A review and theoretical framework. Music and emotion., Oxford University Press., 309-337

Mito, Y., Kawakami, H., Miura, M. and Shinoda, Y. (2017). Relationship between performance motion of the grand piano and pianist by each emotion. Proc. The Asia-Pacific Society for the Cognitive Sciences of Music 6, 201-204

Mito, Y., Kawakami, H., Miura, and M. Shinoda, Y. (2017). RUDIMENTARY STUDY OF ASSOCIATION BETWEEN THE MOTION AND EMOTION BY THE CHINESE PIPA PERFORMANCE. Proc. The International Symposium on Performance Science 2017, 65-66

Miura, M., Mito, Y., and Kawakami, H. (2012). “Motion analysis for emotional performance of snare drums using a Motion Averaging Method”, Proc. of ACOUSTICS 2012, 2pMU9.

Miura, M., Mito, Y., and Kawakami, H. (2012). " A Motion Analysis Method for emotional performance on the snare drums", Proc. of 12th International Conference on Music Perception and Cognition(ICMPC), 682-685.

TIAN, C., Mito, Y., and Kawakami, H. (2015). " Emotional communication and motion analysis by Chinese pipa", Proc. of spring meeting of the Acoustical Society of Japan, 745-746.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

It's Only Rock 'n Roll, But I Like It: Computer Simulation Based on an Auditory Short-Term Memory Model Helps Explain Chord Rating Data

Lincoln G. Craton,[1] Jane Hyo Jin Lee,[1] Amanda Beauregard,[1] and Casey Lyons[1]

[1]*Stonehill College, USA*

[1]lcraton@stonehill.edu, [1]jlee7@stonehill.edu

Abstract

Listeners readily discriminate chords and tones based on their goodness of fit, that is, their appropriateness for a given musical context. An ongoing debate in music cognition concerns whether this ability requires learning conventional tonal regularities through exposure to music, or whether it emerges instead from the "bottom-up" operation of auditory processes. Computational modeling can help test these competing accounts. For instance, Bigand, Delbe´, Poulin-Charronnat, Leman, & Tillmann (2014) showed that Leman's (2000) bottom-up auditory short-term memory (ASTM) model simulates many behavioral and neurophysiological findings from empirical studies of chord processing. To our knowledge, the ASTM model has not yet been used to simulate chord rating data. Experiments from our laboratory provide a novel opportunity to do so (Craton, Juergens, Michalak, & Poirier, 2016; Craton, Lee, & Krahe, in preparation); these have documented high fitness ratings for chords used in rock music (II, bIII, III, bVI, VI, bVII) that lie outside the traditional harmonic hierarchy. In the present paper, we report that simulation with the ASTM model closely approximates fitness ratings for major target chords within and outside the basic diatonic set. The evidence suggests that listeners experience rock chords as normal components of a major key, and that they may do so without drawing on implicit knowledge of tonal regularities acquired through prior musical exposure.

Background

People know a good chord when they hear one. Once they recognize the musical key of a passage, listeners distinguish the basic set of harmonies—that is, "the seven triads built on the diatonic scale" (Krumhansl, 1990, p. 175)—from other chords. Adults tend to rate basic diatonic chords as better fitting (Krumhansl, 1990), and as generating less tension (Bigand & Parncutt, 1999; Bigand, Parncutt, & Lerdahl, 1996; Steinbeis, Koelsch, & Sloboda, 2006) and less emotionality (Steinbeis et al., 2006) than other chords. Even young children judge unfamiliar major-key chord sequences that end on the tonic chord (I) to be "good" more often than sequences ending on i (Corrigall & Trainor, 2009, 2010) or bII (Corrigall & Trainor, 2014).

Chord rating data also point to a hierarchy within the set of basic diatonic chords. In major key contexts, adults rate the tonic (I) as better-fitting (Krumhansl, 1990) and as generating less tension (Bigand et al., 1996) than both the dominant (V) and the subdominant (IV), which are in turn better-fitting and less tension-generating than the other basic diatonic chords (ii, iii, vi, and vii^{o}). Compared to IV, adults also rate passages ending on I as more complete (Bigand & Pineau, 1997; Tillman and Lebrun-Guillaud, 2006) and they more frequently judge I as belonging to a prior musical sequence (Bigand & Pineau, 1997). Finally, young children rate unfamiliar major-key chord sequences ending on I as "good" significantly more often than sequences ending on IV (Corrigall & Trainor, 2009, 2010).

In sum, listeners experience basic diatonic chords occurring in musical contexts as 1) better-fitting or more stable than many chords outside the basic diatonic set, and as 2) forming a hierarchy defined by their within-key fit or stability. However, framing the chord rating literature in this way raises a host of new questions. For starters, how sharp is the psychological division between basic diatonic chords that are low in the hierarchy and chords outside the basic diatonic set? Are all chords lying outside the basic diatonic set equally poor fitting and/or unstable? If not, why not?

Scholarship on popular mainstream music (hereafter, rock) has sharpened these questions. Several music theorists have pointed out that rock routinely employs many chords from outside the basic diatonic set (Moore, 2001; Stephenson, 2002; Tagg, 2014); this has been confirmed by a corpus analysis of rock harmony (de Clercq and Temperley, 2011; see also Acevedo, 2016). Evidently, basic diatonic chords are not quite as uniquely privileged as music cognition researchers used to think. Like the drunk who has lost his keys on a dark street but searches for them under the streetlamp because "that is where the light is," researchers may have formerly focused their attention on basic diatonic chords because this is what traditional music theory accounts do. For those interested in how listeners experience popular music, however, the scholarship on rock harmony has effectively brightened the streetlamp and broadened our focus beyond the traditional harmonic hierarchy.

In fact, recent chord rating data confirm that listeners experience many chords outside the basic diatonic set as musically appropriate (Craton, Juergens, Michalak, & Poirier, 2016). Craton et al. (2016) obtained surprise and liking ratings for major, minor, and dominant 7 target chords that followed a key establishing passage (a major scale + tonic major triad, or a major pentatonic scale + tonic major triad). They used Stephenson's (2002) rock harmony framework and unpublished data from the de Clercq and Temperly rock corpus (de Clercq & Temperley, 2011; D. Temperley, personal communication, August 13, 2013) to operationalize a "rock-typical" chord condition consisting of chords that are used in rock but that lie outside the basic diatonic set. The results showed that, particularly for major chord targets, liking ratings for basic diatonic chords (IV, V) and rock-typical chords (II, bIII, III, bVI, VI, and bVII) did not differ; both were higher than ratings for "atypical" major chords, that is, chords outside the basic diatonic set that are rarely used in rock (#IV, VII).

We recently replicated and extended these findings in two experiments with large online samples (Craton, Lee, & Krahe, in preparation). Participants were recruited via Amazon's Mechanical Turk (MTurk) and directed to an online experiment hosted by Qualtrics. In Experiment 1 (N = 199), participants provided liking ratings (1 = *dislike extremely*; 10 = *like extremely*) for major chords that followed a short key-establishing musical context (C major scale + C major triad), all played using the classical piano setting in Finale 2014. The 20 target chords were presented in root position and included every chromatic root from IV (the F below the tonic chord in the prior musical context) to I (the C an octave above the tonic chord in the prior musical context). The pattern of mean liking ratings closely replicated our earlier findings.

Experiment 2 (N = 188) was identical, except that participants rated how well each target "fit" with the preceding musical context. The results revealed a highly similar pattern, except that fitness ratings for the tonic were particularly high.

The Present Study

Given that listeners can discriminate chords based on how well they fit with a musical passage, how do they do it? Do they first need to learn tonal regularities through exposure to a particular musical system? Alternatively, might they rely exclusively on bottom-up processes operating on the auditory signal alone? Versions of both views are represented in the current literature (Bigand, Delbe´, Poulin-Charronnat, Leman, & Tillmann, 2014; Collins, Tillmann, Delbe´, Barrett, & Janata, 2014; Krumhansl & Cuddy, 2010; Large, Kim, Flaig, Bharucha, & Krumhansl, 2016; Milne, Laney, & Sharp, 2015; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010).

An example of the first view is Krumhansl's cognitive theory (Krumhansl, 1990; Krumhansl & Cuddy, 2010), which proposes that individuals exposed to their culture's music acquire implicit knowledge of its tonal relations, such as chord frequencies and transitional probabilities, which is stored in long-term memory and drawn on during subsequent listening. From the perspective of this sort of enculturation or statistical learning account, our chord rating data are perhaps unremarkable. If listeners learn the goodness of fit of chords from exposure to a particular musical system, then it is not surprising that people who listen to rock will perceive the chords it employs as better fitting than atypical chords! At best, from this point of view, our findings highlight the need for music cognition researchers interested in everyday musical experience to look beyond the basic diatonic set of chords: Western listeners may typically acquire a harmonic hierarchy that is more inclusive than is commonly thought (Acevedo, 2016; Craton et al., 2016; deClercq & Temperley, 2011).

On the other end of the spectrum from a pure statistical learning view is the notion that musical qualia can be computed "bottom-up" from the auditory signal alone, without prior exposure to music or abstract knowledge of tonal regularities (Bigand et al., 2014; Large et al., 2016; Milne et al., 2015). We find this perspective interesting because it points to an explanation of how harmonic conventions may arise in the first place. If the operation of the auditory system itself creates a perceptual ranking for the goodness-of-fit of chords, then this ranking would provide the harmonic palette from which composers and improvisers working in specific musical systems choose. Some types of music may tend to limit harmonic choices to the highest fitting chords, whereas other musical systems may extend the palette considerably. Rock's relatively liberal harmonic palette may simply reflect a more inclusive use of chords from this perceptually based hierarchy than is the case for more conservative musical systems, such as common-practice music.

Computational modelling provides a useful tool for testing the plausibility and relative strengths of these competing theories of tonal sequence processing (Bigand et al., 2014; Collins et al., 2014, Milne et al., 2015). Bigand et al. (2014) recently showed that Leman's (2000) bottom-up model—which had previously simulated human responses in classic probe-tone experiments (Krumhansl & Kessler, 1982)—is also successful in simulating behavioral and neurophysiological data from empirical studies of the processing of chord sequences. They concluded that "the explanatory power of the ASTM model may have been underestimated or misunderstood by the research community" (p. 20).

To our knowledge, the ASTM model has not yet been used to simulate chord rating data. We thus conducted simulations with the model using the experimental stimuli from the online replication study described above (Craton et al., in preparation). To the extent that the model performed like our human listeners, a bottom-up account would become increasingly credible as an explanation for the current findings.

Simulation: Leman's (2000) ATSM Model

The ATSM model is conveniently available online as a MATLAB toolbox (www.ipem.ugent.be/Toolbox). The model takes audio signals as input and computes representations at four main processing stages: 1) peripheral auditory system, 2) pitch periodicity analysis, 3) echoic memory, and 4) tonal contextuality (for more detailed descriptions and useful figures, see also Bigand et al, 2014; Collins et al., 2014).

The first stage transforms the audio signal into an auditory nerve image (ANI), simulating cochlear transduction of the audio signal into neural firing rate-codes across the different critical bands.

The second stage takes the neuronal firing probabilities in the ANI as input and estimates the periodicities in the firing patterns to produce a pitch image (PI), simulating the neural autocorrelation and pooling processes thought to underlie pitch processing in mammals (Cariani & Delgutte, 1996a,b; Schnupp, Nelken, & King, 2011).

The third stage takes the pitch periodicity pattern in the PI as input and uses "leaky" integration over time to produce two new pitch images. A local pitch image (LPI) based on a short integration time (0.1 s) represents the current musical event. A global pitch image (GPI) based on a longer integration time (1.5 s) represents neural information about the preceding musical context that has accumulated in ASTM or echoic memory.

The fourth and final stage determines the goodness of fit of a target event (the current chord or tone) with the preceding musical context by calculating its "tonal contextuality" (TC), which is the correlation coefficient (Pearson's r) between the LPI and the GPI sampled just after the target event occurs.

Method

We conducted the simulations by using the IPEM toolbox in MATLAB (Release 2015b, The MathWorks, Natick, MA). The key process of analyzing the audio data requires the Auditory Modeling library included in the toolbox. The auditory model is a C-library and had to be compiled to work with MATLAB 2015b (the IPEM toolbox available online was developed for an earlier version of MATLAB). We installed the MinGW compiler from TDM-GCC to build a MEX file that provided an interface between MATLAB and functions written in C.

After installing the necessary compiler and compiling the auditory model, the original 20 musical sequences from Craton et al. (in preparation) were presented as input to the model. Each sequence lasted 8 s (6 s key-establishing passage + 2 s target chord). The mean TC value for the final 2 s time frame during which the target chord played corresponds to the simulation's fitness rating for that target.

Results

A comparison plot of mean TC values from the simulation (where TC = 1 indicates the highest possible rating) and listeners' mean fitness ratings (1 = *fits poorly*; 10 = *fits well*) reveals striking similarity (see Figure 1). Kendall's W revealed statistically significant agreement between the model and listeners' ratings, $W = .899$, $p = .018$.

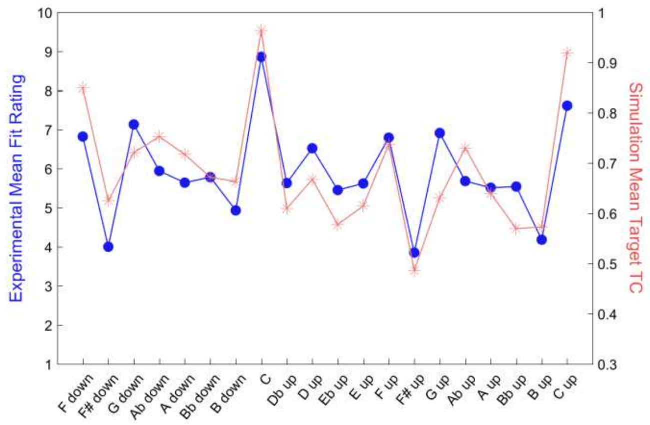


Figure 1. Comparison of listeners' mean fitness ratings (blue circles) and mean tonal contextuality values from the simulation (red stars).

Conclusion

The present simulation findings certainly do not rule out an important role for statistical learning in chord perception, but they do show that a bottom-up model by itself can account reasonably well for chord rating data. Thus, listeners may judge chord fitness without drawing on abstract knowledge of tonal regularities acquired through enculturation. We propose that bottom-up processes create a perceptual ranking of chord fitness for all chords. This hierarchy then provides the harmonic palette from which composers/improvisers in different musical systems may conservatively (common-practice) or liberally (rock) choose. By this account, the perceptually given harmonic hierarchy is relatively impervious to musical exposure—it remains in place and influences music perception throughout a listener's life.

Time will tell how well this simple speculative account holds up to empirical scrutiny. A logical way to start—because we are not wedded to any particular bottom-up model—is to compare the performance of the ASTM model with other biologically plausible bottom-up models. This would help narrow down plausible candidates for the perceptually based harmonic hierarchy we have in mind, which we regard as the starting point of chord perception. Next, further studies of chord perception can explore when this harmonic hierarchy emerges during development—we predict quite early—and test the hypothesis that the hierarchy is relatively unaffected by musical exposure. Finally, corpus analyses of harmonic conventions in different musical systems can test the hypothesis that chord choice is not arbitrary but rather shows a particular pattern based on the perceptually given harmonic hierarchy. Specifically, conservative systems such as common-practice music will limit themselves to the best-fitting chords near the top of the hierarchy; musical systems with a more liberal harmonic palette, in contrast, will include chords at the top of the hierarchy and systematically extend their harmonic palette down the hierarchy to a greater or lesser extent.

Acknowledgements. We wish to thank Joren Six, Micheline Lesaffre, and Marc Leman for their assistance in setting up the toolbox.

References

Acevedo, S. (2016, July). *Computational modelling of chord successions in popular music*. Paper presented at the 14th International Conference for Music Perception and Cognition (ICMPC14), San Francisco, CA, USA.

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience*, *8:94*. doi:10.3389/ fnsys.2014.00094

Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research, 62,* 237-254.

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception and Psychophysics, 41,* 519-524.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception and Psychophysics, 59,* 1098-1107.

Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones: I. Pitch and pitch salience. *Journal of Neurophysiology, 76,* 1698–1716.

Cariani, P. A., & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones: II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology, 76,* 1717–1734.

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: from audio signals to behavior. *Psychological Review*, *121*(1), 33-65.

Corrigall, K. A., & Trainor, L. J. (2009). Effects of musical training on key and harmony perception. *Annals of the New York Academy of Sciences*, *1169*, 164-168. doi:10.1111/j.1749-6632.2009.04769.x

Corrigall, K. A., & Trainor, L. J. (2010). Musical enculturation in preschool children: Acquisition of key and harmonic knowledge. *Music Perception*, *28*, 195-200.

Corrigall, K. A., & Trainor, L. J. (2014). Enculturation to musical pitch structure in young children: Evidence from behavioral and electrophysiological methods. *Developmental Science*, *17*(1), 142-158. doi:10.1111/desc.12100

Craton, L. G., Juergens, D. S., Michalak, H. R., & Poirier, C. R. (2016). Roll Over Beethoven? An initial investigation of listeners' perception of chords used in rock music. *Music Perception*, *33*, 332-343.

de Clercq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music*, *30*, 47-70.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Krumhansl, C. L., & Cuddy, L. L. (2010). A theory of tonal hierarchies in music. In M. Riess Jones, R. R. Fay, & A. N. Popper (Eds.), *Music perception* (Vol. 36, pp. 51–87). New York, NY: Springer. doi:10.1007/978-1-4419-6114-3_3

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89,* 334–368. doi:10.1037/0033-295X.89.4.334

Large, E. W., Kim, J. C., Flaig, N. K., Bharucha, J. J., & Krumhansl, C. L. (2016). A neurodynamic account of musical tonality. *Music Perception*, *33*, 319-331.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception, 17,* 481–509. doi:10.2307/40285830

Milne, A. J., Laney, R., & Sharp, D. B. (2015). A spectral pitch class model of the probe tone data and scalic tonality. *Music Perception*, *32*, 364-393.

Moore, A. (2001). *Rock: The primary text: Developing a musicology of Rock.* Aldershot: Ashgate Publishing.

Schnupp, J., Nelken, I., & King, A. (2011). *Auditory neuroscience: Making sense of sound*. Cambridge, MA: MIT press.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience, 18*, 1380-1393.

Stephenson, K. (2002). *What to listen for in rock: A stylistic analysis.* New Haven, CT: Yale University Press.

Tagg, P. (2014). *Everyday tonality II: Towards a tonal theory of what most people hear.* Huddersfield, NY: Mass Media Scholar's Press.

Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research, 70*, 345-358.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Analysis of Intonation in Unison Choir Singing

Helena Cuesta[1], Emilia Gómez[1,2], Agustín Martorell[1], Felipe Loáiciga[1]

[1]*Music Technology Group, Universitat Pompeu Fabra, Barcelona (Spain)*

[2]*Joint Research Centre, European Commission*

helena.cuesta@upf.edu

Abstract

Choir singers synchronize their fundamental frequency (f_0) and timing when they perform together. In this work, we investigate several expressive characteristics of choir singing, with special emphasis on unison performances, to study how singers blend together and interact with each other by analyzing their f_0 dispersion, intonation, and vibrato. We use the spectral autocorrelation algorithm to obtain the f_0 trajectories of each individual singer and compute the f_0 dispersion as the standard deviation of the f_0 distribution. Using the Pearson correlation, we study the synchronization of singers in terms of their intonation and vibrato. Together with this study, we present a new choir singing dataset of voices singing together but recorded individually using directional close microphones. This dataset contains the individual recordings of 16 singers from a semi-professional 4-part choir (SATB) performing 3 different songs.

We found the f_0 dispersion in unison sections to be in the range of 16 to 30 cents, and the largest average f_0 dispersion was found in the basses. We observed that the intonation correlation varies a lot and is around 0.32 on average in all cases. The slowest piece was the one with the highest percentage of notes with vibrato, which had a frequency around 5 Hz for the whole dataset.

1. Introduction

Choir singing is probably the most widespread type of singing (Sundberg, 1987). The European Choral Association, for instance, represents more than 2,5 million singers, conductors, composers, and managers in over 40 European countries, reaching more than 37 million people in Europe active in the field of collective singing. Although these numbers suggest the relevance of choir singing research, there have been few research studies focusing on voice ensembles compared to the ones addressing solo singing.

Choir singing and solo singing differ in several acoustic characteristics such as loudness or formant frequencies. In a study about choir singing by Zadig et al. (2016), they state that the voices of a choir are commonly meant to blend together, which means that when they sing together, the audience is supposed to perceive one single voice, even though there are a greater number of them. A very relevant review on choir acoustics was published by Ternström (2003), who discussed several related studies. Some of these include an analysis of the intonation of intervals (Lottermoser and Meyer, 1960) or his own research on fundamental frequency (f_0) dispersion (Ternström, 1993).

Intonation is a complex phenomenon (Devaney, 2011) that describes how a pitch is sung in tune (Dai and Dixon, 2017). It is commonly referred to as the accuracy of pitch in singing or playing with respect to a specific tuning system; for example, the most typical in Western music is the equal tempered tuning system. Fundamental frequency (f_0) dispersion is defined as the small deviations in f_0 between singers producing the same note at a unison performance. The agreement between all the voice sources is referred to as *degree of unison* (Sundberg, 1987): the larger the dispersion, the smaller the degree of unison. The investigation by Jers and Ternström (2005) is concerned with intonation, synchronization, and to what extent the singers of a voice section agree to each other. The authors carried out a multi-track recording of a 16-singer choir to analyze individual voices in terms of f_0 and found the dispersion to be between 25-30 cents[1]. Another popular study on barbershop quartets (Hagerman and Sundberg, 1980) reported very small f_0 dispersion values: 4,3 to 16,9 cents. More recent studies include the analysis of interactive intonation in vocal ensembles (Dai and Dixon, 2017), where authors investigate how singers of an unaccompanied ensemble interact with each other in terms of intonation (i.e. pitch accuracy). Dai and Dixon designed a novel experiment and tested the intonation accuracy of individual singers by pitch and interval error. They found that singers interact with each other and that this interaction influences their intonation.

In addition to the intonation and f_0 dispersion, a very relevant aspect of choir singing is vibrato. In the early 30s, (Seashore, 1931) carried out one the first studies on vibrato in singing using two approaches: a phono-photographic recording of musical performance and speech, and a psychophysical measurement on the perception of the vibrato produced synthetically by instruments. As a result of this study, he provided a definition of a vibrato as *"a periodic oscillation in pitch in which the extent of oscillation for the best singers averages approximately a half-tone and for string instruments approximately a quarter-tone"*. Seashore also claims that *"a good vibrato in music is a periodic pulsation, generally involving pitch, intensity, and timbre, which produces a pleasing flexibility, mellowness, and richness of tone"*. According to (Driedger et al., 2016), who proposed a template-based system for vibrato characterization, musicians extensively use vibrato as a musical effect to make their performance more expressive. Herrera and Bonada (1998) present a framework for the extraction and parameterization of vibrato and claim that long sustained notes become boring and uninteresting if their steady states have a strictly constant f_0. This is why performers - and especially well-trained performers - tend to modulate both the frequency and the amplitude of the notes they produce, resulting in *vibrato* and *tremolo*, respectively. In the present study, we consider

[1] A *cent* is a logarithmic frequency unit. A musical semitone is subdivided into 100 cents (Driedger et al., 2016).

vibrato as characterized by its frequency or rate in Hertz and its extent or amplitude in cents. According to (Sundberg, 1987), an aesthetically reasonable range for the vibrato frequency is between 5.5 Hertz and 7.5 Hertz; this is why most methods to extract vibrato from audio signals usually restrict the search to this frequency range, approximately. Another relevant study about vibrato is the one by Prame (1992), who analyzed the performance of ten singers and found that the frequency of the vibrato typically increases at the end of the tones and that the average vibrato frequency across singers was 6.1 Hz.

The aim of our work is to study several expressive characteristics of choir singing, with special emphasis on unison performances, to study how singers blend together and interact with each other in terms of f_0 dispersion, intonation, and vibrato. We formulate three main hypotheses: (1) if we characterize unison performances by instantaneous mean f_0 and f_0 dispersion values, we can identify the perceived f_0 of the performance and quantify the degree of unison, respectively, (2) singers within the same choir section interact with each other in terms of intonation: they adapt their intonation to other singers' intonation, and (3) singers synchronize the frequency of their vibratos for the choir to be perceived as a single entity. We believe that obtaining information about these specific aspects of choir singing would be beneficial for the choir singing community, both for the singers in their practice and for the conductors to better understand their choir and design rehearsals accordingly.

2. Choral-singing Dataset Creation

A novel choral-singing dataset has been created for this study[2]. We collaborated with the Anton Bruckner Choir[3] from Barcelona (Spain) and organized a set of recording sessions in a professional studio with sixteen singers from the choir: four singers per section, i.e. four sopranos (S), four altos (A), four tenors (T) and four basses (B). We recorded the singers in groups, i.e. one session per section, singing in unison, with individual close microphones (with a cardioid polar pattern for directivity purposes) because we were interested in having separate tracks for each individual. In the first session, we recorded a video of the conductor of the choir while she conducted the performance. This video was displayed in the three remaining sessions and used for synchronization purposes. Singers also had the possibility to hear a piano reference (through headphones) for tuning purposes: *a cappella* singing often has the problem of f_0 drift during the performance, which happens when intonation moves away from the reference (Dai and Dixon, 2017). Since this phenomenon was not the focus of this study, we created a piano reference synchronized with the conductor movements.

Three pieces were selected based on the specific needs of the study, which were basically related to the language of the lyrics. These are the pieces we chose:

- *Locus Iste*, written by Anton Bruckner (Latin).
- *Niño Dios*, written by Francisco Guerrero (Spanish).
- *El Rossinyol*, popular Catalan song.

[2] The dataset is published here: http://doi.org/10.5281/zenodo.1286570.

[3] Anton Bruckner Choir: http://cdcantonbruckner.weebly.com.

Overall, this dataset contains, for each of these pieces, the tracks of each individual singer (16 singers), together with the synchronized MIDI files of each choir section. Having the individual tracks also allows the user to create the unison mix for each section, as well as the whole choir performance. The dataset covers the frequency range between 87 Hz and 783 Hz, and is especially dense between 150 and 450 Hz. The notes have durations from 0.15 to 6.21 seconds, with an average of 0.84 seconds.

3. Methodology

Our methodology has three main parts: in the first one we study the characteristics of f_0 dispersion for the different choir sections; in the second one we analyze the correlation between the singers of the unison in terms of intonation, i.e. the degree of synchronization of their f_0 deviations; finally, in the last part of this research, we focus on describing vibrato in unison for singing, which is a very relevant aspect of singing.

Fundamental Frequency (f_0) Dispersion Analysis

The first part of this work aims at studying the f_0 dispersion found in unison choir singing. We obtained the f_0 trajectories for each singer individually using the spectral autocorrelation (SAC) method in (Villavicencio et al., 2015) as an f_0 estimation algorithm. We wanted to describe unison performances using a set of two descriptors, i.e. mean f_0 and f_0 dispersion, both computed from the individual f_0 trajectories of each singer.

The first step was to align audio recordings with their associated MIDI files. These MIDI files were time-synchronized with the audio recordings except for an offset at the beginning, which was manually corrected for all the cases. Then, we used SAC algorithm to obtain the f_0 predictions for each singer. The output of the f_0 estimation algorithm is a set of f_0 values computed frame-wise with a hop size of 5 ms: for each note of the score we have several values.

We divided the f_0 array into notes to study singer behavior per note. We were interested in the average dispersion within the note, following the study by Jers and Ternström (2005). To compute the note boundaries, we extracted the note onsets and offsets from the synchronized MIDI files using the Python library PrettyMidi (Raffel and Ellis, 2014) and segmented the f_0 array.

At this point, we had a set of f_0 values for each note of the score, for each piece and for each singer. We defined the f_0 dispersion as the standard deviation, σ_{f0}, of the distribution of frequencies within a time frame. Besides the standard deviation, we also computed the mean f_0 (μ_{f0}), which we hypothesize it corresponds to the pitch we perceive when we listen to the unison performance. We computed μ_{f0} and σ_{f0} for each note; however, in order to smooth the results, instead of computing the statistical metrics at each analysis frame (one value every 5 ms), we used a sliding window of 8 frames, therefore using 32 values at a time.

To obtain one single value per note for each metric we computed the average along the results for each window. Regarding the σ_{f0}, we used a threshold before averaging the results: since the f_0 annotations were obtained with SAC and not manually, we had some errors in the f_0 trajectories that made the standard deviation reach very high values by mistake. In order to avoid this, we thresholded the values obtained at each window before computing the average. The threshold was set empirically to 60 cents by visualizing several examples and also taking into account the highest dispersion values obtained in other studies.

After the overall analysis of f_0 dispersion, we performed a finer analysis on some of the long notes - longer than 2 seconds, selected after a listening analysis of the pieces - of our dataset in order to study the trend of the f_0 dispersion in these cases. We expected this magnitude to be higher in the attack and release of the notes; however, we did not find any systematic pattern but only in some cases the dispersion is very high because of the attack imprecision, which would be an interesting topic for a study about the synchronization of singers in time.

Intonation Correlation

As mentioned above, singers produce slightly different frequencies even when singing the same note, i.e. f_0 dispersion. Since choir singing is about interacting with the other singers, it is likely that they adjust their intonation to others' intonation. We estimate here to what extent a singer is affected by how the other singers in the same section change their intonation, i.e. increasing or decreasing the produced f_0. We expect these adjustments to be more or less linear: the goal of the intonation synchronization between singers is that the unison dispersion is as low as possible, which means that if a singer increases the produced f_0, the singers around will probably adapt their intonations towards the same direction, either increasing or decreasing their f_0. Assuming a linear correlation, we used the Pearson correlation coefficient, which according to (Papiotis et al., 2014), is the most common method used for quantifying linear dependence between two sets. This coefficient ranges between -1 (complete inverse linear correlation) and 1 (complete linear correlation), being 0 the lack of linear dependence between the two series.

Here, we were interested in studying if the general trend of the intonation was correlated between singers, not in the specific f_0 values. To capture this, we computed the derivatives of each of the f_0 trajectories, removing the continuous part of our data, and then segmented it into notes using the synchronized MIDI files. Instead of computing the Pearson correlation for each note, we also used small windows within each note in order to capture more precise information. When computing the correlation window-wise, we allowed for a small delay in the f_0 adjustments: if a singer modifies the intonation, it will take a while, i.e. a few samples, for the other singers to perceive this change and adapt accordingly. This is why we used a small window (20 frames, about 100 ms) around each sample. In Figure 1 we illustrate this effect with an example of one window: each line corresponds to the derivative of the f_0 trajectory for one singer. We see, for example, that alto 1 decreases around sample 7 and alto 2 imitates this behavior right after. To compute the Pearson correlation coefficient, we used the statistics module of Scipy, a scientific library for Python.

We computed the average of the coefficients of each window to obtain with a single correlation value for each note. We then analyzed the results manually to look for sections or pairs of singers that stand out from the others in terms of correlation. Notice that although coefficients can be either positive or negative, we are interested in their absolute value, since two sets that show an inverse linear dependence might as well be correlated: if one voice decreases the intonation and another voice increases it to match the first one, we will get a negative correlation.

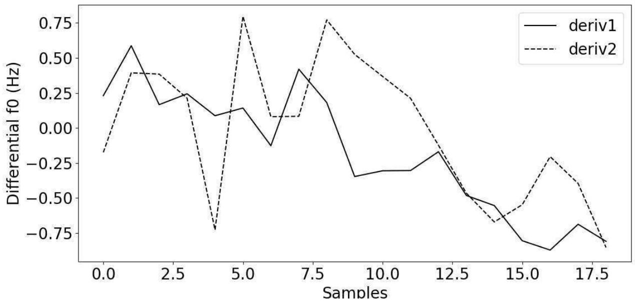


Figure 1. An example of the derivatives of the f_0 trajectories of two voices. This figure illustrates how one singer (deriv2) adjusts her intonation to another singer (deriv1). This window has a Pearson correlation coefficient of 0.57.

Vibrato Analysis

Vibrato is a very important aspect of singing, and most professional and semi-professional singers use it for expressivity purposes (Driedger et al., 2016). Seashore (1931) claims that it is present in the voices of all great artists in about 95% of their phonated time, even in transitions, attacks, and releases.

We used the vibrato extraction algorithm implemented in Essentia (Bogdanov et al., 2013) to find the segments of the performances that contain vibrato and to obtain the corresponding vibrato rates and amplitudes. The vibrato extraction algorithm from Essentia is based on the vocal vibrato detection from MELODIA (Salamon and Gómez, 2012), which is itself based on the method for vibrato characterization described by (Herrera and Bonada, 1998). Following the procedure from the previous stages, we segmented our data (i.e. f_0 trajectories) into notes using the score information, and then carried out a set of experiments described below.

Percentage of vibrato. We first measured the percentage of notes of the dataset that contained vibrato, sorted by singer and piece. Using the vibrato extractor described above, we checked the number of notes reported by the algorithm as having vibrato, and computed the percentage using the total amount of notes. This data is relevant to find out aspects such as whether a singer uses more vibrato than the others, or if singers are more likely to use vibrato in one of the pieces of our dataset.

Vibrato frequency and extent. Using the information extracted in the previous step about the notes that have vibrato, we then studied the vibrato parameters: frequency (or rate) and amplitude (or extent). We iterated through all the notes and kept only the ones with vibrato; notice that although a note is reported to contain vibrato, it does not mean that it is present all along the note. Instead, vibrato usually appears in different parts of the notes. Therefore, to study the frequency and amplitude of the vibrato we first located the segments where it was present, and then extracted both parameters given by the vibrato extractor. For each individual f_0 trajectory, we computed the average and standard deviation of vibrato frequencies and amplitudes.

Vibrato correlation. In previous steps, we studied the correlation between singers in terms of intonation. Following a very similar procedure, we studied the vibrato synchronization between singers of the same choir section (unison). We chose to use the vibrato frequency in this case; however, the vibrato extent could also be used to check whether singers adapt the parameters of the vibrato to the other singers. We computed the correlation between the vibrato produced by each pair of singers of the same section. We used the Pearson correlation coefficient as a metric to quantify, for each note, if the two singers produce similar vibrato rates, which is the same as in the study by Daffern (2017): she computes the correlation for specific relevant notes, i.e. the last note of the performance; however, we were more interested in the overall correlation values, so we averaged our results for each performance.

4. Results

In this section we present the results we obtained, organized following the same structure as the methodology: f_0 dispersion analysis, intonation correlation, and vibrato analysis.

F_0 Dispersion Analysis

Following the methodology described above, we extracted the f_0 trajectories and then computed the dispersion note-wise. The dispersion phenomenon is illustrated in Figure 2, where we display the f_0 trajectories corresponding to a single note of the soprano section. We observe that different and varying frequency values are produced by each singer, resulting in a set of similar yet different values of f_0 at each time instant. In Figure 3 we show the averaged results of the f_0 dispersion by choir section and by piece.

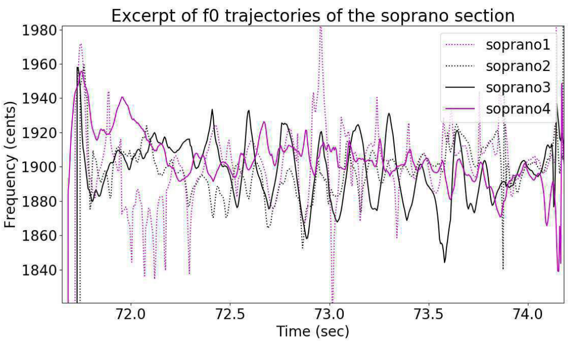


Figure 2. Excerpt of a unison performance by 4 sopranos singing a E5 (around 660 Hz) from *Locus Iste* (minute 1:12). Each line corresponds to the f_0 trajectory of an individual.

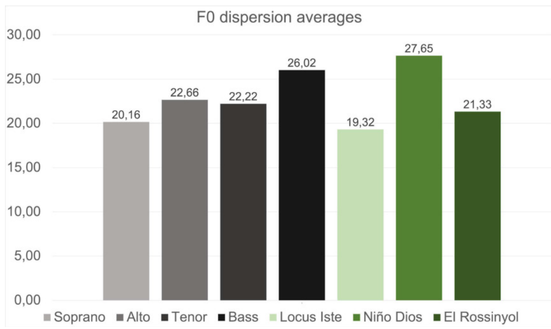


Figure 3. f_0 dispersion results averaged by section (grayscale) and by piece (greenscale).

By analyzing the individual results, we found the dispersion to vary between 16 and 30 cents, depending on the voice, which agrees with the results in (Ternström, 2002). In the Figure 3 we observe that although there are not very strong differences in dispersion between sections nor pieces, basses tend to show a higher dispersion, while sopranos obtained the lower values. Another relevant result is that the second piece was reported by the singers to be the most difficult to sing and has a higher average dispersion. This piece has a higher level of complexity in terms of rhythm, tempo, and intervals, so this result suggests that a more difficult piece might lead to higher dispersion values, therefore decreasing the degree of unison.

Intonation Correlation

Following the procedure detailed in the methodology, we obtained one correlation value for each analysis window, thus getting several correlation values for each note of the score. By manually inspecting some of the results, we realized we got significantly different correlation values between windows of the same note. This is illustrated in Figure 4, which includes a plot of the correlation results of all the analysis windows within a single note, i.e. one value for each window. We observe that the values oscillate, which suggests that the intonation adjustments are not systematic nor constant, and do not follow any specific pattern. However, we can see that there are highly correlated windows (very close to either 1 or -1) in these results, which also suggests that at some points, singers adjust their intonations.

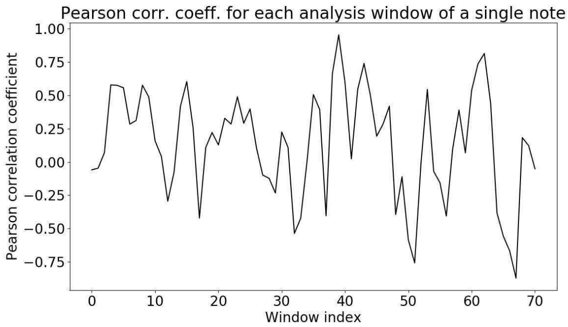


Figure 4. Evolution of Pearson correlation coefficient along one note.

Given the varying nature of the correlation in this context, it is difficult to generalize its trend. Our first approach was computing the average for each note and then manually study which pairs of singers are most correlated and which of the pieces had the highest correlation values. We did not average our results by singer or piece because some relevant information, such as maximum values, was lost. In Table 1 we summarize the results: for each piece and choir section, we report the maximum average correlation (max average) and also the maximum among the maximum correlation values (highest maxim.). For each of these two metrics we also show to which pair of singers they correspond (columns "pair of singers") in order to check for repeated patterns, i.e. a pair of singers that has the highest correlation for all pieces.

Table 1. Summary of the results of the intonation correlation analysis.

Piece	Choir section	Max average	Pair of singers	Highest maxim.	Pair of singers
Locus Iste	Soprano	0.31	3-4	0.6	2-3
	Alto	0.33	3-4	0.7	1-4
	Tenor	0.36	1-2	0.58	1-2
	Bass	0.34	1-3	0.52	3-4
Niño Dios	Soprano	0.34	1-3	0.88	2-3
	Alto	0.33	3-4	0.78	1-4
	Tenor	0.33	2-3	0.82	1-2
	Bass	0.32	1-3	0.78	3-4
El Rossinyol	Soprano	0.33	1-4	0.66	2-3
	Alto	0.32	3-4	0.57	3-4
	Tenor	0.33	2-3	0.73	2-3
	Bass	0.32	2-3	0.58	2-3

On one hand, we observe in Table 1 that the average correlation is similar and around 0.32-0.33 in all cases. This particular result first suggests that there is not a specific pair of singers that is more synchronized than the rest. Additionally, it also shows that, on average, the correlation between singers in terms of intonation is quite low, even lower that 0.5. On the other hand, if we analyze the maximum correlation values (5th column in Table 1) we observe that if we look at the first two pieces (*Locus Iste* and *Niño Dios*), the pairs of singers with the highest maximum correlation are the same: sopranos 2 and 3, altos 1 and 4, tenors 1 and 2, and basses 3 and 4. In the last piece (*El Rossinyol*), however, this changes except for the sopranos. Although this might be a coincidence and this highest value is not representative of the whole data, it is likely that a deeper analysis of these singers would reveal more relevant patterns and more interesting information about the intonation interaction between singers.

Vibrato Analysis: Percentage of vibrato

Using the vibrato extractor described in previous sections we computed the percentage of notes that had vibrato for each piece and choir section. Notice that we computed the percentage for each singer, and then averaged the results for each section. The results are displayed in Figure 5 in the form of a bar chart, where we observe that in general, singers produce vibrato in less than 50% of the notes. However, these results need to be analyzed taking into account the context: these singers are part of a semi-professional choir, and although most of them have some kind of singing training, in general they are not professional singers, which probably means that they do not have a complete control over their vibrato. In Figure 5, we also observe that the first piece (*Locus Iste*) has a significantly higher percentage of notes with vibrato, which might be a result of the piece the slowest tempo.

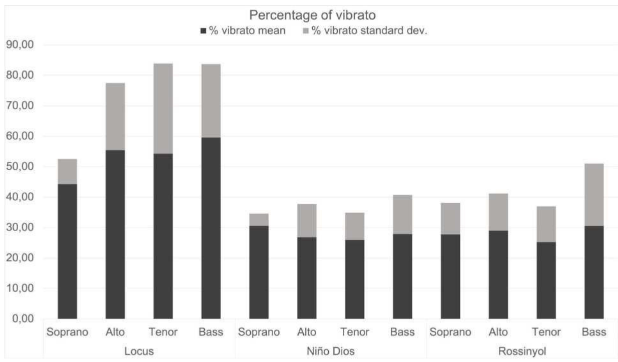


Figure 5. Percentage of notes with vibrato separated by piece and choir section.

Vibrato frequency and extent. In Figures 6 and 7 we display the results of our analysis of the frequency and amplitude of the vibrato, averaged and separated by section and piece. We observe that all vibrato frequencies are around 5 Hz, with quite low standard deviations, which is consistent with the results presented by Sundberg (1987) and Prame (1992). Regarding vibrato amplitude (Figure 7), we do not find any repeated pattern in the results, although we observe that for the second piece, in general, the amplitude of the vibrato is larger, especially for the basses. This is a surprising result, since this piece was reported to be the most difficult to sing by the singers, as well as the one with a fastest tempo, therefore we would expect the vibrato to be subtler or even missing.

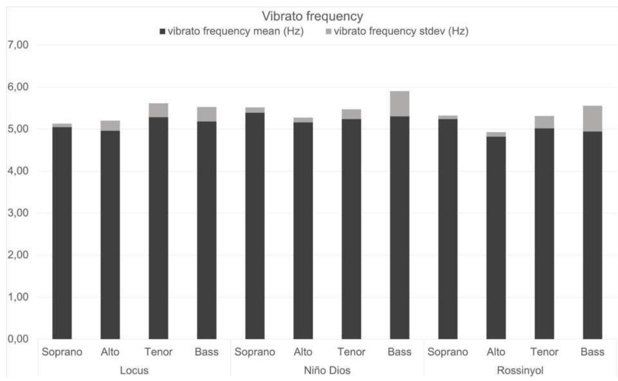


Figure 6. Vibrato rate/frequency averaged by choir section and piece. The values are displayed in Hz.

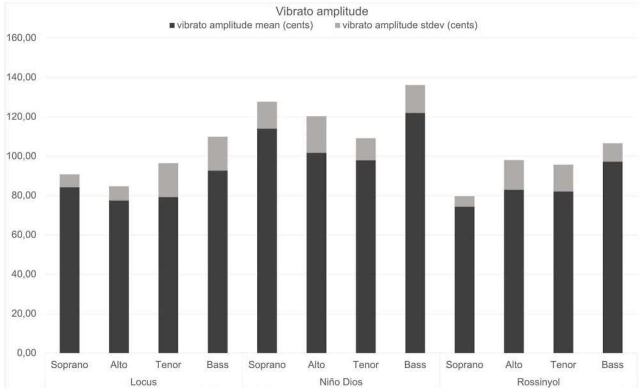


Figure 7. Vibrato amplitude/extent averaged by choir section and piece. The values are displayed in cents.

Vibrato correlation. For this part we present the pairs of singers that have an average vibrato correlation higher than 0.3. This threshold was chosen by sorting all the values (high to low) and keeping only those which differed less than 0.1 from their previous ones. The most relevant result is that the piece *Niño Dios* has the highest vibrato correlation: 0.56 for basses 4 and 2, and several values ranging from 0.36 to 0.47 for soprano pairs. However, these results are not very consistent nor explanatory, so we would need more data and more singers to corroborate and broaden them.

Conclusion

The main aim of this work was to study the synchronization and interaction between singers of a choir in terms of their f_0 dispersion, intonation, and vibrato. We characterized the unison performances by computing the mean f_0 and the f_0 dispersion and found the latter to lie in the range between 16 and 30 cents, being especially large in the bass section of the choir. We used the Pearson correlation coefficient to estimate how singers' intonation is affected by other singers when performing together, and our results show that the intonation correlation fluctuates a lot. Although we observed that a few pairs of singers had high correlation values in a few examples, the average was around 0.32 in all cases. With more data, we would be able to extract more relevant information. We have already planned a new recording session with a choir to extend our working datasets. We finally used a vibrato extractor to compute its presence in our dataset and compare it between different singers. The synchronization of the vibrato rate is a characteristic we would expect from more professional singers that have a much higher control of their voices.

Acknowledgements. This work is partially supported by the Spanish Ministry of Economy and Competitiveness under the CASAS project (TIN2015-70816-R) and the European Commission under the TROMPA project (H2020 770376). Many thanks to all the singers of the Anton Bruckner Choir who took part in this project.

References

Bogdanov, D., Wack, N., Gómez Gutiérrez, E., Gulati, S., Herrera Boyer, P., Mayor, O., ... & Serra, X. (2013). Essentia: An audio analysis library for music information retrieval. In *Proceedings of the 14th ISMIR Conference*.

Daffern, H. (2017). Blend in singing ensemble performance: Vibrato production in a vocal quartet. *Journal of Voice, 31*(3), 385-e23.

Dai, J., & Dixon, S. (2017). Analysis of Interactive Intonation in Unaccompanied SATB Ensembles. In *Proceedings of the 18th ISMIR Conference*.

Devaney, J. (2011). An empirical study of the influence on musical context on intonation practices in solo singers and SATB ensembles. *Doctoral dissertation*. McGill University. Montreal, Canada.

Driedger, J., Balke, S., Ewert, S., & Müller, M. (2016). Template-based vibrato analysis in music signals. In *Proceedings of the 16th ISMIR Conference,* p. 239-245.

Hagerman, B., & Sundberg, J. (1980). Fundamental frequency adjustment in barbershop singing. In *TMHQPSR Speech, Music and Hearing: Quarterly Progress and Status Report, 21*(1), 28-42.

Herrera, P., & Bonada, J. (1998). Vibrato extraction and parameterization in the spectral modeling synthesis framework. In *Proceedings of the Digital Audio Effects Workshop (DAFX98)*.

Jers, H., & Ternström, S. (2005). Intonation analysis of a multi-channel choir recording. In *TMHQPSR Speech, Music and Hearing: Quarterly Progress and Status Report, 47*(1), 1-6. ISO 690

Lottekmoser, W., & Meyer, F. (1960). Frequenzmessungen an gesungenen Akkorden. *Acta Acustica united with Acustica, 10*(3), 181-184.

Papiotis, P., Marchini, M., Perez-Carrillo, A., & Maestre, E. (2014). Measuring ensemble interdependence in a string quartet through analysis of multidimensional performance data. In *Frontiers in psychology,* 5, 963.

Prame, E. (1994). Measurements of the vibrato rate of ten singers. *The journal of the Acoustical Society of America, 96*(4), 1979-1984.

Raffel, C., & Ellis, Daniel P.W. (201) Intuitive analysis, creation and manipulation of MIDI data with pretty_midi. In *Proceedings of the 15th ISMIR Conference*. Late breaking demo paper.

Salamon, J., & Gómez, E. (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. In *IEEE Transactions on Audio, Speech, and Language Processing, 20*(6), 1759-1770.

Seashore, C. E. (1931). The natural history of the vibrato. In *Proceedings of the National Academy of Sciences, 17*(12), 623-626.

Sundberg, J. (1987). *The Science of the Singing Voice*. Northern Illinois University Press.

Ternström, S. (1993). Perceptual evaluations of voice scatter in unison choir sounds. *Journal of Voice*, *7*(2), 129-135.

Ternström, S. (2003). Choir acoustics: an overview of scientific research published to date. *International Journal of Research in Choral Singing, 1*(1), 3-12.

Villavicencio, F., Bonada, J., Yamagish, J., & Pucher, M. (2015). Efficient Pitch Estimation on Natural Opera-Singing by a Spectral Correlation based Strategy. In *IPSJ Technical Report*.

Zadig, S., Folkestad, G., & Lyberg Åhlander, V. (2016). Multi-track recordings of choral singers: Development and validation of a method to identify activities and interaction in the choral voice, based on recordings of the individual singers. *Bulletin of empirical music education research, 7*(1), 1-20.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Importance of Felt Mood and Emotion for Expressive Movement Characteristics in Pianists

Anna Czepiel[1] and Geoff Luck[2]

[1, 2] *Department of Music, Art and Culture Studies, University of Jyväskylä, Finland*

[1]anmaczep@student.jyu.fi, [2]geoff.luck@jyu.fi

Abstract

Expressive intentions and perceived emotion can be shown through gestures in a musical performance, but emotional engagement and felt emotions can also influence performers' movement. In exploring this further, we studied movement features in pianists performing a piece in three conditions with low, medium and high emotional engagement (EE), where pianists' movement was recorded by an optical motion capture system. After each performance in each condition, positive affect and negative affect scores were obtained. We hypothesised that 1) Felt emotions and EE with the music would influence expressive movement (operationalised as amount of movement in typically expressive locations, such as head and shoulders, and number of postural fluctuations), and 2) positive and negative felt affects (moods) would have different movement characteristics. In support of Hypothesis 1, we found that performers elicited significantly more expressive movement in medium and high EE compared with low EE. Jerkiness was lowest in the high EE condition in locations associated with technical movement (wrists and fingers). In support of Hypothesis 2, positive-affect scores positively correlated with movement features associated with expressive movement and negative-affect scores negatively correlated with amount of expressive movement and positively correlated with jerkiness in the left wrist. Stepwise regressions further revealed that movement associated with negative affect depends on whether it was evoked by the performance (e.g. anxiety) or by the music (e.g. feeling the sadness of the music). Performance-related negative affect was associated with jerkiness in wrists, whereas music-related negative affect was associated with stillness in posture and more hunching of the shoulders. These results may benefit musicians in professional and educational domains, especially if struggling with performance anxiety, by suggesting they engage more in the emotional aspects of the music (as well as enjoy it) to provide a more visually expressive and smooth performance.

Introduction

A music performance can contain a wealth of visual information and a variety of different body movements, where expressivity is associated with amount of movement (Davidson, 1991; Wanderley, Vines, Middleton, McKay, & Hatch, 2005) and communication of emotion is associated with quality of movement (Dahl & Friberg, 2004; 2007). Subsequently, body movement in a musical performance has a great impact on evaluation of the performance quality (Griffiths & Reay, 2018; Huang & Krumhansl, 2011; Juchniewicz, 2008; Tsay, 2013).

Studies show that expressivity is linked to the amount of movement in musical performance. More specifically, increase in expressive intention can increase the amount of movement in locations such as the torso and head (Castellano, Mortillaro, Camurri, Volpe, & Scherer, 2008; Thompson & Luck, 2012) and the left hand (Davidson, 2007). Indeed, expressive intentions can be perceived more strongly in the visual domain compared to the audio-visual domain (Vuoskoski, Thompson, Clarke, & Spence, 2014), although this is dependent on the genre of the music (whether Baroque, Romantic or modern, Huang & Krumhansl, 2011) and on the viewers' musical training (Juchniewicz, 2008; Kwan, 2016).

The quality of movement features can communicate emotion in a music performance (Dahl & Friberg 2004, 2007; Krahé, Hahn, & Whitney, 2015). In particular, large movements tend to represent happiness, small movements sadness; large and jerky movements can represent anger and small, fast and jerky movements can represent fear (Dahl & Friberg, 2004). Emotional intensity can be exhibited by fluctuating forward and backward lean towards and away from the piano keyboard (Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004).

These studies demonstrate how emotion and expression are conveyed by performers. However, more recent studies show that movement can also convey music-related felt emotions (Glowinski et al., 2008; Van Zijl & Luck, 2013). Indeed, performers are usually encouraged to really "feel" the emotion of music (Reimer, 2004). The ability to feel and engage with emotion when playing music has been of importance for centuries: in relation to music performance practice, C.P.E. Bach wrote in a keyboard treatise: "a musician cannot move others unless he too is moved… in sad passages, the performer must languish and become sad" (cited in Persson, 1993). Additionally, emotional engagement with the music seems vital in constructing an expressive performance (Van Zijl & Sloboda, 2010). Van Zijl & Luck (2013) found movement differences when performers *expressed* sadness (upright posture, more movement, greater speed, velocity and jerkiness of movement) compared to when they really *felt* sadness (bent posture, significantly smaller movement, and slower speed, acceleration and jerkiness of movement). However, in addition to music-related emotions, felt emotions in a music performance may also consist of performance-related emotion (Van Zijl & Sloboda, 2013). These performance-related felt emotions can also be embodied in movement, for example shaking when feeling nervous (Van Zijl & Sloboda, 2013). Additionally, movement exhibited due to these performance-related felt emotions (e.g. performance anxiety) can subsequently affect audience ratings of expressiveness (Kwan, 2016).

The main aim of the present study was to further explore the influence of felt emotion on movement features in piano performances. Several studies exploring gestures and expressivity in music performance have used complete musical pieces (Davidson, 2007; Thompson, 2007; Wanderley et al., 2005) and a previous study into felt emotion and movement features used short excerpts (Van Zijl & Luck, 2013). A motivation for the present study, was to further Van Zijl & Luck's (2013) findings to the same level as expressivity studies by using complete music compositions. We additionally decided to use music pieces that were already in a pianist's repertoire, in order to evoke genuine emotion. One possible problem when studying emotion in performance settings, which has been highlighted in some theatre-performance studies, is that emotion is "acted" rather than really felt (Wallbott, 1998). Therefore, we asked the performer to play a piece that they felt emotionally attached to (this would aid the Emotional Recollection task, see Methods) as we assumed that a performer would be able to engage in a piece more successfully if it were part of their (favourite) repertoire, rather than sight-reading or learning a piece prior to the study.

Based on previous research showing that expressive intentions increase expressive movement and that felt emotions change amount and quality of movement, we posed the following hypotheses for the current research: 1) Felt emotion and emotional engagement (EE) with the music would influence expressive movement; and 2) felt mood would also influence movement features, where positive affect would be associated with more expressive movements whereas negative affect would be associated with less expressive and more subtle movement. "Expressive movement" would be operationalised by the amount of movement in locations such as the head and torso, (Castellano et al., 2008; Davidson, 2007; Thompson & Luck, 2012) and posture fluctuations (Camurri et al., 2004).

Method

Participants

Ten pianists (7 females, mean age = 33 years, standard deviation (SD) = 11.39), 5 of whom were professionals, were recruited from Jyväskylä, Finland.

Stimuli

Participants were asked to play a piece with which they had an emotional attachment. Pieces chosen were Beethoven's Adagio cantabile from *Sonata Pathetique* Op.13, Debussy's Arabesque No. 1, *La fille aux cheveux de lin,* and *L'isle joyeuse*, Hannikainen's Valse No. 1 Op.17, Kuusisto's *Berceuse* from Trois Miniatures Op. 4, Rachmaninoff's Etudes-Tableaux Op. 39 No. 5 in E-flat minor, Wagner/Liszt's *Isoldens Liebestod* S. 447, Sibelius's *Romance* from 10 pieces Op. 24, and a waltz written by a participant's uncle.

Apparatus

Movement (three-dimensional position of reflective markers) was recorded using an optical motion capture system (Qualysis Pro Reflex) with eight cameras at 120 frames per second. Each participant wore 22 reflective markers: four around the head, two on the neck (front and back), two on the mid-torso (front and back), four around the hip, one on each shoulder, one on each elbow, two on each wrist and one ring marker on each middle finger.

Measures

The Positive Affect Negative Affect Schedule (PANAS, Watson, Clark, & Tellegen, 1988) was used to measure the felt affective state (felt mood) of the performers.

Procedure

Participants first completed the PANAS, then performed their chosen pieces in three conditions that were selected to elicit different levels of EE (based on the conditions used in Van Zijl & Luck, 2013):

- **Technical (T):** low EE (instruction to play the piece as accurately as possible);
- **Expressive (Ex):** medium EE (instruction to "communicate" the music);
- **Emotional (Em):** high EE (Emotion Recollection Task, then instruction to perform as if they were just playing the emotion of the music).

After performing in each condition, participants were instructed to complete the PANAS to convey how they felt *during* their performance, then they were asked some reflective questions. Once data collection was complete, participants were offered cakes to counter any negative effects experienced in the experiment, as food has been shown to induce positive emotions (Isen, Daubman, & Gorgoglione, 1987).

Emotion Recollection Task

As it has been shown that performers would often recall biographical memories and use imagination to create an expressive musical performance (Persson, 2001; Van Zijl & Sloboda, 2010), the pianists were asked to describe the emotion(s) of their piece, then asked to recall a personal memory and/or imagine a situation associated with that emotion. They were instructed to take at least a minute to absorb themselves in this emotion as fully as possible.

Analysis

PANAS scores underwent baseline correction. Pieces were categorised generally into either positive or negative valence based on the participants' comments (Positive: the three Debussy pieces, participant's uncle's piece, Sibelius; Negative: Kuusisto, Beethoven, Hannikainen, Rachmaninoff and Wagner/Liszt).

Motion Capture data was pre-processed by trimming the movement data to fit the length of the performance, gap-filling any missing trajectories (using polynomial or linear interpolation depending on which represented more naturalistic movement) and transforming the original markers to form a set of 12 (secondary) markers for the head, neck, left and right shoulders, left and right elbows, mid-torso, hip, left and right wrists, and left and right (middle) fingers. The following features were extracted using MATLAB Motion Capture (MoCap) Toolbox (Toiviainen & Burger, 2015):

Amount of movement (AM) represented by total cumulative distance;

Jerkiness of movement (J) represented by norm of third time derivative (calculated using numerical differentiation and a Butterworth smoothing filter, second-order zero-phase);

Postural features (PF):

- **Posture (P)** represented by angle along the y dimension between hip marker and neck marker;
- **Postural lean (PL)** represented by distance of head from piano;
- **Shoulder hunch** represented by distance between head and mean location of shoulders;
- **Head tilt left (HTL)** represented by distance between head and left shoulder;
- **Head tilt right (HTR)** represented by distance between head and right shoulder.

The mean jerkiness for each performance was calculated. Additionally, both means and standard deviations (representing fluctuation) were calculated for each postural feature. These 34 movement features (AM for each marker = 12, jerkiness for each marker = 12, means of different postural features = 5, standard deviation (SD) of different postural features = 5) were extracted for each performance (10 pieces × 3 conditions). To obtain representative data that would allow comparisons between participants, the movement features were also normalised using feature scaling (Min-Max normalisation), a technique that has been used in other kinematic studies to account for individual differences (Best & Begg, 2006).

To gauge if emotional engagement differed between conditions, a mixed design analysis of variance (ANOVA) with the repeated measures factor Condition (Technical / Expressive / Emotional) and between-subjects factor Piece Valence (Positive / Negative) was conducted for the PANAS scores. To test the hypotheses, mixed design ANOVAs with repeated measures factor Condition (Technical / Expressive / Emotional) and the between-subjects factor Piece Valence (Positive / Negative) were also conducted for movement features. Spearman correlations and stepwise regressions were also run to examine associations between the movement features and positive or negative felt mood (PANAS scores) of the performer.

Results

A significant main effect of Condition on PANAS scores confirmed that participants' felt emotion (and therefore emotional engagement) was affected by the conditions (positive affect scores: $F(2,16) = 8.41$, $p < .01$; negative affect scores: $F(2,16) = 8.26$, $p < .01$), even when considering Piece Valence (significant interaction of Condition × Piece Valence for positive affect scores, $F(2,16) = 4.047$, $p < .05$).

With regard to the first hypothesis, ANOVA results for main effect of Condition (with Greenhouse-Geisser corrections for sphericity violations), estimated marginal means and pairwise comparisons (with Bonferroni corrections) for movement features are reported in Table 1. We found a significant main effect of Condition on AM values in the head, neck, shoulders, right elbow and hip. Estimated marginal means showed that AM was highest in Expressive conditions in these locations. Pairwise comparisons showed that AM was significantly different between Technical and Expressive conditions, and between Technical and Emotional conditions.

A significant main effect of Condition on jerkiness of movement was found in locations associated with technical movement (in both wrists and both fingers). Estimated marginal means showed that normalised jerkiness values were lowest in the Emotional condition in these locations. Pairwise comparisons showed that jerkiness was significantly different between Technical and Emotional conditions, and between Expressive and Emotional conditions.

There was a significant main effect of Condition on fluctuations (SD) on the postural features. Estimated marginal means showed that fluctuations of back posture, postural lean and shoulder hunch were highest in the Expressive condition, whereas fluctuations of head tilt left and head tilt right were highest in the Emotional condition. Pairwise comparisons showed that posture was significantly different between Technical and Expressive conditions, and between Technical and Emotional conditions.

Table 1. ANOVA results (for main effect of Condition) for amount of movement (AM), jerkiness (J) and fluctuation (standard deviation, SD) of postural features (PF) with estimated marginal means of normalised movement features and pairwise comparisons between conditions.

Movement feature		ANOVA results (main effect of Condition)	Estimated marginal means of normalized features			Pairwise Comparisons		
			Technical	Expressive	Emotional	T-Ex	T-Em	Ex-Em
AM	Head	$F(1.22, 10.82) = 10.82$**	.40 (*SE* = .07)	.52 (*SE* = .07)	.49 (*SE* =.07)	**		
	Neck	$F(2, 16) = 18.83$***	.21 (*SE* = .03)	.30 (*SE* = .04)	.28 (*SE* =.04)	**	*	
	Mid-torso	$F(2, 16) = 13.07$***	.09 (*SE* = .01)	.14 (*SE* = .02)	.13 (*SE* =.02)	**	*	
	Left shoulder	$F(2, 16) = 18.06$***	.28 (*SE* = .04)	.37 (*SE* = .04)	.34 (*SE* =.05)	***	*	
	Right shoulder	$F(2, 16) = 18.22$***	.25 (*SE* = .03)	.34 (*SE* = .04)	.32 (*SE* =.04)	**	*	
	Right elbow	$F(2, 16) = 5.70$*	.45 (*SE* = .04)	.50 (*SE* = .04)	.47 (*SE* =.04)			
	Hip	$F(2, 16) = 6.21$**	.00 (*SE* = .00)	.03 (*SE* = .01)	.02 (*SE* =.01)	*		
J	Left wrist	$F(2, 16) = 7.88$**	.60 (*SE* = .02)	.59 (*SE* = .03)	.54 (*SE* =.03)		*	*
	Right wrist	$F(2, 16) = 4.24$*	.54 (*SE* = .02)	.53 (*SE* = .03)	.49 (*SE* =.03)			*
	Left finger	$F(2, 16) = 11.51$**	.99 (*SE* = .01)	.95 (*SE* = .02)	.88 (*SE* =.03)		*	*
	Right finger	$F(2, 16) = 4.33$*	.79 (*SE* = .05)	.77 (*SE* = .04)	.72 (*SE* =.04)			*
PF	SD of Posture	$F(2, 16) = 77.70$***	.00 (*SE* = .00)	.92 (*SE* = .05)	.81 (*SE* =.07)	**	**	
	SD of PL	$F(1.20, 9.60) = 24.84$***	.00 (*SE* = .00)	.80 (*SE* = .09)	.78 (*SE* =.11)	***	***	
	SD of SH	$F(1.12, 9.50) = 18.01$***	.00 (*SE* = .00)	.84 (*SE* = .11)	.66 (*SE* =.11)	***	***	
	SD of HTL	$F(2,16) = 9.06$ **	.10 (*SE* = .10)	.57 (*SE* = .12)	.84 (*SE* =.10)	*	*	
	SD of HTR	F(1.22, 9.73) = 33.16***	.00 (*SE* = .00)	.70 (*SE* = .10)	.91 (*SE* =.08)	***	***	

***p* < .05, ***p* < .01. ****p* < .001**

With regard to the second hypothesis, no significant interactions for Condition × Piece Valence were found. However, an interaction was nearing significance for the hip, $F(2,16) = 3.45$, $p = .057$. The largest difference appeared to be in the Emotional condition: high EE with positive emotions resulted in higher AM in the hip compared to high EE with negative emotions (see Figure 1).

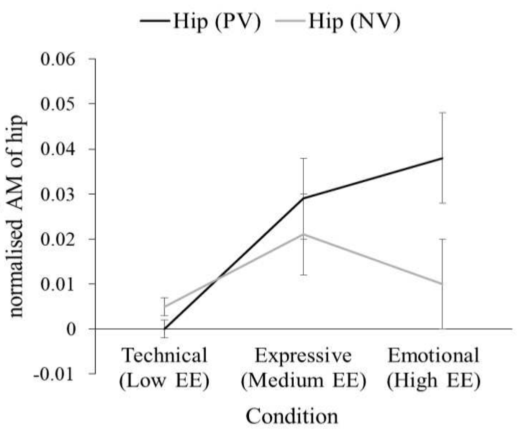


Figure 1. Differences AM in hip across conditions with music of positive valence (PV) and with music of negative valence (NV).

Spearman correlations and stepwise regressions were run to further gauge the influence of positive and negative felt emotions on performer movement. We found that positive affect scores positively correlated with movement features associated with expressive movement: AM in head ($r_s = .450$, $p < .01$), right shoulder ($r_s = .458$, $p < .01$), left shoulder ($r_s = .490$, $p < .01$), right elbow ($r_s = .376$, $p < .05$) and hip ($r_s = .599$, $p < .001$), and also fluctuations in posture ($r_s = .385$, $p < .05$), postural lean ($r_s = .42$, $p < .05$), shoulder hunch ($r_s = .44$, $p < .05$) and head tilt right ($r_s = .43$ $p < .05$).

Negative affect scores negatively correlated with expressive movement: AM in head ($r_s = -.40$, $p < .05$) and right shoulder ($r_s = -.37$, $p < .05$), also fluctuations in postural lean ($r_s = -.41$, .05), head tilt left ($r_s = -.39$, $p < .05$) and head tilt right ($r_s = -.40$, $p < .05$) and also positively correlated with jerkiness of left wrist ($r_s = .37$, $p < .05$). Stepwise regressions were run to see whether movement features could predict positive or negative affect. A significant regression equation was found for positive affect ($F(2,24) = 11.54$ $p < .001$) with an R^2 of .49, where AM of hip ($\beta = .63$, $p < .001$) and jerkiness of neck ($\beta = -.49$, $p < .001$) were significant predictors of positive affect. A significant regression equation was also found for negative affect ($F(3,23) = 11.33$, $p < .001$), with an R^2 of .60, where jerkiness of hip ($\beta = .49$, $p < .01$) and jerkiness of left wrist ($\beta = .48$, $p < .01$) were significant predictors of negative affect, with fluctuation of postural lean ($\beta = -.28$, $p = .06$) almost as a significant predictor. We also ran stepwise regression within conditions. We found that in the Technical condition, negative affect was significantly predicted by jerkiness of wrists ($F(2,7) = 18.37$, $p < .01$), with an R^2 of .84, where more jerkiness of left wrist predicted higher negative affect ($\beta = 6.02$, $p < .001$) and less jerkiness of right wrist predicted higher negative affect ($\beta = -3.805$, $p < .001$). However, in the Emotional condition negative affect was significantly predicted by different movement features ($F(2,7) = 11.11$, $p < .01$), with an R^2 of .76, namely fluctuation of postural lean ($\beta = -1.14$, $p < .01$) and fluctuation of shoulder hunch ($\beta = .73$, $p < .05$).

Discussion

The aim of this study was to further explore how movement features in a piano performance are influenced by 1) felt emotion and emotional engagement (EE) and 2) differences in positive and negative affect. To this end, ten pianists performed a piece that they felt emotionally attached to. Movement features were compared in conditions with differing EE. Correlations and stepwise regressions were run between movement features and positive/negative affect scores obtained in these conditions.

In support of our first hypothesis, we found that with increased EE performers made more movement in typically expressive locations (body areas not directly related to making the sound, but associated with ancillary gestures of head, torso and arms, Wanderley et al., 2005). This supports previous studies that also found that expressivity in performance is conveyed by amount of movement in the torso and head of pianists (Davidson, 1991; Thompson & Luck, 2012).

In addition, we found that with increased EE there was reduced amount of jerkiness in locations associated with technical movements for playing the piano (wrists and fingers). As shakiness (which could also be measured by jerkiness) is linked to nervousness (Van Zijl & Sloboda, 2013) or associated with less instrumental expertise (Nusseck & Wanderley, 2009), our results suggest that higher EE may result in more fluid playing, ensuing a less stressful (or at least a less visibly stressful) performance.

However, this reduction in jerkiness in the Emotional condition may be due to the order of conditions pianists performed in. Although the pianists may have become more relaxed as time went on, we still believe the decrease in jerkiness was due to high EE for two reasons. First, significant differences were found between Expressive (second) and the Emotional (third) condition, whereas if participants were getting used to the experiment, one might expect a more significant decrease from the first to the second condition (Technical to Expressive), rather than from the second to the third (Expressive to Emotional). Second, participants described in the interviews how they felt “freer” and as if “mistakes did not matter” in the Emotional condition. It should also be noted that from observing and listening to the recordings, the number of mistakes did not necessarily decrease across conditions, but the participants felt that mistakes were less important and this was reflected in the reduction of jerkiness. Our results suggest then that engaging with emotion could lead to smoother movements despite mistakes occurring in a performance. Considering the research by Waddell & Williamon (2017), who found an audience’s overall judgement score was lower when they heard a mistake followed by a visual movement reflecting the mistake (negative facial reaction) compared to the same performance mistake with no visual cue reflecting the mistake, we further propose that higher EE could also provide a more convincing performance, even with mistakes.

We also found differences in EE in terms of posture and head tilt. Our results suggest that when playing in the Technical condition, performers had a more stable torso and head posture, whereas in the Expressive and Emotional conditions, their torso posture, head posture, shoulder hunch and head tilting movement fluctuated significantly more. Head tilt fluctuated the most in the Emotional condition, supporting the findings that head tilt is used as a device for expressing emotions in acting (Dael, Mortillaro, & Scherer, 2012) as well as in music performance (Davidson, 1991; 2012; Delalande, 1995). Moreover, in previous studies exploring human-robot interaction, the lateral head tilt of robots increased perception of "naturalness" (Liu, Ishi, Ishiguro, & Hagita, 2012). We tentatively suggest, therefore, that head tilt may reflect a more genuine emotion, rather than just "expressing" it.

In testing our second hypothesis, we found movement differences within positive and negative felt emotions. Engaging with the emotion of the music influenced the amount of movement in the lower torso (although the observed differences fell just short of statistical significance). When pianists strongly engaged with negative emotions (i.e. feeling sadder), there was reduced lower torso movement. When pianists strongly engaged with positive emotions (i.e. feeling more joyful), lower torso movement became more animated. This supports the findings of Dahl and Friberg (2004), who suggested happy emotion leads to larger movement and sadness to lesser movement. It also extends their results as we found that the movement differences occurred in the torso, suggesting that felt emotions might be exhibited more subtly, such as in inner body and torso locations, rather than in extremities (such as head and shoulders). It is in line with findings of Van Zijl & Luck (2013), who found that when actually feeling negative (sad) emotions, posture is more arched compared to more upright when "expressing" sadness, extending their findings to piano performances. Additionally, it supports research by Saarikallio, Luck, Burger, Thompson, & Toiviainen (2013), who found that dance movements (to an emotionally neutral piece of music) reflected felt affect, where positive affect correlated with larger amount of movement.

Correlation and regression analyses revealed that positive affect was associated with more expressive movement (AM in head, shoulders, right elbow, hip and fluctuation of postural features). Negative affect, on the other hand, exhibited less expressive movement (e.g. AM in hip), but more jerky movement in wrists and a more stable posture.

Further investigation within conditions suggested that movement is exhibited differently depending on whether the negative felt affect is music-related or performance-related. In the Technical condition, the instruction was to play the piece as accurately as possible, therefore felt emotions would be most likely performance-related. On the other hand, instruction for the Emotional condition was to become as absorbed in the music as possible, thus it is assumed the most music-related emotions would have been felt here. This highlights the difference between music-related and performance-related emotions in negative felt emotions: *performance*-related negative felt affect was associated with jerkiness of wrists, whereas *music*-related negative felt affect was associated with a more stable posture (less leaning towards and away from the piano) and more fluctuations in shoulder hunching. In terms of negative affect and jerkiness in wrists, an increase of jerkiness was associated with more negative emotion only in the left wrist, with the opposite effect in the right wrist. We suggest this may have been due to the fact that (generally speaking and from observing the participants' music scores) the melody was usually in the right hand and therefore given more attention. As the left hand tends to be the less dominant hand, and has less dominant lines, it may have received less attention in practice and therefore perhaps may not have been as stable in a performance compared to the right hand.

We feel that these interpretations of performance-related and music-related emotions which account for movement differences remain relatively speculative, as the PANAS does not specifically account for either performance-related or music-related emotions. This highlights the need for an instrument to measure felt emotions in a music performance as well as one that specifically distinguishes between music-related and performance-related emotions, in order to more systematically confirm the results found in this study.

Instruction to the performers to choose their own piece might be seen as a limitation of this study. This allowed for strong felt emotions in the Emotional condition. Although movement features were feature-scaled to account for these differences, movement features may have still been affected by this uncontrolled variable. For example, the genre differences (such as between the Rachmaninoff and Debussy pieces) may have elicited different movements, as different musical genres generate different appropriate levels of movement (Huang & Krumhansl, 2011). Additionally, emotion of the piece was not controlled and the different emotions can generate movement differences (Burger, Saarikallio, Luck, Thompson, & Toiviainen, 2013; Dahl & Friberg, 2004). Analysis of effect of different emotions on movement features is still ongoing.

Furthermore, movement may vary greatly between professional performers from very physically active to still, or even change over time. We had a relatively small sample size that may not have accounted for a large enough range of playing styles, thus further study of a wider range of participants is required to confirm the reliability of our results.

The fact that the features extracted were low-level features may also be considered a limitation. Future studies could extract a wider range of movement features as well as using feature reduction techniques such as principle component analysis to explore how higher level features may be related to felt emotion.

Conclusion

Our results suggest that felt mood and emotions influence movement features in a music performance. Higher emotional engagement can create more movement in expressive locations (e.g. head and shoulders), more postural fluctuations and smoother technical movements, though further perception studies are required to confirm if these would be identified in creating a visually convincing performance. We also found that larger expressive movement was associated with positive affect, whereas movement associated with negative affect depended on whether it was performance-related (and therefore related to jerkiness in wrists) or music-related (and therefore related to stillness in posture and more hunching of the shoulders). These results may be beneficial for musicians in the educational as well as the professional domain, especially with performance anxiety, by suggesting one engages more in the emotion of the music to give a more visually expressive and smooth performance. Our study highlights the need for an instrument that would more reliably distinguish between performance and music-related felt emotions, and for further study of other (combinations of) movement features in musical performance.

Acknowledgements. Many thanks to the participants who agreed to take part in this study. We also extend our thanks to our MMT colleagues at the University of Jyväskylä and our anonymous reviewers for their valuable comments.

References

Burger, B., Saarikallio, S., Luck, G., Thompson, M. R., & Toiviainen, P. (2013). Relationships between perceived emotions in music and music-induced movement. *Music Perception*, *30*(5), 517–533.

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In *International Gesture Workshop*, 20–39. Springer, Berlin, Heidelberg.

Castellano, G., Mortillaro, M., Camurri, A., Volpe, G., & Scherer, K. (2008). Automated Analysis of Body Movement in Emotionally Expressive Piano Performances. *Music Perception*, *26*(2), 103–120.

Dael, N., Mortillaro, M., & Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion*, *12*(5), 1085–1101.

Dahl, S. & Friberg, A. (2004). Expressiveness of musician's body movements in performances on marimba. *Lecture Notes in Computer Science*, *2915*(3), 479–486. Retrieved from http://vbn.aau.dk/en/publications/expressiveness-of-musicians-body-movements-in-performances-on-marimba(61ca8680-9eb7-11de-b96d-000ea68e967b).html.

Davidson, J. W. (1991). The perception of expressive movement in music performance. (Unpublished doctoral dissertation). City University London, London, England.

Davidson, J. W. (2007). Qualitative insights into the use of expressive body movement in solo piano performance: a case study approach. *Psychology of Music*, *35*(3), 381–401.

Glowinski, D., Camurri, A., Volpe, G., Noera, C., Cowie, R., Mahon, E. M., & Knapp, B. (2008). Using induction and multiodal assessment to understand the role of emotion in musical performance. In C. Peter, E. Crane, M. Fabri, H. Agius, & L. Axelrod (Eds.) *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2*. (pp. 8-12). Fraunhof: Fraunhofer Verlag.

Griffiths, N. & Reay, J. (2018). The relative importance of aural and visual information in the evaluation of western canon music performance by musicians and nonmusicians. *Music Perception*, *35*(3), 364–375.

Huang, J. & Krumhansl, C. L. (2011). What does seeing the performer add? It depends on musical style, amount of stage behavior, and audience expertise. *Musicae Scientiae*, *15*(3), 343–364.

Isen, A. M. & Levin, P. F. (1972). Effect of feeling good on helping: cookies and kindness. *Journal of Personality and Social Psychology*, *21*(3), 384-388.

Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychology of Music*, *36*(4), 417–427.

Krahé, C. Hahn, U., & Whitney, K. (2015). Is seeing (musical) believing? The eye versus the ear in emotional responses to music. *Psychology of Music*, *43*(1), 140–148.

Kwan, P. Y. (2016). The effect of music performance anxiety, context, modality and observers' music expertise on judgment of musical performances. (Unpublished master's thesis). University of Jyäskylä, Jyväskylä, Finland.

Nusseck, M. & Wanderley, M. M. (2009). Music and movement-how music-related ancillary body movements contribute to the experience of music. *Music Perception*, *26*(4), 335–354.

Persson, R. S. (1993). The subjectivity of musical performance: an exploratory music-psychological real world enquiry into the determinants and education of musical reality. (Unpublished doctoral dissertation). University of Huddersfield, Huddersfield, UK.

Saarikallio, S., Luck, G., Burger, B., Thompson, M. R., & Toiviainen, P. (2013). Dance moves reflect current affective state illustrative of approach-avoidance motivation. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(3), 296–305.

Thompson, M. R. (2007). Expressive gestures in piano performance. (Unpublished master's thesis). University of Jyäskylä, Jyväskylä, Finland.

Thompson, M. R. & Luck, G. (2012). Exploring relationships between pianists' body movements, their expressive intentions, and structural elements of the music. *Musicae Scientiae*, *16*(1), 19–40.

Toiviainen, P., & Burger, B. (2011). *MoCap Toolbox Manual.* Retrieved from University of Jyäskylä: Jyväskylä, Finland: https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mocaptoolbox.

Tsay, C. J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(36), 14580–5.

Van Zijl, A. G. W. & Luck, G. (2013). Moved through music: The effect of experienced emotions on performers' movement characteristics. *Psychology of Music*, *41*(2), 175–197.

Van Zijl, A. G. W., & Sloboda, J. A. (2010). Performers' experienced emotions in the construction of expressive musical performance: An exploratory investigation. *Psychology of Music*, *39*(2), 196–219.

Van Zijl, A. G. W., & Sloboda, J. A. (2013). Emotions in concert: performers' experienced emotions on stage. In G. Luck & O. Brabant (Eds.) *Proceedings of the 3rd International Conference on Music & Emotion (ICME3)*. Finland: University of Jyväskylä.

Vuoskoski, J. K., Thompson, M. R., Clarke, E. F., & Spence, C. (2014). Crossmodal interactions in the perception of expressivity in musical performance. *Attention, Perception, & Psychophysics*, *76*(2), 591–604.

Waddell, G. & Williamon, A. (2017). Eye of the beholder: Stage entrance behavior and facial expression affect continuous quality ratings in music performance. *Frontiers in Psychology*, *8*, 1–14.

Wanderley, M. M., Vines, B. W., Middleton, N., McKay, C., & Hatch, W. (2005). The musical significance of clarinetists' ancillary gestures: an exploration of the field. *Journal of New Music Research*, *34*(1), 97–113.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Expectation Theories: Towards a Noumenal Tonality

Connor Davis

Department of Music, Louisiana State University, United States of America

cdav224@lsu.edu

Abstract

Beginning with John Dewey's "conflict theory of emotion", theories of expectation and their role in aesthetics and affect have been deeply explored (1894). Some examples include Meyer (1956), Narmour (1990), Margulis (2005), and Huron (2006), with significant contributions in between. These theories have common threads of gestalt principles, statistical learning, and the objectivity of the subjective experience via the examination of the stimulus that produces that experience (Meyer, 1956, p. 32). However, even with these common threads, the status of their empirical claims often have philosophical shortcomings that pertain to the categories of phenomenal vs. noumenal (perceived vs. the object/event in itself) and *a priori* vs. *a posteriori* (self-evident vs. empirical)(Kant, 1998). This study aims to unify expectation theories by explicating the component that is shared between all theories. That is, what is the normative principle that allows us to have affective responses via our expectations? In other words, what is it that universally creates expectations apart from genre or surface details? I call this component "*Tonality* as a noun" (Hyer, 2006): the organizing principle that exists before and separate from the music. I argue that through the application of the aforementioned metaphysical and epistemological categories we can address the question of *what* is tonality as a noumenal event (its metaphysics) and *how* we know about this noumenal event through its phenomenology (its epistemology). By finding the common element(s) between various theories of expectation, we begin to understand the root (i.e. noumenal/metaphysical) properties of the component that are expressed in the phenomenal events that have been systematized in various theories of expectation.

Definitions

When considering aesthetic theories, it is helpful to examine the various perspectives from which aesthetics can be defined; that is, we can broadly define aesthetics from the perspective of response theory (RT) (Iser, 1978), from the perspective of the object and its fulfillment of certain analytical traits independent of the consumer's response (essentially formalism), or from the perspective of prototype theory (PT) (Rosch, 1973; Zbikowski, 2002). To be sure, each perspective must be understood before exploring the operationalization of that definition. Moreover, it is apparent from any discussion of aesthetics that these definitions overlap, and none of them are epistemologically sufficient in and of themselves. While my aim for this study is primarily concerned with the relationship of RT to aesthetics, I will include references to its relationship to other perspectives.

Response theory of aesthetics is likely the most inclusive definition of aesthetics as it necessarily focuses on the influences upon a consumer of artwork, which does not necessarily include engaging the aesthetic object as such. This might lead a researcher to explore the sociological conditions, historical conditions, artistic conditions, cultural conditions, cognitive conditions, among many others, when determining the cause for an object's being defined as possessing aesthetic value. Additionally, this definition includes certain assumptions as to how aesthetic value is created, which is to say the object is essentially irrelevant in so far as it is the cognitive organization of that object that is important, and the context in which it is consumed is primary. This assumption carries with it problems that I will examine as relating to the categories of noumenal and phenomenal, which are things in and of themselves and events in which we experience those things, respectively.

Philosophical considerations

One significant problem with response theories taken independently of other considerations is that they do not answer metaphysical questions and ultimately find their roots in Quine's naturalized epistemology (1969). Response theories often do not engage with the noumenal object and thus fall into a sort of circular argument. That is to say, if this epistemology is tasked to explain the possibility of empirical knowledge, it then cannot rely solely on empirical knowledge to justify itself. This problem can be resolved through the exploration and definition of normative features (similar to formalism) in aesthetic response theory. If all that exists is experience of the object, then we cannot actually know that we are experiencing reality. We cannot know that this aesthetically pleasing *thing* is actually aesthetically pleasing because all that exists is our perception of it. This critique of the lack of normativity is clearly articulated by Jaegwon Kim (1988):

> [Quine] is asking us to set aside the entire framework of justification-centered epistemology. That is what is new in Quine's proposals. Quine is asking us to put in its place a purely descriptive, causal-nomological science of human cognition (p. 388).

By this, Kim asserts that a purely naturalized epistemology asks us to ignore *a priori* normativity for an exclusively deductive epistemology of perception and cognition. In short, normativity is central to epistemic matters per se, so to eliminate normativity of the object is to lack certain epistemic groundings in reality as such.

A central feature of normativity that can be identified in music is tonality as a noun (TN) (Hyer, 2006). Tonality as a noun is meant to describe the organizing principle that exists before the music is written or perceived. This is in opposition to tonality as an adjective, which would be a specific cultural organization such as western tonality, or the organizing features of Indian Classical Music. Tonality as a noun is more broadly the temporal organization of sound waves that exist independently of the human cognitive organization of these sounds. Additionally, this normativity as I will demonstrate in TN should not aim to define both necessary and sufficient conditions for aesthetic theory (from all three perspectives), but it should merely seek to define necessary conditions for aesthetic theory. I posit that this necessary condition should be a normative feature, and this normative feature is best expressed in current theories of expectations. I will trace the teleological path of these theories beginning with John Dewey's conflict theory of emotions (1894) through Elizabeth Hellmuth Margulis' model of melodic expectations (2005). Some of the narrative of this teleological explanation is based on Elizabeth Hellmuth Margulis' 2005 publication *A Model of Melodic Expectation.*

Theories of Expectation

The conflict theory of emotions was originally proposed by John Dewey (Dewey, 1894). In short, it states that emotions are results of the termination of a period of anticipation, in which that period of anticipation has multiple conflicting outcomes. The resulting "emotion" is a reaction to how the period of anticipation was actually terminated. This theory is nestled within some of the utilitarian arguments of evolutionary theory, which is to say, he intended to ask of emotion what purpose it served in the preservation of the species (Dewey, 1894, p. 553). Dewey hopes to naturalize emotions and tie them to certain biological responses. He writes, "the rational hypothesis is to suppose that these [emotional responses] are survivals of certain acts, and not symbolic indications of certain emotions" (Dewey, 1894, p. 568). By this, Dewey means to say that emotions in and of themselves do not correspond to their current emotional definitions such as humor, fear, excitement, etc. Instead, he hypothesizes the following, "emotions are, in reality, the reduction of movements and stimulations originally useful into attitudes" (Dewey, 1894, 569). It is important to note the shift in categories; Dewey is critiquing the existence of emotions in the classical, or even Darwinian sense, and creating a new type of emotion which is defined as the biological and psychological result of the fulfillment of expectations given the potential outcomes of a situation.

Another key feature of this theory is the relationship of the affective response to an object. Dewey emphasizes the importance of the affective response being tied to an object (broadly defined to include events). Even if the situation does not contain the object in which the response occurs, the response is still tied to the potentiality of an object and the response is trained from the past encounters with said object. This is important to my metaphysical discussion because even as Dewey's psychological definition of emotion relies on the preservation of the species through its ability to perceive the world and react to it, it is dependent on the object being actual; which is to say, it is dependent on the object possessing noumenal characteristics that are known through the phenomenal realm, and thus it is not solely dependent on the subject's psychological/cognitive organization of their surroundings.

Leonard Meyer took this theory, refined by MacCurdy (1925), and applied it to music. His theory is defined as "the inhibition of a tendency to respond, or, on the conscious level, the frustration of expectation [is] found to be the basis of the affective and the intellectual aesthetic response to music" (1956, p. 43). The first musical example in Meyer's book, recreated below in Example 1, is an exemplar of the psychological underpinnings of this hypothesis at work in music. We are presented with a partial phrase in the key of C-major that moves through the chords $I - ii^6 - V_{4-3}^{6-5}$. What does this progression generate in the mind of (western) listeners? The answer if transparently the expectation of resolution to some degree or another. Moreover, it is more than simply an expectation of certain tones to resolve, but also an expectation that they resolve on beat one. Meyer points out that the resolution of this chord is the fulfillment of the previous progression and that it is ultimately irrelevant as to what or when this resolves (1956, p. 26). Ultimately, the only concern is that this pattern generates a conditioned response within the listener that informs the response to the consequent resolution.

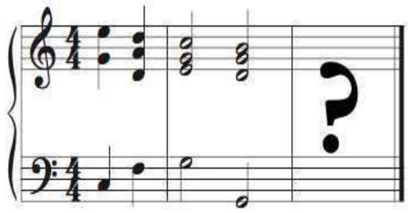


Figure 1. Recreated from *Emotion and Meaning in Music* (Meyer, 1956, p. 25).

Beyond this very clear example, we can have less clear expectations that are created by less universal idioms than those in Example 1. Meyer asks the reader to consider a simple repeated pattern. The mere repetition of the pattern creates expectations that the pattern will eventually end, that expectation generates a response dependent upon the conditions of the repeated pattern (1956, p. 26). These two examples of the role of expectation on musical affect are presented as exemplars of expectation theory that address the object and its relationship to the psychology of the subject. By extension, by identifying what in music actually creates these emotional responses, we can begin to understand aspects of TN as such (i.e. noumenal tonality via its phenomenology).

The preceding examples raise the question of the effect of learning on the conditions of expectation in music. We think in the musical grammar of our culture (1956, p. 43). Meyer defines this grammar as arising through the relationship of five conditions (1956, p. 45);

a) Only some sounds or "unitary sound combinations" are possible

b) Those sounds possible within the system may be plurisituational within the defined limits

c) The sounds possible within the system can be combined in certain ways to form compound terms

d) The conditions stated in a), b) and c) are subject to the probability relationships obtaining within the system

e) The probability relationships prevailing within the system are a function of the context within a particular work as well as within the style system generally.

The above list of conditions hinges on two things, mainly the importance of statistical learning, but also the presence of a system of organization within any given style. If one is to argue that this system is solely the result of cognitive organization and it does not actually exist as such, then that person will likely fall into some of the circular arguments regarding a naturalized epistemology stated above. That is not to make the claim that any given style "exists in nature", it is to make the claim that something in the musical object actually exists independent of the cognitive organization, and it elicits the generally replicable responses in the subject. Additionally, the patterns in a genre, which are the results of the genre, result in the ability of a subject to learn and make predictions based on that genre.

The claims made above are not intended to say that the cognitive organization of the object plays no part in its reality, but merely to point out that cognition cannot be the only thing that exists even as it is vitally important when discussing the phenomenal relationship to the noumenal reality. This is where gestalt principles show themselves to be imperative. According to Meyer (and many others), the organization of stimuli is not "arbitrary", and is actually directed by the various gestalt principles which group seemingly unrelated objects in our surroundings (also see Clark, 2001). See Figure 2, below, which is a graphic of the Kanisza Triangle. This triangle is representative of the law of closure. This law states that objects grouped together are seen as a whole. We can see a floating triangle that doesn't "actually" exist. However, if we deform the things that actually exist on the page (Figure 3), then can study the relationships of the things that actually exist via our cognitive organization of it.

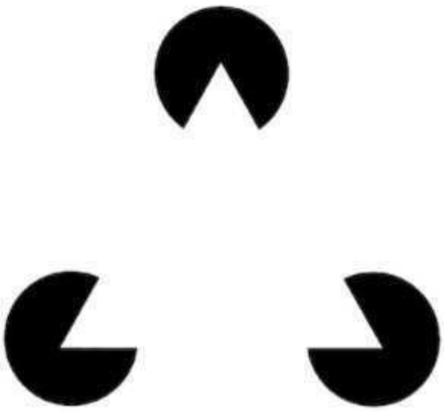


Figure 2. Kanizsa Triangle.

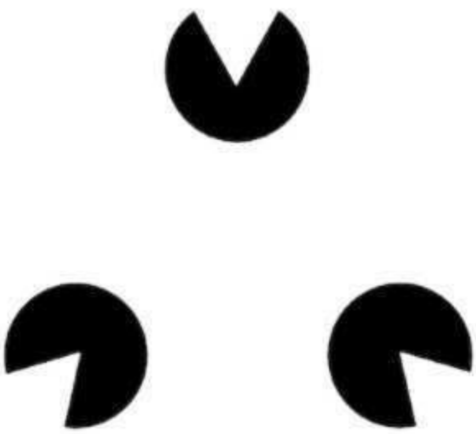


Figure 3. Deformed Kanizsa Triangle.

This presents a key element of my argument for a noumenal tonality, which is that even though these graphics are perceived differently, we can study the "Pac-Men" to tease out what it is we are cognitively organizing. Meyer presents musical analogs to the various gestalt principles, in which particular characteristics of the music cause us to cognitively organize the music in certain ways. For example, grouping step-wise musical line together as a single entity (1956, p. 93-94). While this may seem to argue for TN to be a solely a product of cognition (like the floating triangle), it is not so. The reason is that this perspective is similar to that of studying the "Pac-Men" to tease out cognition vs. what exists. This statement is qualified below in the Synthesis.

Narmour took these gestalt principles to consider expectation theory in the abstract and applied them to melodic expectation. His theory is the implication realization model (I-R), and at its root it contains two logical expressions. The first aspect of his theory is that A + A → A, meaning that one repetition of an event creates the expectation for another repetition. The second aspect is A + B → C, meaning two different events creates the expectation of change. Beyond these larger categories are subcategories of events that relate to gestalt grouping principles, which Margulis points out are all conditioned on interval size and direction (2005, p. 669). This theory of expectation is helpful because it sets out to explore the relationship between gestalt groupings and the potential expectation that is created by these groupings. However, one clarification needs to be made which is that this theory only applies to schematic listening.

Barucha defines two types of listening that would influence one type of expectation. The first is schematic and the second is veridic (1987). Schematic listening occurs when a listener has never listened to a piece (or has no memory of it). In this case, the listener's expectations are dependent upon their knowledge of the particular tonal system/genre. In a veridic listening, the listener is familiar with the piece. The influence of these two types of listening is clearly demonstrated in the V-vi deceptive motion that exists in western-tonality. In the case of schematic listening, the listener will not be aware of the deformation of the prototypical cadence, so the vi chord resolution would likely have influence on the degree to which the listener is surprised. If listening to the same resolution, and the listener knows the vi chord is coming, they would likely have a different type of surprise at least then when listening schematically. The existence of schematic listening adds another fold to my argument for a noumenal tonality as our expectations are based on some organizing principle which necessarily involves our engagement with an object that is organized. Like the "Pac-Men" in the Kanizsa Triangle, we are organizing something that exists.

Margulis built upon the I-R model but sought to make it more empirically predictive rather than potentially predictive. She accomplished this by creating a formula that models the probability that x will follow y. This formula assigns ratings to various aspects of music such as tonal events, metric placement, distance between the notes, etc… These factors are all combined under the heading of Stability (s), Proximity (p), Mobility (m), and Direction (d), and these 4 variables are then combined in the following formula to create a prediction that an event will occur: $z = (s\ p\ m) + d$. There is a prediction value for each possible melodic value, and "a pitch x is expected to follow pitch y by amount z" (2005, p. 683). Beyond this z-value for individual expectedness, Margulis presents the following formula for the overall expectedness based on the hierarchic level of a given prediction.

$$\frac{\Sigma w_i[(s_i \times p_i \times m_i) + d_i]}{\Sigma w_i}$$

Figure 4. "Where i = the level under consideration, w_i = the weight of the level under consideration (15 for the note-to-note level, 5 for levels with spans up to 2 s, 2 for levels with spans up to 6 s), s_i = the stability rating for the pitch or head candidate at that level, p_i = the proximity rating for the pitch or head candidate at that level, m_i = the mobility rating for the pitch or head candidate at that level, and d_i = the direction rating for the pitch or head candidate at that level (Margulis, 2005, p. 690)."

Hierarchy carries with it many different variables to consider including the determination of what might be considered structural and what might not be considered structural. Again, enter gestalt principles. Margulis utilizes Lerdahl and Jackendoff's time-span, which involves choosing a "head" and then determining the influence of that head on a segment of music (Lerdahl and Jackendoff, 1983, p. 124). Margulis mentions that the head is often chosen in regards to a preference-rule, which is arguably not the most empirical method, but, nonetheless, it could feasibly become empirical if one sought to test the preference rules empirically under the construct of Margulis' model. However, Margulis utilizes these preference rules, which are based on gestalt principles, to determine the head and relative time spans that pertain to that head.

Again, here we have a model that is built upon gestalt organizational principles, but I would ask of this model, and all of the other theories discussed, what is it getting at on a deeper level? What are the "Pac-Men" in auditory stimuli that cause us to organize music in the way that we do? Why might we hear a scale as a single unit and a Bach cello suite (i.e. single line polyphony) as several units working together? I would argue the answer is in exploring the concept of TN as it pertains to phenomenology and the noumenal. Kitcher has defended the position of Transcendental Psychology to answer some of the questions of the circular arguments contained in naturalized epistemology. In short, this theory is that the systems of a Kantian Transcendental Idealism map onto certain cognitive conceptions of organization (Kitcher, 1990). In particular, gestalt principles help to combat some of the categorical concerns of the phenomenal realm addressing aspects of the noumenal realm because we can trust our perceptions as pertaining to reality as such.

Synthesis: towards a noumenal tonality

This idea of a noumenal tonality, or TN, can begin to be seen in the consistency and common traits that exist between these various theories of expectation. Contained within each is an understanding that our brain is organizing something in the world, though many, if not all, reject the existence of tonality outside of our cognition of auditory stimuli. Here I would posit that if we are to talk about tonality solely as an organization of "lifeless stimuli" that do not "perpetuate [themselves]" (Meyer, 1956, p. 92), what is it that we are saying about music? How do we rationalize the presence of music and the universal traits of music (or acoustical properties) that seem to be mostly replicable across cultures (see Savage, *et al* 2015)? I would argue that since these theories of expectation deal with auditory signals in time that can be manipulated to produce replicable results, we should explore what it is in the object that is causing this organization. What are the "Pac-Men" in the case of music? I believe that if we do not acknowledge a reality of something in music that is outside of ourselves that we remove an element of musical knowing; the epistemological grounding for musical universals and musical meaning is in question.

The full definition of a noumenal tonality is beyond the scope of this paper as it involves both a fuller explication of expectation theories, as well as an interaction with musical universals and clear and precise definitions of consonance and dissonance via their relationship to tonotopic mapping. None the less, I hope to have demonstrated the epistemological problem and paved a path for a way forward. While this idea of musical universals is a sticky topic given the relationship of theories of expectation and consonance/dissonance are slightly confounded by statistical learning, I believe if we can find the necessary conditions for musical expectation as found in the object, we are a step closer to understanding the world as it is, rather than the world as we perceive it to be.

If tonality can be defined noumenally, then it would be normative for both epistemology and aesthetic judgements. This normativity would unite the formalist definition of aesthetics with prototype theories and response theories. It would provide an element of objectivity when discussing beauty and even a definition for music that is grounded in philosophy, art, and science, drawing from each to inform this definition. It would provide the necessary, but not sufficient definition so as to discuss the relationship of musical objects to one another even if they use different tonal organization systems (i.e. tonality as an adjective). We could also begin to discuss the relationship of TN to other social considerations of aesthetic judgement; that is, by better understanding what it is in the object that causes an aesthetic response we could more accurately talk about the social, societal, and cultural conditions that inform that aesthetic response.

There are currently shortcomings with this theory that need to be addressed. Most clearly is the lack of empirical evidence. Additionally, there is a need for a refined expectation theory that accounts for properties in the sound waves rather than the western tonal system. Also, there consideration of rhetorical parallels to musical expectation would necessarily need to be explored even if they were not part of the necessary definition of noumenal tonality.

To conclude, given the need to explicitly unite historical considerations of aesthetics and philosophy with modern scientific inquiry, it becomes important to unite the concerns of each discipline so they actually inform each other. Science is a form of epistemology, and philosophy can help clarify what science is trying to know via its concern with metaphysics as they relate to epistemology. My future work will explore the relationship of aesthetics, science, and philosophy and refine this theory to provide a better answer to philosophical concerns of music cognition and vice versa.

References

Allison, H. (1996). On naturalizing Kant's transcendental psychology. In idem, Idealism and freedom: Essays on Kant's theoretical and practical philosophy (pp. 52–66). Cambridge: Cambridge University Press.

Altieri, C. (2003). *The particulars of rapture: An aesthetics of the affects*. Ithaca, NY: Cornell University Press.

Angier, R. (1927). The Conflict Theory of Emotion. *The American Journal of Psychology, 39*(1/4), 390-401.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5,* 1–30.

Clarke, E. (2001). Meaning and the Specification of Motion in Music. *Musicae Scientiae: The Journal of the European Society for the Cognitive Sciences of Music*, 5(2), 213–234.

Dewey, J. (1894). The theory of emotion. *Psychological Review, 1,* 553–569.

Granot, R., & Donchin, E. (2002). Do re me fa sol la ti—constraints, congruity, and musical training: An event-related potentials study of musical expectancies. *Music Perception, 19,* 487–528.

Huron, D. (2006). *Sweet anticipation: music and the psychology of expectation.* Cambridge, MA: MIT Press.

Hyer, B. (2006). Tonality. In T. Christensen (Eds), *The Cambridge History of Western Music Theory* (pp.726-752). Cambridge: Cambridge University Press.

Iser, W. (1978) *The Act of Reading: A Theory of Aesthetic Response*. Baltimore, MD: The Johns Hopkins University Press.

Kant, I. (1998). Critique of pure reason (W. Pluhar Trans.). Indianapolis/Cambridge: Hackett Publishing Company.

Kant, I. (2000). Critique of the power of judgment (P. Guyer, Ed.; P. Wood, & E. Matthews, Trans.). Cambridge: Cambridge University Press.

Kitcher, P. (1990). Kant's transcendental psychology. New York: Oxford University Press.

Kitcher, P. (2006). Kant's philosophy of the cognitive mind. In P. Guyer (Ed.), The Cambridge companion to Kant (pp. 169–202). Cambridge: Cambridge University Press.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Lerdahl, F. (2001). *Tonal pitch space.* New York: Oxford University Press.

MacCurdy, J. (1925). *The psychology of emotion.* New York: Harcourt, Brace and Co.

Margulis, E. H. (2005). A model of melodic expectation. *Music Perception,* 22, 663-714.

Meyer, L. B. (1956). Emotion and meaning in music. Chicago: University of Chicago Press.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures.* Chicago: University of Chicago Press.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience,* 6, 688–691.

Quine, W.V.O. (1969), *Ontological Relativity and Other Essays*, New York: Columbia University Press.

Rosch, E.H. (1973). Natural categories. *Cognitive Psychology* 4, 328-350

Savage PE, Brown S, Sakai E, Currie TE (2015) Statistical universals reveal the structures and functions of human music. *Proc Natl Acad Sci USA, 112*, 8987–8992.

Schellenberg, E. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition, 58,* 75–125.

Schellenberg, E. G. (1997). Simplifying the implication-realization model of melodic expectancy. *Music Perception, 14,* 295–318.

Schellhous, R. (1991). Fétis's "Tonality" as a Metaphysical Principle: Hypothesis for a New Science. *Music Theory Spectrum, 13*(2), 219-240.

Zbikowski, L. M. (2002). *Conceptualizing music: cognitive structure, theory, and analysis*. Oxford; New York: Oxford University Press, 2002.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Investigating the Associations between Nonverbal Skills and Musical Abilities in Children without Musical Training

Delmolin, G[1]., Andrade, O.V[2]., Andrade, P[3]., Vanzella, P[4,5].

[1] *Universidade Federal do ABC,*
[2] *National Union of Private Higher Education Institutions, Brazil,*
[3] *Goldsmiths University of London, United Kingdom.*
[4] *Interdisciplinary Unit for Applied Neuroscience, Universidade Federal do ABC, São Bernardo do Campo, São Paulo, Brazil*
[5] *Department of Music, Universidade de Brasília, Brasília, Brazil*
[1]guiharris@gmail.com, [2]paulo_sustain@yahoo.com, [3]olguinhavandrade@yahoo.com, [1]p.vanzella@ufabc.edu.br

Abstract

Musical practice in childhood is associated with significant gains in spatial-temporal reasoning abilities, even when controlling for socioeconomic status (SES) (Bilhartz, Bruhn, & Olson, 1999). This study investigated the relationship between simple music perception abilities and nonverbal reasoning in Brazilian children from public schools. Our goal was to investigate the association between simple melodic and rhythmic perceptual abilities, processing of musical sequences, nonverbal IQ, visual memory, and SES in children from public schools without musical training in Brazil. Seventy-three (n=73) students from public schools (mean age: 8.43, SD: 0.689) completed the Nonverbal Intelligence Test for Children (R-2) and visual and visuospatial memory tasks from the Cognitive-Linguistic protocol (adapted from the Benton Visual Retention Test) (Capellini, Smythe, & Silva, 2012). The Montreal Battery for Evaluation of Musical Abilities (MBEMA) (Peretz et al., 2013) and the Music Sequence Transcription Task (MSTT) (Zuk, Andrade, Andrade, Gardiner, & Gaab, 2013) were chosen to measure the musical perception skills from musically untrained children given their simplified paradigm and because MSTT is a valid music-based task suited for classroom settings. Both the MSTT ($r = .367$, $p < .001$) and the Rhythm task from MBEMA ($r = .324$, $p = .005$, $p < .01$) were significantly correlated with the R-2. The MSTT showed a correlation with Figure order ($r = .371$, $p < .005$). When controlling for SES and sex, rhythm task ($r = .350$, $p = .024376$) showed a correlation with the R-2 and MSTT ($r = .330$, $p = .05$) showed a correlation with Figure order. A multiple regression was run to predict Figure order from MBEMA and MSTT scores. These variables significantly predicted Figure Order, $F(3, 73) = 4.786$, $p < .005$, $R^2 = .164$. Only MSTT was statistically significant to the prediction, $p < .05$. Correlational analysis corroborate the notion that both the melodic and rhythmic dimensions may involve visuospatial mechanisms relevant to executive functions regardless of musical training, SES, and sex. Multiple regression analysis suggests that musical tasks such as MSTT may be relevant to predict visual memory in this population. Unlike MBEMA, MSTT does not require fine tonal discrimination and may involve sequential auditory processing relevant to executive functions.

Introduction

Several authors have investigated the associations between the development of musical abilities and cognitive skills such as intelligence, language, memory and visuospatial abilities (Miendlarzewska, Trost, Sammler, & Planck, 2014; Schellenberg & W. Weiss, 2013; Swaminathan, Schellenberg, & Venkatesan, 2018). There is a comprehensive field of evidence suggesting that learning to play a musical instrument in childhood stimulates cognitive development and leads to improved skills in a wide array of areas (Kraus & Chandrasekaran, 2010; Rajalakshmi, 2018; Tierney & Kraus, 2013). Some authors have shown that musical practice in childhood is associated with enhanced nonverbal reasoning and verbal abilities (Forgeard, Winner, Norton, & Schlaug, 2008), and significant gains in spatial–temporal reasoning abilities as well, even when controlling for sex, ethnicity, parental education, and socioeconomic status (SES; Bilhartz et al., 1999).

Although musical abilities are innate and normally distributed in the population, even in the absence of musical training (Schellenberg & W. Weiss, 2013; Stalinski & Schellenberg, 2012), assessing musical skills in childhood is challenging. Both training and normal maturation occur simultaneously, making our understanding of the effects of musical experience of cognitive and motor development difficult. As Galván (2010) suggests, development and learning would not be two completely separate constructs, but instead they exist on a continuum. In this sense, several variables should be taken into account when investigating child development. Hackman & Farah (2009), for example, showed how socioeconomic status (SES) can interfere in the development of a variety of neurocognitive abilities (most robust disparities between SES and cognitive skills are evident in the domains of language and executive functions); such findings should be considered critical in the investigation of the basic functions of the human brain.

In this study, we investigated the relationship between simple music perception abilities and nonverbal reasoning in a sample of musically untrained children from public schools in Brazil. Given the scarcity of studies with this population and the disparities in academic achievement between public and private schools in Brazil, our aim was to investigate if basic musical and perceptual skills would be related to nonverbal reasoning abilities in this population. Perceptual abilities such as listening and comparing melodic and rhythmic content were tested using the Montreal Battery for Evaluation of Musical Abilities (MBEMA) – a task developed as an objective, short and up-to-date test of musical abilities used in a variety of situations, from the identification of children with musical difficulties to the assessment of the effects of musical training in typically developing children of different cultures (Peretz et al., 2013). To access a motor response given by the transcription of musical sequences, we used the Music Sequence Transcription Task (MSTT), a task designed to preferentially engage cognitive and perceptual mechanisms dedicated to "auditory pattern sequencing" including auditory working memory but also involving other operations, such as the mapping of perceived sounds onto written symbols as well as a decision-making component followed by a subsequent motor response (Zuk et al., 2013). Cognitive skills were measured using both the Nonverbal Intelligence Test for Children – R-2 (Rosa & Alves, 2000), a Brazilian adaptation for children of the *Raven's Progressive Matrices* and the Cognitive-Linguistic Protocol (Capellini et al., 2012), developed in Brazil with the aim to identify the cognitive-linguistic profile in the first stages of reading acquisition of Brazilian children (Andrade, Andrade, & Capellini, 2015).

Materials and Methods

Participants

This study was carried out with a non-referred school-based sample of 73 (35 girls) native speakers of Brazilian Portuguese enrolled in the 2nd and 3rd grades of elementary education (corresponding to the same grade levels in the American education system) in two public schools in Santo André (State of São Paulo, Brazil). Age ranged from 7.3 to 11.5 years (M = 8 and 5 months, SD = 8 months), and the whole sample comprised of 37 2nd graders (M = 8, SD = 8 months, 17 girls) and 36 3rd graders (M = 8 and 10 months SD = 4 months, 18 girls). Selection criteria for participation in the study were: parental permission granted, no major sensorimotor handicaps (deafness, blindness), no pervasive neurodevelopmental disorders (psychosis, autism), an adequate command of their native language, and normal intelligence as assessed by the Nonverbal Intelligence Test – R2 for children (Rosa & Alves, 2000) which is a Brazilian adaptation for children of the *Raven's Progressive Matrices* (Raven, Raven, & Court, 1993).

Behavioral Measures

Psychometric measures. Visual processing tasks were assessed with the Cognitive-Linguistic Protocol (Capellini et al., 2012), hereafter referred to as CS&S protocol, developed in Brazil with the aim of producing an effective instrument to identify the cognitive-linguistic profile in the first stages of reading acquisition of Brazilian children. Because of time constraints imposed by the schools and based on results of previous investigations, we have opted to use in this study only those tasks of the protocol with higher loadings on the linguistic and visual components (Andrade et al., 2015; Zuk et al., 2013).

Visual memory (individually administered). The visual short-term memory task from CS&S protocol is based on the Benton Visual Retention Test (BVRT) (Sivan, 1992). In the original test, one or more simple geometric figures are drawn from memory after a brief exposure (typically 5 seconds for adults and 10 seconds for children). Two types of scores are computed: the number of correct representations and the number of errors. In the visual short-term memory task from CS&S protocol, instead of drawing from memory, children look at the figures for 10 seconds and then receive the same figures on cards in random order and are asked to organize them in the same order and position they just have seen.

(i) *Figure order*: Ordering of solid abstract figures, i.e., after seeing two to five-figure sequences over a 10-second period, the student was asked to reassemble the pattern of figures in the same previous order and rotation. The score is given by the number of figures correctly ordered, with a maximum score of 28.

(ii) *Figure rotation hit:* Number of hits made in rotation of the shapes in figure order task. 28 figures were presented throughout the sequences, leading to a maximum of 28 hits.

Nonverbal Intelligence Test – R2 for children (individually administered). Fluid intelligence was assessed by the Nonverbal Intelligence Test – R2 for children (Rosa & Alves, 2000), a Brazilian adaptation for children of the *Raven's Progressive Matrices* (Raven, Raven, & Court, 1993). The R-2 Test consists of 30 boards, or cards, with colored figures, to be used in the sequence of their numbering, with figures of concrete objects or abstract shapes. The task to be performed by the child is similar to that of Raven's Colored Progressive Matrix Test, in which there is a figure in each item to be completed with one of the alternatives presented below the figure, according to the type of reasoning involved on that item. Some items have six alternatives, others have eight, and only one is correct. The application is performed individually and the responses are recorded by the examiner. The test has no time limit since it is a capacity/power test. In the standardization sample, the mean time was 8 minutes and the test was considered adequate for children between 5 and 11 years of age.

Socioeconomic measures. SES was determined using three indicators: socioeconomic class (ABEP index), monthly family income, and parental education. Socioeconomic class was determined by the Brazilian Association of Research Companies questionnaire (ABEP, 2016). The questionnaire includes questions on the education level of the head of the household, ownership of goods (radio, TV, refrigerator, washing machine, VCR/DVD and car), presence of housemaid and housing features. Most of these items are used to calculate an index of SES called the "Criteria for Economic Classification", which is widely used in publicity, political polls and research in Brazil. The Criteria for Economic Classification is a point system with established weights for each item. The total number of points is used to place the individual in a SES category designated as A (A1, A2), B (B1, B2), C (C1, C2), D or E, where A is the highest and E is the lowest category.

Musical Tasks

Abbreviated MBEMA (collectively administered). We used the Montreal Battery of Evaluation of Musical Abilities (MBEMA), a novel tool for assessing musical abilities in children (Peretz et al., 2013). Musical stimuli in the MBEMA are shorter versions of the melodies used in the adult MBEA (Peretz, Champod, & Hyde, 2003); melodies are 5–9 tones long, 3–4 seconds in overall duration, and are played in ten different keys (half major, half minor) and ten different timbres as well (e.g., piano, marimba, guitar, flute) (Peretz et al., 2013, p. 3). Due to time constraints, we opted to administer the abbreviated version of the MBEMA consisting of one musical *rhythm* test and a *melody* test, which requires the identification of differences with regard to scale, contour and interval in pairs of short melodies. Each test comprises two practice trials and 20 test trials constructed from the same 20 melodies, being 10 trials with identical (same) comparison melodies and 10 trials with violating (different) comparison melodies randomly presented, thus requiring just a same-different response.

In the *rhythm* test, the original meter is always preserved in the comparison melody, and the violations consist of duration alterations between two adjacent notes altering only the temporal grouping. Each trial is preceded by a warning tone followed by 500ms, and consists of a target melody and a comparison melody separated by a 1.5-s silent interval; trials are separated by 4-s silent intervals (for more details see Peretz et al., 2013).

Musical Sequence Transcription Task (MSTT) (collectively administered). The MSTT is a task that requires the active transcriptions of simple auditory sequences and was first described in Zuk et al. (2013). The MSTT trials consist of four-sound isochrounous sequences comprised of only two-note chords (dyads), one dyad in the low register [an interval of a perfect fifth, with fundamental frequencies 110 (A) and 165 Hz (E)] and the other dyad in the high register [a perfect fourth, with 330 (E) and 440 Hz (A)] which children learned to code as the "thick" and the "thin" sounds, respectively (for more details see Zuk et al., 2013). Whereas in the previous study MSTT stimuli were played on the acoustic guitar by the examiner administering the task collectively in the classroom, in the present study the same sound stimuli were recorded and played over speakers. The stimuli were synthesized using the acoustic guitar timbre of the Guitar Pro 6 (AROBAS MUSIC, 2010) which is a digital audio workstation that is used to record, manipulate or synthesize music. Children listened to each one of the 20 trials of four-sound sequences presented in a slow, isochronous manner, consistent in tempo throughout the entire task (80 beats per min) and were instructed to code ("to write") each of 20 four-sound (the lower pitched sound) and a circle "O" for the "thick" sound (the lower pitched sound). As an example, a sequence of two consecutive dyads in the high register, one dyad in the low register, and one last dyad in the high register should be coded as "| | O |". The time interval between each sequence was determined by the examiner, taking into account the moment when all students were prepared to listen to the next sequence. Children received detailed instructions and performed two to five training sequences before the start of the test trials.

Scoring criteria for the MSTT. There were 20 trials (sequences). Only the accurate coding of the four sounds in the correct order was considered a correct response and was given a score of one point, thus leading to a maximum score of 20 for the total task. When less than four sounds were marked for a particular trial (insufficient response, IR), even if partially correct, the trial was incomplete, and thereby scored as zero; if more than four sounds were recorded, even with the four first sounds being in the correct order, this was also considered incorrect and scored as zero for the trial and coined as superfluous responses (SR).

Procedure

An initial sample of 77 participants was tested on the tasks from the CS&S protocol, the abbreviated version of MBEMA, the MSTT task and the Nonverbal Intelligence Test – R2 for children (Rosa & Alves, 2000). Because four participants performed at the 30th percentile or lower on the R2 test they were excluded from the subsequent data analysis leaving 73 in the sample. Data collection occurred over the course of the middle of the academic calendar year and took place during school hours. Families who accepted study participation received a packet containing the ABEP questionnaire and were instructed to complete and return the instrument within two weeks.

Statistical Analysis

To investigate the extent and nature of relationships between Musical tasks and Nonverbal assessments, Bonferroni-corrected correlations partialling out age, sex and SES were implemented through SPSS software package, version 20.0 (SPSS Inc., Chicago, IL, USA) for Windows. A multiple regression was run to describe data and the relationship between musical tasks and nonverbal domains. An alpha level of .05 (two-tailed) was set for statistical significance.

Results

Most families were categorized in the middle class B2 (35.6%): A (5.1%), B1 (20.3%), C1 (23.7%), C2 (13.6%) and D (1.7%). With regard to family income, 30.4% reported monthly earnings of $830, 23.2% reported earning up to $550 monthly, 21.4% reported earning up to $1675, 14.3% reported earning up to $310 and 10.8% reported earning between $2200 and $3050 monthly. Most parents had completed high school (mothers: 40.7%; fathers: 43.4%), 22% of the mothers had a college degree whereas only 11.3% of fathers had a college degree. 8.5% of mothers and 11.3% of fathers did not finish the elementary school. Descriptive statistics are show at Table 1, providing means and SDs on the music and nonverbal measures.

Table 1. Means (and SDs) on the Music and Cognitive-Linguistic Measures

	n	Mean (maximum value)	*SD*
Age in years	73	8.43	8.268
MSTT	73	10.56 (20)	5.679
MBEMA Melody	72	13.37 (20)	3.313
MBEMA Rhythm	72	14.76 (20)	4.313
Figure Order	73	16.79 (28)	3.295
Rotation	73	22.26 (28)	4.919
R-2	73	78.138 (100)	20.127

As shown in the Table 2, Bonferroni-corrected correlations of music tasks with nonverbal abilities revealed that both the

MSTT ($r = .416$, $p < .001$) and the Rhythm task ($r = .385$, $p <.01$) were significantly correlated with the R-2 whereas the correlation between Melody task and R-2 ($r = .264$, $p = .024$) did not survive Holm-Bonferroni correction. The MSTT showed a correlation with Figure order ($r = .362$, $p < .01$). Rhythm task had a marginally significantly correlation with Rotation of figures ($r = .290$, $p = .0283$, $p = .07$). Finally, correlation between Melody task and Rotation of figures was marginally significant ($r = .218$, $p = .068$).

Table 2. Holm-Bonferroni corrected partial correlations between musical tasks and nonverbal abilities

	Nonverbal Dimension		
	Figure Order	Rotation	R-2
MSTT	.371***	.232	.367***
Melody	.236	.237	.226
Rhythm	.238	.350**	.324**

Footnote. Values of Spearman's correlation coefficient are reported in this table. * indicates a p-level of <.05, ** a level of, <.005, *** indicates a p-level of, <.001. Correlations controlled for Sex, Age and SES.

When controlling for SES and sex, rhythm task ($r = .370$, $p = .024376$) correlated with the R-2 and MSTT ($r = .330$, $p = .046275$) correlated with Figure order. A multiple regression was run to predict Figure order from MBEMA and MSTT scores. These variables significantly predicted Figure Order, $F (3, 73) = 4.786$, $p < .005$, $R^2 = .164$. Only MSTT was statistically significant to the prediction, $p < .05$.

Conclusion

These data suggest a relationship between performance in a variety of musical perception tasks and nonverbal skills in a sample of musically untrained children from low-income families. Correlational analysis corroborate the notion that both the melodic and rhythmic dimensions may involve nonverbal reasoning relevant to executive functions regardless of musical training, SES and sex. Multiple regression analysis suggests that musical tasks such as MSTT may be relevant to predict visual memory in this population. Unlike MBEMA, MSTT does not require fine tonal discrimination and may involve sequential auditory processing relevant to executive functions. In this paper, we present evidence of associations between basic musical perception skills and non-verbal reasoning in a population of Brazilian children that had not yet been investigated, taking into account socioeconomic status. Another important contribution of this work was the use of musical tasks of a collective application, especially the MSTT, which presents an ecologically valid musical experience for the context of classrooms.

References

ABEP. Alterações na aplicação do Critério Brasil, 1 ABEP 1–5 (2016). São Paulo. https://doi.org/10.12820/rbafs.v.6n2p5-18

Andrade, O. V. C. a., Andrade, P. E., & Capellini, S. a. (2015). Collective screening tools for early identification of dyslexia. *Frontiers in Psychology*, *5*(January), 1–13. https://doi.org/10.3389/fpsyg.2014.01581

AROBAS MUSIC. (2010). Guitar Pro. AROBAS MUSIC.

Bilhartz, T. D., Bruhn, R. A., & Olson, J. E. (1999). The effect of early music training on child cognitive development. *Journal of Applied Developmental Psychology*, *20*(4), 615–636. https://doi.org/10.1016/S0193-3973(99)00033-7

Capellini, S. A., Smythe, I., & Silva, C. (2012). *Protocolo de Avaliação de Habilidades Cognitivo-Linguísticas*. (1º, Ed.). Book Toy.

Forgeard, M., Winner, E., Norton, A., & Schlaug, G. (2008). racticing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. *PLoS ONE*, *3*(10), 1–8. https://doi.org/10.1371/journal.pone.0003566

Galván, A. (2010). Neural plasticity of development and learning. *Human Brain Mapping*, *31*(6), 879–890. https://doi.org/10.1002/hbm.21029

Hackman, D., & Farah, M. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, *13*(2), 65–73. https://doi.org/10.1016/j.tics.2008.11.003.Socioeconomic

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Miendlarzewska, E. A., Trost, W. J., Sammler, D., & Planck, M. (2014). How musical training affects cognitive development : rhythm , reward and other modulating variables. *Frontiers in Neuroscience*, *7*(January), 1–18. https://doi.org/10.3389/fnins.2013.00279

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14681118

Peretz, I., Gosselin, N., Nan, Y., Caron-Caplette, E., Trehub, S. E., & Béland, R. (2013). A novel tool for evaluating children's musical abilities across age and culture. *Frontiers in Systems Neuroscience*, *7*. https://doi.org/10.3389/fnsys.2013.00030

Rajalakshmi, M. A. (2018). A Review of the Effects of Early Exposure to Music on Brain Development A Review of the Effects of Early Exposure to Music on Brain Development. *International Journal of Advanced Scientific Research and Management*, *3*(February), 56–59.

Rosa, H. R., & Alves, I. C. B. (2000). *Coleção R-2 - Teste Não-Verbal de Inteligência para Criança*. São Paulo: Editora Vetor.

Schellenberg, E. G., & W. Weiss, M. (2013). Music and Cognitive Abilities. In *The Psychology of Music* (pp. 499–550). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00012-2

Stalinski, S. M., & Schellenberg, E. G. (2012). Music Cognition: A Developmental Perspective. *Topics in Cognitive Science*, *4*(4), 485–497. https://doi.org/10.1111/j.1756-8765.2012.01217.x

Swaminathan, S., Schellenberg, E. G., & Venkatesan, K. (2018). Explaining the Association Between Music Training and Reading in Adults. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000493

Tierney, A., & Kraus, N. (2013). *Music Training for the Development of Reading Skills*. *Changing Brains - Applying Brain Plasticity to Advance and Recover Human Ability* (1st ed., Vol. 207). Elsevier B.V. https://doi.org/10.1016/B978-0-444-63327-9.00008-4

Zuk, J., Andrade, P. E., Andrade, O. V. C. a., Gardiner, M., & Gaab, N. (2013). Musical, language, and reading abilities in early Portuguese readers. *Frontiers in Psychology*, *4*(June), 1–12. https://doi.org/10.3389/fpsyg.2013.00288

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Musical emotional processing in patients with Alzheimer's Dementia

Detlefsen, Veronica,[1] Moltrasio, Julieta[2], Mora Milena[3], Rubinstein Wanda[4]

[1,2,3,4]*Universidad de Buenos Aires, Argentine Republic*

[4]*Consejo de Investigaciones Científicas y Técnicas (CONICET), Argentine Republic*

[4]*Hospital Interzonal de Agudos (HIGA) Eva Perón*

[1]verodetlefsen@hotmail, [2]julietamoltrasio@gmail.com, [3]milenamora93@gmail.com, [4]wrubinstein@live.com

Abstract

Music produces intense emotional reactions, similar other stimuli. The ability to process musical stimuli requires a group of processing components, and involves different brain regions. Patients with Alzheimer's Disease (AD) present episodic memory deterioration and deficits in the emotional processing of visual stimuli. However, music emotional judgment is preserved. The aim of this study is to carry out a bibliographic review of preserved and impaired musical abilities in patients with AD. Studies showed that patients with AD show impaired performance in some aspects of musical processing, such as global. In spite of that, the emotional processing of melodies is preserved. The emotional processing component would be preserved in AD. And the areas activated are similar to healthy subjects, such as the cingulate gyrus, the cerebellum, and frontal areas. Music seems to aim to processes and areas that are preserved in AD. The study shows the importance of studying emotions in music, focusing on what is preserved in AD. This could be the key to explain the relative conservation of musical memory in this pathology, and to develop music-based treatments.

Keywords: Alzheimer's Disease, Emotional judgement, Music

Introduction

Emotions and motivation involve a stimulus and the events that move an organism to action. A classic way of conceptualizing emotions consists of the dimensional model (Bradley & Lang, 2000): we categorize the world according to how much a stimulus promotes (positive valence) or threatens (negative valence) life, and how much it intensifies the approach or flight towards or from a stimulus (arousal). These parameters have been used in conditioned visual stimuli, such as photographs (Lang, Bradley, & Cuthbert, 1999), using different age groups and populations with neurodegenerative pathologies, such as AD (Kesinger, Brierley, Medford, Growdon, & Corkin 2002, Satler et al., 2007, Li et al., 2016, Gómez-Gallego & Gómez-García, 2017). That is, different populations perform similar emotional judgements of these stimuli.

At a neuroanatomical level, the study of the emotional processing, focused on the responses of the autonomic nervous system (Bradley & Lang, 2000, Kandel, Jesell, & Schwarts, 1997): sweating, heart rate, blood pressure, temperature. Moreover, various structures of the limbic system have been associated with the emotional response, such as the hypothalamus and the amygdala (Bradley & Lang, 2000, Damasio, 1994, LeDoux, 1999). It is worth mentioning the participation of dopamine in the pleasure and reward circuit (Cummings & Coffey, 1994). Stimuli that naturally produce a reward, such as food, drink and sex, increase dopamine release in the nucleus accumbens, which are projected into areas of the limbic system (Bressan & Crippa, 2005).

Music produces intense emotional reactions, similar to those produced by other stimuli (Blood, Zatorre, Bermudez, & Evans, 1999, Blood & Zatorre, 2001). In addition, it can increase the emotional experience (in psychophysiological measures, such as skin conductance) of visual stimuli (Baumgartner, Lutz, Schmidt, & Jancke, 2006). The emotional judgement of a piece or musical extract (in terms of valence and arousal) is done quickly, immediately and automatically (Eschrich, Münte, & Atenmüller, 2008, Peretz, Gagnon, & Bouchard, 1998). We be moved by a novel melody and, subjects of the same culture, perform a similar qualification (Peretz, Gagnon, & Bouchard, 1998, Vieillard et al., 2008). This facility for musical emotional judgment is preserved in pathologies such as AD (Omar, Hailstone, Warren, Crutch, & Warren, 2010, Hsieh, Hornberger, Piguet, & Hodges, 2012).

The aim of this study is to carry out a bibliographic review of preserved and impaired musical abilities in patients with AD. As well as the cognitive models and neurobiological bases involved in musical processing in relation to emotions. We hypothesize that the recognition of emotions in music is preserved in AD.

Methods

Cualitative search in databases of scientific publications, such as Google Scholar and PubMed. The articles considered had to comply with the following characteristics: published articles, both national and international, in journals with peer review process. The articles that were included are group studies on musical emotional processing in subjects without pathology and patients with AD, including neuroimaging, psychophysiological and / or psychometric measurements. The articles related to music and episodic memory improvement were excluded from the search. There was no criterion of exclusion due to seniority, as classic articles were also included.

Modularity of Musical Processing and Neural Bases

Peretz and Coltheart (2003) proposed an architectural model of musical processing: an information processing system including a musical module, whose specific operation is musical processing. The model contains components whose domains are specific to music, and components whose domains aren't specific to music. According to the model, the musical input modules are organized in two subsistems: the temporal (involving both tempo and rhythm) and melodic (involving the contour analysis of melody, scale and intervals), which send outputs to the musical lexicon and the emotion expression analysis component. The first contains all the representations of the specific musical phrases to which one has been exposed during one's lifetime. The emotion expression analysis component, allows the subject to recognize and experience the emotion aroused by the music. Two structural properties determine the emotional judgment of a piece: the mode (minor, major) and the tempo (slow, fast). The tempo would be the determinant of the level of arousal (activation or relaxation), while the modality would determine the valence (pleasant / positive or unpleasant / negative), as well as the emotional categories: joy, sadness, threat or terror and peace (Vieillard et al., 2008). They claim that the specifity to music of the emotion expression analysis is currently unknown.

Koelsch and Siebel (2005; 2011) also proposed a modular neurocognitive model for musical perceptual processing, but they incorporated the anatomical areas responsible for the specific processing. According to their model, musical perception begins in the cochlea, then involves the auditory brainstem and thalamus, where a primary processing of the auditory signal register is performed. After that, the primary and secondary auditory cortex allow a more refined processing of the acoustic characteristics, such as pitch, timbre, intensity and roughness. At the same time, acoustic information enters auditory sensory memory, long-term memory and working memory and representations of Gestalt are formed. In this module, melodic, timbral, rhythmic and spatial grouping processes are carried out, such as the conformation of the musical contour. Its function is to recognize and follow an acoustic object in an auditory environment on the temporal plane. At a higher level, although related to the previous one, the analysis of the intervals continues, which involves the processing of the chords, the individual tones and the temporal intervals of a melody. The processing of the chords involves bilateral and prefrontal temporal regions, while the areas involved in the analysis of the intervals are the supratemporal anterior and posterior bilateral cortex.

The next level of processing corresponds to the construction of the musical syntactic structure, which involves elements of a musical work, such as tones, intervals, chords and rhythm, present a relational organization (Limb, 2006). This level gives the creation of expectations, from the internalization of the rules of a musical culture, without the need of formal academic study. The areas activated in the musical syntax are the opercular pairs of the bilateral inferior frontolateral cortex, especially right hemispheric areas (Koelsch & Siebel, 2005). The processing of the tempo occurs mainly in the left hemisphere, activating the premotor, left parietal and right cerebellar areas (Limb, 2006). When listening to an irregular or unexpected element, the process of structural analysis and repair can be initiated, both for the melody (Besson & Faïta, 1995) and for harmonic accompaniment (Patel, Gibson, Ratner, Besson, & Holcomb, 1998).

At a neuroanatomical level, music, as an emotional stimulus, activates the neural systems of reward and emotion, mentioned in the first section (Blood & Zatorre, 2001). The areas of the left ventral striatum and the dorsomedial midbrain, and in the supralimbic regions such as the right orbitofrontal cortex, the bilateral insula, and other regions associated with excitation processes (such as the thalamus and anterior cingulate cortex), motors (motor area) supplementary, and cerebellum, have been activated during listening to pleasant music. In addition, a decrease in activation of the right amygdala left hippocampus and the ventromedial prefrontal cortex, has been detected.

Blood and collaborators (1999) studied the emotional responses to melodies, according to the degree of consonance and dissonance. The harmonic base of the melody was versioned in different levels of decreasing consonance to increasing dissonance. According to the rules of western music, the more consonant a melody is, the more pleasant it sounds, and the more dissonances it possesses, the more unpleasant it seems. Paralímbic and neocortical areas where involved in musical emotional processing. The areas activated during the increase of the dissonances were the right hippocampal and right precuneus gyrus, while the regions activated by the increase of the consonances (or decreasing dissonances) were the bilateral orbitofrontal cingulate, the medial subcallose cingulate region and the right frontal pole. They concluded that the anatomical structures involved in emotional processing differ from the areas responsible for musical perception.

Alzheimer Disease

AD is the most frequent form of dementia, being responsible for more than 60% of cases of dementia (Alzheimer Foundation Spain, 2016). The disease has an insidious onset and a gradual progression. One of the main characteristics is the deterioration in episodic memory (Bäckman & Small, 1998, Carlesino & Oscar-Berman, 1992), and, for its diagnosis, a decline of at least another cognitive domain must be present (American Psychiatric Association, 2013). In addition, cognitive deficits interfere with the performance of activities of daily living. The brain areas affected in the AD are mainly those related to memory deficits: medial temporal lobe (Russo, 2015), as well as the amygdala, and temporal, parietal and frontal multimodal association cortices (Robles et al., 2002; Bartoloni & collaborators, 2015).

Despite the deficits that characterize this pathology, the recognition of emotions in music seems to be preserved. Using musical parameters (Peretz & Colheart, 2003; Vieillard et al., 2008), as mentioned in the previous section, patients with mild AD perform an emotional attribution similar to that performed by healthy adults (Drapeau, Gosselin, Gagnon, Peretz, & Lorrain, 2009, Gagnon, Peretz & Fülöp 2009, Samson, Dellacherie & Platel, 2009, Omar et al., 2010, Hsieh et al., 2012). That is, they are able to distinguish between happy and sad melodies (Gagnon et al., 2009; Samson et al., 2009), between melodies that evoke feelings of joy, peace, sadness and threaten (Drapeau et al., 2009), and they perform better in this type of task than patients with other types of dementia (Omar et al., 2010; Hsieh et al., 2012). In addition, alteration in the recognition of emotions in the face was observed, with a preserved emotional musical recognition (Drapeau et al., 2009). The mentioned studies used diverse musical stimuli, according to the western culture rules, and all of them were novel for the patients.

Other studies focused on the brain structures involved in the emotional processing of music in patients with AD, in relation to the emotions caused by familiar music. Leggieri et al. (2018) carried out a listening program of familiar and non-familiar songs in a small group of patients with mild AD. They found, through functional magnetic resonance imaging, that family music activated more areas related to emotional processing and frontal areas, such as the cerebellum, inferior frontal gyrus, and putamen.

In a study by Jacobsen and collaborators (2015), the areas involved during the listening/encoding of familiar melodies were analyzed in comparison with recently heard melodies and new melodies, using functional magnetic resonance imaging. They found greater activation of non-atrophied areas in AD, during the listening of known melodies, such as the anterior cingular caudate and the presuplementary motor area. Other studies have suggested the importance of the familiarity of the melodies in the activation of limbic and paralimbic areas and the reward circuit (Pereira et al., 2011).

In relation to other skills related to music processing, some studies suggest that the ability to discriminate changes in the melody and, to a lesser extent, the rhythm, would not be preserved in AD. Campanelli and collaborators (2016), used, in this population, a battery for evaluation of amusia based on the model of modular processing of Peretz and Colheart (2003): Montreal Battery of Evaluation of Amusia (MBEA, Peretz, Champod, & Hyde, 2003). They evaluated patients with mild AD and healthy controls, and found a significant difference in the performance of both groups. The deterioration was greater for the processing of melodic aspects (contour, intervals and scales). Golden and collaborators (2017), found similar results, although they used a test developed by themselves. Their results indicate that, in AD, the processing of global aspects of the melody (contour changes) is impaired, but the processing of local aspects ("intrusive" notes or changes of intervals) is preserved.

Another study, however, had not found any alteration in the ability to discriminate changes in the melody, using the MBEA battery (Johnson et al., 2011).

There seems to be a dissociation between the processing of certain formal musical aspects (such as the global aspects of melody), and emotional recognition. In relation to the latter and the musical memory in AD (not addressed in this work), Peck, Girard, Russo, & Fiocco (2016) raise the question of whether music strengthens the connections of affected brain areas or, rather, points to processes that are preserved in AD. This question arises from the aforementioned work by Jacobsen and colleagues (2015). Both the anterior cingulate caudate and the proximal motor area, involved in listening to familiar musical extracts in healthy subjects, present less atrophy and hypometabolism in patients with AD. Ergo, suggesting that these regions are relatively preserved in this type of patients, despite the growing neuronal atrophy.

Conclusion and Discussion

Music acts as a powerful emotional stimulus, capable of producing emotional activation (arousal), a certain degree of pleasure or displeasure (valence); and to enhance the emotional effect of other stimuli (Baugmanter et al., 2006, Eschrich et al., 2008). In addition, it produces the activation of brain areas related to pleasure and reward, while decreasing the activation of areas related to fear and anger (Blood et al., 1999, Blood & Zatorre, 2001).

The processing of a musical stimulus, however, is not simple: it involves various specialized modules and components (Peretz & Colheart, 2003), and various areas of the brain (Koelsch & Siebel, 2005; Koelsch, 2011). This processing includes the emotion that a melody awakens or the emotional judgment that can be made of it. In addition, the importance of familiarity in the emotional response has been highlighted (Koelsch & Siebel, 2005; Pereira et al., 2011).

AD mainly affects episodic memory, but impairment of emotional recognition of stimuli, such as faces, has been reported (Rubinstein, Cossini, & Politis, 2016). Despite this, the emotional processing of melodies is (Drapeau et al., 2009, Gagnon et al., 2009, Samson et al., 2009). It could be considered that the emotional processing component (Peretz & Colheart, 2003), would be preserved in this pathology, and that, moreover, it would be specific for the processing of emotions in music.

The processing of formal musical aspects in AD, suggests that there would be an alteration, at least, in some of these aspects. Golden et al. (2017) affirm:

> Impaired global processing of melody information in music is in line with other defective formation of representations of other coherent global stimuli in AD: this deficit could reflect a greater demand for coordinated integrative computations between the temporo-parietal association cortices vulnerable to Alzheimer's disease (p. 12).

Koelsch & Siebel (2005) suggest that correct melodic processing requires, among other elements, the analysis of the outline of the melody and the relationship of the intervals. Here, at least one area of multimodal association stands out as responsible for the formation of the auditory Gestalt: the temporal plane seems to be a relevant structure for said analysis. Although there are still no studies that emphasize the importance of areas of parietal association in the cognitive processing of music, studies highlight the implication of temporal areas for an adequate processing and recognition of musical phrases. If these regions are commonly affected in the pathology of AD, it is not uncommon to expect a faulty melodic perception in this population (Campanelli et al., 2016).

At neuroanatomical level, there are few studies that show what happens in patients with AD, in relation to music and emotions. The few studies cited consider two points. First, the areas activated during listening are similar in patients and in healthy subjects, such as the cingulate gyrus, the cerebellum, and frontal areas (Blood & Zatorre, 2001, Jacobsen et al., 2015, Leggieri, 2018). These areas, in turn, are relatively conserved in the early stages of the disease. On the other hand, they emphasize the role of the known melodies, in contrast to the novel ones, in the emotional response: the emotional response is much greater when the melodies used are known, both in healthy subjects and in AD.

In conclusion, the proposed hypothesis has been confirmed: the emotional processing of music in AD is preserved, as well as the brain areas involved. Still remains to confirm a question raised by Peck and collaborators (2016) in relation to musical memory: does music strengthen the connections of affected brain areas or points to processes that are preserved in AD?

Focusing on the studies addressed in this work (emotions in music) the second option could be considered.

Finally, we emphasize the importance of studying emotions, focusing on what is preserved in AD. Many studies have used music to diminish the behavioral symptomatology in AD (aggression, anxiety, agitation, etc.), using family music (Cooke, Moyle, Shum, Harrison, & Murfield, 2010, Eggert et al., 2015, Särkämö et al., 2016). Likewise, the emotions that the music awakens for a person with ATD or awoke in a premorbid period, could be the key to explain the relative conservation of musical memory in this pathology, as opposed to the alteration in the episodic memory of other modalities (Cuddy & Duffin, 2005).

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub.

Bäckman, L., & Small, B. J. (1998). Influences of cognitive support on episodic remembering: Tracing the process of loss from normal aging to Alzheimer's disease. *Psychology and aging, 13*(2), 267

Bartoloni, L., Russo, M.J., Mesa y Morteo, M., Iturry, M., Román, F., Sevlever, G. (2015). In Roqué, M.L (Ed). *Deterioro Cognitivo, Alzheimer y otras Demencias: formación profesional para el equipo socio-sanitario* (pp. 28-41). Ciudad Autónoma de Buenos Aires. Argentina.

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: how music enhances the feeling of affective pictures. *Brain research*, *1075*(1), 151-164.

Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal Of Experimental Psychology: Human Perception And Performance*, *21*(6), 1278-1296.

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nature neuroscience*, *2*(4), 382.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, *98*(20), 11818-11823.

Bradley, M. M., & Lang, P. J. (2000). Emotion and motivation. *Handbook of psychophysiology*, *2*, 602-642.

Bressan, R. A., & Crippa, J. A. (2005). The role of dopamine in reward and pleasure behaviour–review of data from preclinical research. *Acta Psychiatrica Scandinavica*, *111*(s427), 14-21.

Cooke, M. L., Moyle, W., Shum, D. H., Harrison, S. D., & Murfield, J. E. (2010). A randomized controlled trial exploring the effect of music on agitated behaviours and anxiety in older people with dementia. *Aging and mental health*, *14*(8), 905-916.

Campanelli, A., Rendace, L., Parisi, F., D'antonio, F., Imbriano, L., Lena, C., & Trebbastoni, A. (2016). Musical cognition in Alzheimer's disease: application of the Montreal Battery of Evaluation of Amusia. Annals of the New York Academy of Sciences, 1375(1), 28-37.

Carlesino, G. A., & Oscar-Berman, M. (1992). Memory deficits in Alzheimer's patients: a comprehensive review. *Neuropsychology review, 3(2)*, 119-169.

Cummings, J. L., & Coffey, C. E. (1994). Neurobiological basis of behavior. In *Textbook of geriatric neuropsychiatry* (pp. 72-96). American Psychiatric Press, Inc, Washington, DC.

Cuddy, L. L., & Duffin, J. (2005). Music, memory, and Alzheimer's disease: is music recognition spared in dementia, and how can it be assessed? *Medical hypotheses*, *64*(2), 229-235.

Damasio, A. R. (1994). *El error de Descartes: la razón de las emociones*. Andrés Bello.

Drapeau, J., Gosselin, N., Gagnon, L., Peretz, I., & Lorrain, D. (2009). Emotional recognition from face, voice, and music in dementia of the Alzheimer type. *Annals of the New York Academy of Sciences*, *1169*(1), 342-345.

Eggert, J., Dye, C. J., Vincent, E., Parker, V., Daily, S. B., Pham, H., ... & Roy, T. (2015). Effects of viewing a preferred nature image and hearing preferred music on engagement, agitation, and mental status in persons with dementia. *SAGE open medicine*, *3*, 2050312115602579.

Eschrich, S., Münte, T. F., & Altenmüller, E. O. (2008). Unforgettable film music: the role of emotion in episodic long-term memory for music. *BMC neuroscience, 9(1)*, 48.

Fundación Alzheimer España (2016). Qué es la enfermedad de Alzheimer. Madrid, España:Fundación Alzheimer España. Recuperado de http://www.alzfae.org/alzheimer/95-que- es-alzheimer/que-es- alzheimer

Gagnon, L., Peretz, I., & Fülöp, T. (2009). Musical structural determinants of emotional judgments in dementia of the Alzheimer type. *Neuropsychology, 23(1)*, 90.

Golden, H. L., Clark, C. N., Nicholas, J. M., Cohen, M. H., Slattery, C. F., Paterson, R. W., ... & Warren, J. D. (2017). Music perception in dementia. *Journal of Alzheimer's disease*, *55*(3), 933-949.

Gómez-Gallego, M., & Gómez-García, J. (2017) Negative Bias in the Perception and Memory of Emotional Information in Alzheimer Disease. *Journal of geriatric psychiatry and neurology, 30*, 131-139.

Hsieh, S., Hornberger, M., Piguet, O., & Hodges, J. R. (2012). Brain correlates of musical and facial emotion recognition: evidence from the dementias. *Neuropsychologia, 50(8)*, 1814-1822.

Jacobsen, J. H., Stelzer, J., Fritz, T. H., Chételat, G., La Joie, R., & Turner, R. (2015). Why musical memory can be preserved in advanced Alzheimer's disease. *Brain*, *138*(8), 2438-2450.

Johnson, J. K., Chang, C. C., Brambati, S. M.,Migliaccio, R., Gorno-Tempini, M. L., Miller, B. L., & Janata, P. (2011). Music recognition in frontotemporal lobar degeneration and Alzheimer disease. Cognitive and behavioral neurology: official journal of the Society for Behavioral and Cognitive Neurology, 24(2), 74.

Kandel, E. R., Jessell, T. M., & Schwartz, J. H. (1997). *Neurociencia y conducta* (No. 577.25 KAN)

Kensinger, E. A., Brierley, B., Medford, N., Growdon, J. H., & Corkin, S. (2002). Effects of normal aging and Alzheimer's disease on emotional memory. *Emotion*, *2*(2), 118.

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in cognitive sciences*, *9*(12), 578-584.

Koelsch, S. (2011). Toward a neural basis of music perception–a review and updated model. *Frontiers in psychology*, *2*, 110.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1999). International affective picture system (IAPS): Instruction manual and affective ratings. *The center for research in psychophysiology, University of Florida*.

Ledoux, J. (1999). El cerebro emocional. Edit. Planeta, Barcelona.

Leggieri, M., Fornazzari, L., Thaut, M., Barfett, J., Munoz, D. G., Schweizer, T. A., & Fischer, C. (2018). Determining the Impact of Passive Music Exposure on Brain Activation and Functional Connectivity Using FMRI in Patients with Early Alzheimer's Disease. The American Journal of Geriatric Psychiatry, 26(3), S135.

Limb, C. J. (2006). Structural and functional neural correlates of music perception. *The anatomical record*, *288*(4), 435-446.

Li, X., Wang, H., Tian, Y., Zhou, S., Li, X., Wang, K., & Yu, Y. (2016). Impaired White Matter Connections of the Limbic System Networks Associated with Impaired Emotional Memory in Alzheimer's Disease. *Frontiers in aging neuroscience, 8*.

Omar, R., Hailstone, J. C., Warren, J. E., Crutch, S. J., & Warren, J. D. (2010). The cognitive organization of music knowledge: a clinical analysis. *Brain, 133(4)*, 1200-1213.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing Syntactic Relations in Language and Music: An Event-Related Potential Study. *Journal Of Cognitive Neuroscience*, *10*(6), 717-733.

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain: familiarity matters. *PloS one*, *6*(11), e27241.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences*, *999*(1), 58-75.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature neuroscience*, *6*(7), 688.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition, 68(2)*, 111-141.

Peck, K. J., Girard, T. A., Russo, F. A., & Fiocco, A. J. (2016). Music and memory in alzheimer's disease and the potential underlying mechanisms. *Journal of Alzheimer's Disease*, *51*(4), 949-959.

Robles, A., Del Ser, T., Alom, J., & Pena-Casanova, J. (2002). Propuesta de criterios para el diagnóstico clínico del deterioro cognitivo ligero, la demencia y la enfermedad de Alzheimer. *Neurología*, *17*(1), 17-32.

Rubinstein, W., Cossini, F., & Politis, D. (2016). Impact of facial emotional recognition alterations in Dementia of the Alzheimer type. Vertex XXVII, 263-269

Russo, M.J. (2015). Criterios diagnósticos. In Roqué, M.L (Ed). *Deterioro Cognitivo, Alzheimer y otras Demencias: formación profesional para el equipo socio-sanitario* (pp. 42-46). Ciudad Autónoma de Buenos Aires. Argentina.

Samson S., Dellacherie, D., & Platel, H. (2009). Emotional power of music in patients with memory disorders. *Annals of the New York Academy of Sciences, 1169(1)*, 245-255.

Särkämö, T., Laitinen, S., Numminen, A., Kurki, M., Johnson, J. K., & Rantanen, P. (2016). Pattern of emotional benefits induced by regular singing and music listening in dementia. *Journal of the American Geriatrics Society*, *64*(2), 439-440.

Satler, C., Garrido, L. M., Sarmiento, E. P., Leme, S., Conde, C., & Tomaz, C. (2007). Emotional arousal enhances declarative memory in patients with Alzheimer's disease. *Acta Neurologica Scandinavica, 116(6)*, 355-360.

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720-752.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Clinical Trials of the Encephalophone Music Prosthetic for the Motor Impaired

Thomas Deuel[1] James Rosenthal,[2] James Wenlock[3], Juan Pampin[4]
[1] *Department of Neurology, Swedish Medical Center, USA*

[1,4] *DXARTS and School of Music, University of Washington, USA*
[2] *Department of Electrical Engineering, University of Washington, USA*
[3] *Individualized Studies, University of Washington, USA*

[1]tasdeuel@gmail.com, [2]jamesdr@uw.edu, [3]wenloj@uw.edu, [4]pampin@uw.edu

Abstract

The Encephalophone is a brain-computer interface which uses elctroencephalogram (EEG) signal to generate music with intentional control, without movement. Mental imagery controls the pitch of the electronic instrument in real time. It's basic efficacy has been demonstrated with healthy subjects. However, with the goal of re-enabling the motor impaired to play music again - as a musical prosthetic - it's use with patients who have lost their musical ability due to ALS, brainstem stroke, spinal cord injury, or MS has not been shown.

To demonstrate real-world efficacy of the Encephalophone brain-music interface to enable patients with motor disability to generate music in real time without movement.

15 patients are undergoing three one hour-long sessions consisting of accuracy testing, free play, and then a second accuracy test. During accuracy testing, subjects are given a target note and must match the note 3 times in a row within 9.5 seconds. They are given 5 minutes to match as many notes as possible. During free play, patients are left to solo over accompanying music of various styles, tempos, and keys. A questionnaire is completed on the final session, to assess tolerability and satisfaction.

Seven patients with various levels of motor disability from a range of neurological disorders have completed the trial, out of a planned total of 15 patients. All patients demonstrated accuracy improvement over the 3 sessions, and 6 of 7 patients demonstrated accuracy significantly higher than chance.

Initial results are encouraging in suggesting that patients with severe motor disability due to ALS, brainstem stroke and other neurological disorders can be enabled to play music with reasonable accuracy without requiring movement.

Introduction

Since early in the history of the use of electroencephalogram (EEG) for measurement of electrical patterns of the human brain, efforts have been made to transform EEG electrical activity into sound.

The earliest example of converting EEG signal to sound appears in the literature shortly after the invention of the EEG. In 1934, Adrian and Matthews, replicating the earliest EEG descriptions of the posterior dominant rhythm ('the Berger rhythm') by Hans Berger in 1929 (Berger, 1929), monitored their own EEG with sound (Adrian and Matthews, 1934). Conversion of EEG signals to not just sound, but musical modalities, followed later: in 1965, the composer and experimental musician Alvin Lucier created a performance involving control of percussion instruments via strength of EEG posterior dominant rhythm (PDR), with the encouragement and participation of composer John Cage (Lucier, 1965). However, they experienced some difficulty in achieving good control, and to overcome this employed a second performer manually adjusting the gain from the EEG output (Rosenboom, 1975).

Eduardo Miranda at the Interdisciplinary Centre for Computer Music Research (ICCMR) at Plymouth University, UK was part of that summer workshop project, and has gone on to contribute significantly in this area of generating music from EEG signal. In 2008, he used the changing patterns of alpha and beta frequency rhythms in EEG to act as a switch between different musical direction to allow visual evoked potentials of EEG to control various musical parameters (Miranda et al., 2007). More recently, Miranda and colleagues used a statistical analysis of subjective emotions and EEG in an attempt to create an emotion sensor to subconsciously allow users to select music which they associate with more subjectively positive emotions (Eaton et al., 2007). Similarly, Scott Makeig and colleagues used EEG and non-EEG signal (scalp muscle and eye movement) from one subject to drive the use of subjective emotions to control a series of musical intervals (Makeig et al., 2001).

Some of the devices described above which use conscious control can be considered Brain Computer Interfaces (BCIs)(Wolpaw and Wolpaw, 2012). BCI research has progressed significantly in advancing towards the goal of using non-invasive EEG scalp electrodes to generate a direct interface from brain signal to a computer to control such actions as moving a cursor on a screen or a word speller (Sellers et al., 2014), driven by signals such as alpha frequency event-related desynchronizations and synchronizations (Roberts et al., 1999)(Pfurtscheller et al., 2000). We sought to use the well-described methods of using posterior dominant rhythm and motor imagery EEG real-time control to create a new scalar musical instrument, and to measure its accuracy for novices.

We created the Encephalophone, a musical instrument and biofeedback device that uses motor cortex mu rhythm (mu) to consciously and volitionally control the generation of scalar music. Mu rhythm was used to generate an instrument that can be controlled without movement (Yuan and He, 2014) for potential applications with patients with motor disabilities. Alpha frequency control using mu rhythms has been well described in the BCI literature (Roberts et al.,1999)(Pfurtscheller et al., 2000) for non-musical control. We previously demonstrated accuracy for novice, healthy individuals that is significantly higher than random in controlling the instrument by conscious cognitive processes. Here we also describe recent results in progress of a clinical trial to demonstrate feasibility of re-enabling hands-free music creating with patients with moderate to severe motor disability from stroke, ALS, MS, and spinal cord injury.

Materials and Methods

EEG Signal Collection

Figure 1 illustrates the experimental setup (Fig. 1). A Mitsar 201 EEG (Mitsar Co., Ltd., St. Petersburg, Russia; distributed by Nova Tech, Inc., Mesa, AZ) and 19-channel ElectroCap electrode cap (Electro-Cap International Inc., USA) were used to collect EEG signal utilizing the International 10-20 system of electrode placement (American Electroencephalographic Society, 1994) from 15 healthy human volunteer subjects.

Subjects were positioned in a relaxed, reclining position with a headrest to minimize muscle artifacts, and were positioned facing away from computer screens and other equipment to eliminate any potential for visual feedback. EEG signal at a sampling rate of 500Hz was initially processed in a HP Pavilion PC (Hewlett-Packard, Palo Alto, CA) with Mitsar EEG Acquisition software, where filters were applied (100Hz low-pass, 0.5Hz high-pass, and 60Hz notch filters). Raw EEG signal was visually verified by a physician clinical neurophysiologist for good signal quality and lack of artifacts. EEG data was then streamed in real time to Matlab (The MathWorks, Inc., Natick, MA) via the Mitsar Matlab API.

Matlab scripts for real-time signal processing were created to apply a fourth order Butterworth filter at the 8-12 Hz band to generate an estimate of power for the posterior dominant rhythm in visual cortex from occipital electrode O1, or motor cortex mu rhythm from electrode C3 (international 10-20 system) for right hand motor imagery, in real time. The delay in the system from EEG signal acquisition to Matlab processing was approximately 20 msec. The filter was applied to incoming segments of 500 ms of data. The bandpass filtered data was rectified and then averaged over the entire segment length to produce a single power estimate for every segment.

Calibration Period

A calibration was created for each individual subject and each individual trial session of the Encephalophone. The five minute long calibration period consisted of twenty 15 second long alternating cued states ("on" or "off"). For visual cortex posterior dominant rhythm, an auditory cue of "on" cued the eyes closed, awake state, and "off" cued the eyes open, awake state. For motor cortex mu rhythm, an auditory cue of "on" cued the awake, resting state, and "off" cued the motor imagery (but not actual movement) state: subjects were instructed to imagine right hand grasping and opening at a rate of approximately 1 Hz as per prior motor imagery BCI methods of Neuper *et al.* (Neuper et al., 2006). This calibration period established the range of values of 8-12 Hz power for an individual subject and individual trial session in the different cued states, then divided these values into eight equal sized 'bins', or ranges of values, based on the calibration period alpha power histogram. After calibration, these 8 possible values generate the 8 scale degrees of the C major musical scale including the octave (C4 to C5).

After the calibration period is used to calibrate the instrument to each individual, the device enters the free-running period, during which a value from 1 to 8 is generated every 500 msec in real-time from the desired 8-12 Hz frequency power (posterior dominant rhythm or mu rhythm) of the user. Subjects were allowed brief (3 minute) free-running practice with note generation before accuracy experiments.

This free-running stream of values from 1 to 8 in Matlab is sent at a rate of one value per 500 msec (120bpm musical tempo for quarter notes) using OSC (Open Sound Control) along an Ethernet cable via a router to a second computer - an Apple MacBook Pro (Apple, Inc. USA) - where it is received by Max/MSP music generation software (Cycling '74, USA). The streaming values from 1 to 8 are used to generate the 8 scale degree notes in the C major musical scale with a synthesized piano tone (8 notes from C4 to C5).

Accuracy Experiments

For note accuracy experiments, the subject is presented with a target note of either a high C (C5) or low C (C4). The subject generates one note every 500 msec and attempts to match the note (C4 or C5) or its nearest neighbor (D4 or B4) 3 times consecutively. If the note is successfully matched 3 times consecutively, a 'hit' is scored and a reward chord (C major) is played, then a new target note is presented. If the subject does not hit the target note 3 times consecutively within 9.5 seconds (19 notes), a 'miss' is scored and an error chord (tritone) is played, then a new target note is presented. This results in a chance probability of 19.03% to score a 'hit' over the interval. A total of 300 seconds, or 5 minutes, is given for each trial, and the results recorded.

Results

15 healthy adult volunteer subjects were trained and tested for musical accuracy using the Encephalophone using both posterior dominant rhythm (PDR) control and motor mu rhythm control (basic subject demographics shown in Table 1). Subjects underwent a 5 minute calibration period, followed by a brief (3 minute) free-run practice period, then a 5 minute accuracy trial for each of PDR and mu control. Results from these musical accuracy experiments were recorded for individual number of hits, trials and percent accuracy for each 5 minute trial using PDR control and mu control (summary shown in Table 2).

Subjects using PDR control had an average of 27.4 hits (standard deviation =11.9, standard error +/-3.2) in an average of 38.7 trials, resulting in an average of 67.1% accuracy (Fig. 2a, standard deviation = 17.42%, standard error +/-4.5%). Subjects using mu control had an average of 20.6 hits (standard deviation = 5.7, standard error +/-1.5) in an average of 35.6 trials, resulting in an average of 57.1% accuracy (Fig. 2b, standard deviation = 11.2%, standard error +/-3.0%). Each individual subject scored significantly higher than random in accuracy for both PDR and mu control (Fig. 2): p values ranged from 6.3 x 10^{-36} to 2.8 x 10^{-3}. Additionally, PDR accuracies (average 67.1%) were significantly higher (p = 1.4 x 10^{-4}) than Mu accuracies (average 57.1%).

We also looked at the correlation between PDR hits and accuracy, and mu hits and accuracy, with years of musical training (Figure not shown). There was a moderate positive relationship between increased PDR hits and accuracy (correlation values 0.58 and 0.41, respectively) - but not mu hits and accuracy (correlation values -0.16 and -0.11, respectively) - with increasing years of musical training.

Initial results of clinical trials (first seven of 15 clinical subjects) show all seven demonstrating improved accuracy over three one-hour sessions with two accuracy tests per session (Figure 3). All but one subject demonstrate accuracy significantly higher than chance probability.

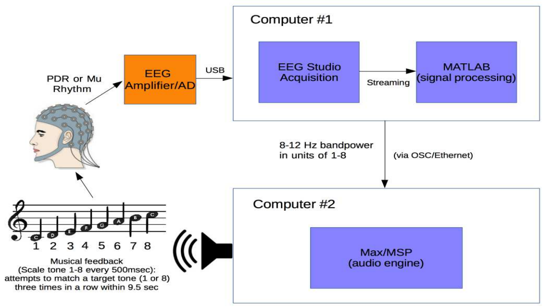


Figure 1. Experimental Setup. EEG signal from subject wearing electrode cap is sent from Mitsar 201 EEG amplifier to Computer #1 where 8-12 Hz PDR or Mu power is converted to a value from 1 to 8. This value from 1-8 is sent via OSC to Computer #2 where it is converted to a musical piano tone in the key of C (seven tones of C major scale and octave, from C4 to C5). Subjects generating tones attempt to match them with a presented target tone.

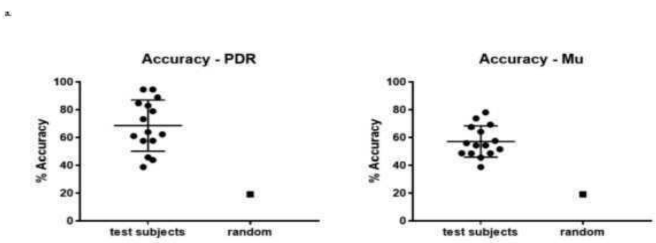


Figure 2. Percent Accuracy for PDR and Mu Control. Individual subjects were given 5 minutes to hit as many target notes as possible using either PDR control or Mu control. Scatter plots of results of all subjects were generated (bars represent mean and standard deviation), with random (chance) control, for each of: (A) Percent accuracy using PDR control (standard error +/- 4.7). (B) Percent accuracy using mu control (standard error +/- 3.0).

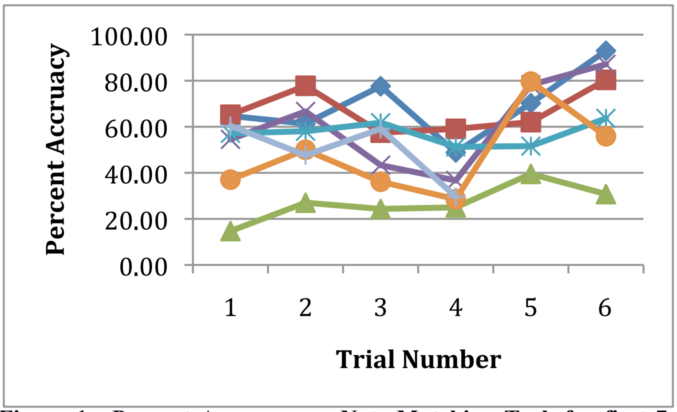


Figure 1. Percent Accuracy on Note-Matching Task for first 7 clinical trial subjects. Trial 1-2 on session 1, trial 3-4 on session 2, trial 5-6 on final session (3).

We initially created the Encephalophone, a musical instrument and biofeedback device, which uses mu rhythm EEG signal to control notes of a musical scale in real time. Prior experiments with 15 normal subjects novice to the device demonstrated accuracy in hitting a target note, with each subject scoring significantly higher than random (Deuel et al., 2017). There was additionally a moderate positive correlation between years of musical training and PDR accuracy, but not mu accuracy.

These studies demonstrated that the Encephalophone allows novices to have some cognitive volitional control of generation of musical notes in real time, without movement. Given the known potential for significant improvement with training in mu-based BCI devices (Neuper et al., 2006), novices such as those tested here have the potential with continued training to significantly improve accuracy and facility with the instrument. The use of scalar musical tones - rather than non-musical sound or visual biofeedback - may confer a training advantage: the benefits of music for arousal motivation for both training and therapeutics have been shown (Bergstrom et al., 2014). Third, the use of a musical feedback-based EEG device with responsiveness noticeable to the user holds promise for patients - such as those with ked-in syndrome - who are severely incapacitated and may more likely to respond to auditory (and specifically musical) nulus and feedback than to visual stimulus and feedback. is is particularly so for those who may have visual pairment (e.g. cortical blindness), and particularly to those o played music before their injury.

We are currently undergoing clinical trials of the cephalophone with 15 patients with moderate to severe motor disability from stroke, ALS, MS, or spinal cord injury. Results thus far show ability to significantly control note accuracy in six of seven patients, and all show improvement of accuracy with training.

Future controlled clinical trials will look at therapeutic and anatomic effects of the Encephalophone on patients with motor disability, in terms of new white matter tract formation (e.g. DTI tractology) as well as measurement of cognitive and motor improvements with use of the Encephalophone in serial training sessions. It has already been demonstrated that neurologic music therapy improves executive function in traumatic brain injury rehabilitation (Thaut et al., 2009). Other patients who might thus benefit would include patients suffering from amyotrophic lateral sclerosis (ALS), brainstem stroke, or traumatic amputation (such as war veterans). The ability to generate music using a portion of the brain that is no longer able to control motor movement of limbs may be beneficial for emotional and cognitive rehabilitation. Also, combining the Encephalophone with physical therapy may improve motor rehabilitation, while cortical 'rewiring' of motor circuits by the recruitment of the motor cortex by the Encephalophone may allow new motor output pathways for regaining some motor control.

Previously, others have reported use of BCI to control not only visual output (e.g. cursor on a computer screen) but also sound, and reported better control and accuracy with visual rather than (non-musical) auditory feedback (Nijboer et al., 2008). However, Bergstrom and colleagues showed musical biofeedback to be better than either simple passive music listening or non-musical sonification biofeedback for control of physiological arousal state (Bergstrom et al., 2014). Our previous results show control with virtually no training, using scalar musical tone feedback rather than non-musical auditory feedback. Thus we hope that with further training involving musical accompaniment between testing sessions, the musical context provided will greatly improve learning and accuracy of control. This is being tested in experiments underway with highly trained musical subjects and serial training, with graded testing with more note target options (3 or 4 possible notes to match rather than 2).

This device is being used as a novel improvisational musical instrument in live performance, accompanied by small ensembles of musicians. Future development will include using multiple soloists performing with Encephalophones together, in a call and response improvisation, as well as performers improvising not only with musical scales, but also with timbre or chordal improvisation. Furthermore, work in computer music using conscious control of sound spatialization is being explored.

References

Adrian, E.D., Matthews, B.H. (1934). The interpretation of potential waves in the cortex. *J Physiol* 81(4), 440-71.

American Electroencephalographic Society (1994). Guideline thirteen: Guidelines for standard electrode position nomenclature. *J Clin Neurophysiol*, 11, 111-113.

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen (On the human electroencephalogram). *Archiv f. Psychiatrie u. Nervenkrankheiten* 87, 527–70.

Bergstrom I, Seinfeld S, Arroyo-Palacios J, Slater M, Sanchez-Vives MV. (2014). Using music as a signal for biofeedback. *Int J Psychophysiol.* 93, 140-149.

Breslow, N., Clayton D. (1993). Appoximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88 (421): 9-25

Brouse A., Filatriau J., Gaitanis K., Lehembre R., Macq B, Miranda E.R., et al. (2006) Use of spectral analysis of multiple frequency bands of EEG brain waves to generate sound. An instrument of sound and visual creation driven by biological signals;
eNTERFACE summer workshop. http://www.enterface.net

Deuel T.A., Pampin J., Sundstrom J., Darvas F. (2017) The Encephalophone: A Novel Musical Biofeedback Device using Conscious Control of Electroencephalogram (EEG). *Frontiers in Human Neuroscience* (11): 213

Eaton, J. Williams, D. Miranda, E. (2011). Affective Jukebox: A Confirmatory Study of EEG Emotional Correlates in Response to Musical Stimuli. Proceedings of Joint ICMC. 11th Sound and Music Conference, 14th International Computer Music Conference, University of Athens, Greece

Hinterberger T, Baier G. (2005). Parametric Orchestral Sonification of EEG in Real Time. *IEEE MultiMedia*. 12(2), 70-79.

Loui P., Koplin-Green M., Frick M., Massone M. (2014). Rapidly Learned Identification of Epileptic Seizures from Sonified EEG. *Frontiers in Human Neuroscience.* 8:820. doi:10.3389/fnhum.2014.00820.

Lucier, A. Music for Solo Performer; 1965. Performance. Available: https://www.youtube.com/watch?v=bIPU2ynqy2Y

Makeig, S., Leslie, G., Mullen, T., Sarma, D., Bigdely-Shamlo, N. and Kothe, C. (2001). First demonstration of a musical emotion BCI. *Affective Computing and Intelligent Interaction.* Berlin: Springer pp. 487-496.

Miranda E.R. and V Soucaret V. Mix-It-Yourself with a Brain-Computer Music Interface; 2008. Proceedings of 7th International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT) & ArtAbilitation, Porto, Portugal

Miranda ER, Magee WL, Wilson JJ, Eaton J, Palaniappan R (2011). Brain-Computer Music Interfacing (BCMI): From Basic Research to the Real World of Special Needs. *Music and Medicine* vol. 3 no. 3, 134-140.

Neuper C., Müller-Putz G.R., Scherer R., Pfurtscheller G. (2006). Motor imagery and EEG-based control of spelling devices and neuroprostheses. *Prog Brain Res*. 159, 393-409.

Neuper C., Scherer R., Reiner M., Pfurtscheller G. (2005). Imagery of motor actions: differential effects of kinesthetic and visual-motor mode of imagery in single-trial EEG. *Brain Res Cogn Brain Res* 25, 668-677.

Nijboer F., Furdea A., Gunst I., Mellinger J., McFarland D.J., Birbaumer N., et al. (2008). An auditory brain-computer interface (BCI). *J Neurosci Methods.* 167(1), 43-50.

Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W., Ramoser, H., Schlögl, A., et al. (2000). Current Trends in Graz Brain-Computer Interface (BCI) Research. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society* 8, 216–9.

Pham M., Hinterberger T., Neumann N., Kübler A., Hofmayer N., Grether A., et al. (2005). An auditory brain-computer interface based on the self-regulation of slow cortical potentials. *Neural Repair.* 19(3), 206-18.

Roberts, S. J.; Penny, W.; Rezek, I. (1999). Temporal and Spatial Complexity Measures for Electroencephalogram Based Brain-Computer Interfacing. *Medical & Biological Engineering & Computing 37*, 93–98.

Rosenboom, D. (1975). In *Biofeedback and the arts: results of early experiments*, ed. D.
Rosenboom (Vancouver, Canada: Aesthetic Research Centre of Canada.) 60-61.

Sellers E., Rayn D., Hauser C. (2014). Noninvasive brain-computer interface enables communication after brainstem stroke. *Sci Transl Med*. October 8; 6(257): 257re7.

Thaut MH1, Gardiner JC, Holmberg D, Horwitz J, Kent L, Andrews G, Donelan B, McIntosh GR. (2009). *Ann N Y Acad Sci.* 1169, 406-16.

Wolpaw, JR, Wolpaw, EW (2012). Brain-computer interfaces: principles and practice. Oxford ; New York: Oxford University Press.

Yuan, H., He, B. (2014). Brain-Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives. *IEEE Trans Biomed Eng.* 61(5), 1425–1435.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Individual differences in Emotional Responses to Music

Laura Edelman, Michael Silverstein, Amanda Hawthorne, Erik Berger, and Christine Cimpian.
Psychology Department, Muhlenberg College, Allentown, PA, USA

lauraedelman@muhlenberg.edu, michaelsilverstein@muhlenberg.edu, amandahawthorne@muhlenberg.edu, erikberger@muhlenberg.edu,christinecimpian@muhlenberg.edu

Abstract

Two motivation systems are known to influence one's perception of the world, the Behavioral inhibition system (BIS) and Behavioral activation system (BAS) (Carver & White, 1994). BIS can be characterized as an internal drive to avoid aversive stimuli whereas BAS motivation can be characterized as an internal desire to seek fun, rewards, and more appetitive stimuli. Loxton, Mitchell, Dingle, and Sharman (2016) found that individuals high on BAS had stronger positive responses to music. Other individual difference factors that may influence responses to music include dispositional optimism and differences in need to belong. Several studies indicate that people respond emotionally more to the melody of songs than to the lyrics (Peynircioglu, Rabinovitz, and Thompson, 2008; Ali and Peynircioglu, 2006). The message of the lyrics is not always consistent with the emotional tone of the melody. The current studies used music that had either congruent or incongruent melodies and lyrics to assess whether these individual differences factors would influence to which aspect of the music participants would respond. All three factors (BIS/BAS, optimism, and need to belong) influenced emotional responses to the music. The emotional responses were primarily to the melody rather than the lyrics.

Introduction

There has been extensive research on individual differences that influence emotional responses to music (Barrett & Janata, 2016; Karreman, Laceulle, Hanser & Vingerhoets, 2017; Ladinig & Schellenberg, 2012; and Vuoskoski & Eerola, 2011). Among these differences are the two motivation systems known to influence one's perception of the world, the Behavioral inhibition system (BIS) and Behavioral activation system (BAS) (Carver & White, 1994). BIS can be characterized as an internal drive to avoid aversive stimuli whereas BAS motivation can be characterized as an internal desire to seek fun, rewards, and more appetitive stimuli. Balconi, Falbo & Conte (2012) showed that people higher on BAS responded more strongly to positive stimuli and people higher on BIS responded more strongly to negative stimuli. Several researchers have explored how BIS and BAS might affect music listening. Loxton, Mitchell, Dingle and Sharman (2016) found that people higher on BAS were more involved in music, and Mori and Iwanaga (2015) found that people who were higher on BAS were more likely to experience chills when listening to music. Ravaja and Kallinen (2004) used background music and showed that both BIS and BAS influenced how participants responded to startling music.

Other individual differences that have been found to play a role in responses to music are optimism and need to belong. Noguchi, Gohm, and Dalsky (2006) showed that optimistic people attend more to positive rather than negative information in a story. We might expect that the same would be true for music, that optimistic people would attend more to the positive than negative attributes of a piece of music. Past research does not show a clear optimistic bias in music perception. Some studies have found that listening to music increases people's feeling of optimism (Elvers, 2016 and Getz, Marks & Roy, 2014). Vella and Mills (2017) found that highly optimistic people tended to prefer listening to energetic and rhythmic music. Additionally, Getz, Marks and Roy 2014) showed that optimistic people have stronger emotional responses to music. The current study explores whether optimistic people attend more to the positive attributes of music.

Baumeister and Leary (1995) argue that humans have a fundamental need to belong to a group. If group membership is a fundamental need, then finding ways to become a member of a group would be a basic human drive. It has been theorized that music has evolved as a tool that humans use to encourage social bonding, thus music would play a role in fulfilling this basic human need (Dunbar, 2012). Hagen and Bryant (2003) suggest that music can enhance group cohesion. Enhanced cohesion would make group membership more secure. Loersch and Arbuckle (2013) explored the role of need to belong in music perception. They found that individuals high on need to belong and individuals whose sense of belongingness was threatened were more sensitive and more reactive to music.

The individual differences that we chose to study should have different effects for the happy versus sad information in the music. To explore the different responses, we chose to use music with mixed and congruent emotional information. Some of the music has melody and lyrics that convey the same emotion, and some of the music has melody and lyrics that convey contrasting emotions. Several studies indicate that people respond more emotionally to the melody of songs than to the lyrics (Peynircioglu, Rabinovitz, and Thompson, 2008; Ali and Peynircioglu, 2006). The message of the lyrics is not always consistent with the emotional tone of the melody. We hypothesized that people who are more BIS oriented will be more sensitive to the negative emotions and therefore respond more strongly to either the music or lyrics depending on which depicts the more negative emotion. Conversely, people who are more BAS oriented will respond more to whichever aspect of the music carries positive emotion. People high in optimism should respond more strongly to the positive information, and people high in need to belong may respond more strongly in general.

General Methods

Participants

One-hundred twenty-four people (67.5% women, 31.7% men) participated in Experiment 1. One hundred and five people (87% female 13% male) participated in Experiment 2.

Design

Both studies used a mixed 2X2X2X2 factorial design to examine the influence of the individual difference factors on music perception. The independent variables were the emotional tone of the melody (happy or sad), the emotional tone of the lyrics (happy or sad), whether the melody and lyrics were consistent or inconsistent, and whether the participant was high or low on the individual difference factor.

Stimuli for both Experiments

Brief, ten second clips, of twelve songs were presented to the participants. There were three songs in each of the four categories based on tone and congruency. Each song was pretested to fit within the category. The songs were:

Happy melody, sad lyrics
1. Third Eye Blind, Semi Charmed Life
2. Goo Goo Dolls, Slide
3. Nena, 99 Red Balloons

Sad melody, happy lyrics
1. Joe Purdy, I love the Rain the Most
2. Augustana, Boston
3. John Legend, All of Me

Happy melody, happy lyrics
1. Frankie Valli and the 4 Seasons, December 1963
2. Katrina & The Waves, Walking on Sunshine
3. Billy Idol, Dancing With Myself

Sad melody, sad lyrics
1. Adele, Someone Like You
2. Sarah McLachlan, In the Arms of an Angel
3. Taylor Swift, All Too Well

Materials and Procedure

For experiment 1, participants were sorted into high or low based on responses to the BIS/BAS scales (Carver & White, 1994.) These scales include 20 items: seven items measure Behavioral Inhibition, four items measure Drive, four items measure Fun Seeking and five items measure Reward Responsiveness. Each item was rated on a four-point scale from not true to very true of me.

For Experiment 2, participants were sorted into high or low in dispositional optimism according to their responses on Scheier, Carver, and Bridges's (1994) Life Orientation Test. The scale consists of 12 statements, including, "It's easy for me to relax" and "I enjoy my friends a lot" (participants divided by median split). Participants then answered these questions using a 1-5 scale, with 1 meaning strongly agree and 5 meaning strongly disagree. Participants were also sorted into high or low in belongingness according to their responses on Leary, Kelly, Cottrell, and Schreindorfer's (2013) Need to Belong Scale. The scale consists of 10 statements including, "I want other people to accept me" and "I do not like being alone" (participants divided by median split). Participants then answered these questions using a 1-5 scale, with 1 meaning strongly agree and 5 meaning strongly disagree. Participants rated each piece of music on the degree to which they found it to be happy, exciting, and familiar, as well as the degree to which they liked the music. These attributes were rated on a 1 to 7 scale with higher numbers meaning more.

Participants took the surveys online through Qualtrics survey software.

Results

For each individual difference scale, participants were divided into two categories (high or low) based on a median split of the data. A four-way mixed analysis of variance was used to assess the effects of melody, lyrics, BIS and BAS on the rating of how happy the music was. There was a significant interaction between the music and the lyrics, F (1,119) = 108.02, $p < .01$, $\eta^2 = .48$. Figure 1 shows that the melody had a stronger effect on the rating, but the lyrics also influenced the rating.

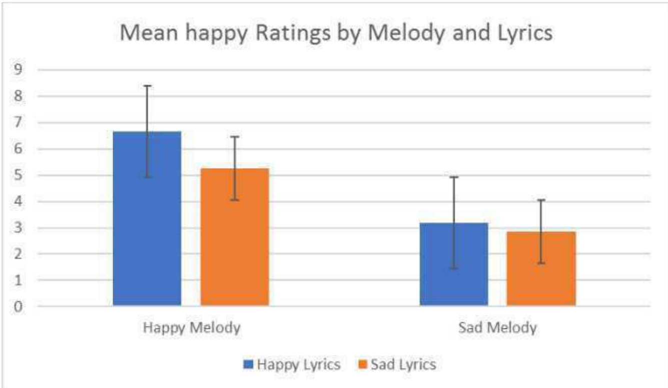


***Figure 1*. The mean rating of happy by melody and lyrics.**

BIS/BAS Results

For the happy ratings, there was a main effect of BAS, such that people high on BAS rated all the music as happier (M = 4.64, SD = 1.47) than people low on BAS, M = 4.35, SD = 1.20, F (1,119) = 4.19, p = .04, $\eta^2 = .03$. There was a significant interaction between melody and BIS, F (1,119) = 13.48, $p < .01$, $\eta^2 = .10$. Figure 2 shows that people high on BIS made a larger distinction in happiness ratings than people low on BIS.

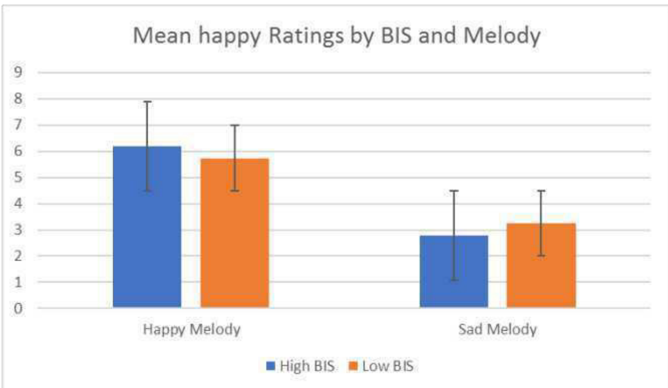


Figure 2. **The mean rating of happy by BIS and lyrics.**

A similar analysis of variance was done on the liking ratings. People high on BIS liked all of the music more (M = 5.78, SD = 1.40) than people low on BIS, M = 5.25, SD = 1.42, F (1,119) = 7.92, $p < .01$, $\eta^2 = .06$. Similarly, people high on BAS liked all of the music more (M = 5.87, SD = 1.41) than people low on BAS, M = 5.16, SD = 1.39, F (1,119) = 13.87, $p < .01$, $\eta^2 = .10$.

The same analysis was done on the ratings of how exciting the music was. People high on BAS rated all of the music more exciting (M = 5.86, SD = 1.69) than people low on BAS, M = 5.15, SD = 1.86, F (1,119) = 8.07, $p < .01$, η^2 = .06. Similar interactions were found between melody and BAS (F (1,119) = 7.14, $p < .01$, $\eta^2 = .06$) and melody by BIS, F (1,119) = 10.71, $p < .01$, $\eta^2 = .08$. Figure 3 shows the means for this interaction for BIS, the numbers were almost identical for BAS. For both BAS and BIS, people high on the scale made a greater distinction in exciting for the happy and sad melody than people low on the scale.

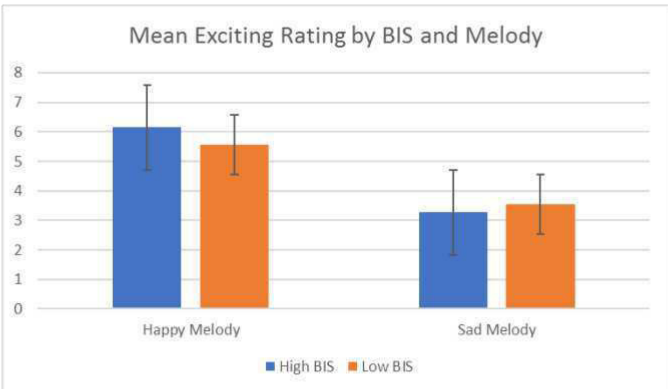


Figure 3. **The mean rating of exciting by BIS and melody.**

Finally, similar analyses were done on the familiarity ratings. Participants seemed to be equally familiar (M = 5.67, SD = 1.40) with all the music and there were no significant effects.

Optimism and Need to Belong Results

The same set of analyses were done for optimism and need to belong for melody and lyrics on the four different response measures. The happy rating resulted in only one interaction between melody, lyrics, and optimism, F (1,101) = 4.85, p = .01, $\eta^2 = .05$. Figure 4 shows that the highly optimistic people tended to rate all the music as happier, except the music with both a sad melody and sad lyrics. Need to belong did not influence the happy ratings.

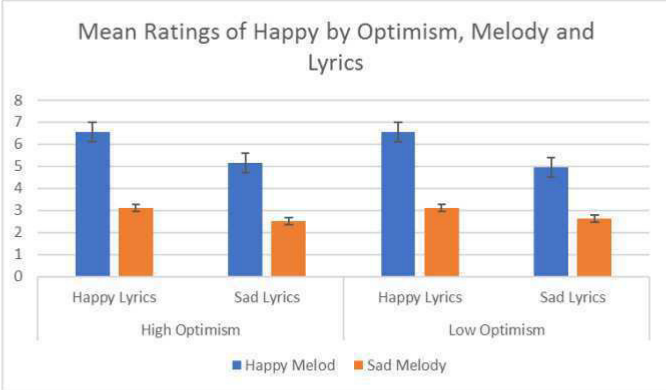


Figure 4. **The mean rating of happy by optimism, melody and lyrics**.

For the liking ratings, there was a main effect of optimism, F (1,101) = 6.65, p < =. 01, $\eta^2 = .06$. People high on optimism liked all of the music more (M = 5.71, SD = 1.16) than people low on optimism, M = 5.24, SD = 1.39. Need to belong did not influence the liking ratings.

For the ratings of exciting, there was an interaction between need to belong and melody, F (1,101) = 7.27, p < . 01, $\eta^2 = .07$. Figure 5 shows that when the melody was happy people with a high need to belong found it more exciting than those low in need to belong. However, when the melody was sad people low in need to belong found it more exciting than people high in need to belong.

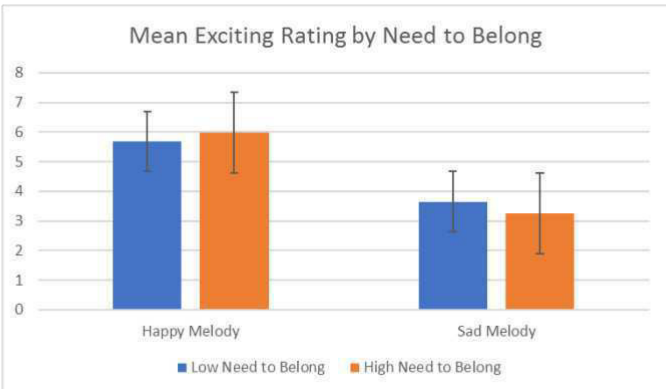


Figure 5. **The mean rating of exciting by need to belong and melody.**

Finally, similar analyses were done on the familiarity ratings. Participants seemed to be equally familiar (M = 5.71, SD = 1.37) with all the music and there were no significant effects.

Discussion

Regardless of the individual difference factors, the participants responded more to the melody than the lyrics of the music. The effect size for the music and lyrics interaction was 48%, indicating that, similar to the results of Peynircioglu, Rabinovitz, and Thompson (2008) and Ali and Peynircioglu (2006), participants responded more strongly to the melody than the lyrics. However, the lyrics did influence the ratings of the music. Happy lyrics made the sad melody less sad and sad lyrics made the happy melody less happy. Our hypothesis that participants high in BIS or low on optimism might respond more to the sad aspects of the songs was not supported. When the individual difference factors did show an interaction with the music, it was always with the melody and not with the lyrics. Again indicating that people respond more to the melody than the lyrics. Participants high on BIS made a greater distinction between the happy and sad melodies than did people low on BIS. People high on BIS rated the happy melodies as both happier and more exciting and the sad melodies as both sadder and less exciting than the people low on BIS. It seems as though people high on BIS, who are expected to monitor the environment for negative stimuli, have stronger emotional reactions or perceive greater intensity of the emotional characteristics of stimuli. However, there is no evidence that the sad lyrics drew their attention away from the happy melody.

The main effects fit better with our expectations of these individual differences. People high on BAS rated the music happier, more exciting, and liked the music more than those low on BAS. Similarly, people high in optimism rated the music as happier and liked it more than those low in optimism. Thus, people expected to attend to positive stimuli in the environment perceived the music more positively, even when the melody or lyrics, or both were sad. It may be that these characteristics (BAS and optimism) influence people to interpret their environment more positively than people low on these characteristics. Need to belong only influenced excitement ratings in that people high in need to belong found the happy music more exciting. Given the strength of Loersch and Arbuckle's (2013) findings, it is surprising that need to belong did not have a stronger effect. However, need to belong may require a context to trigger attention to that need. We expected that filling out the need to belong scales would prime participants to be thinking about that need. Filling out the scales might not have been a sufficient prime and there was no threat to belongingness in the survey.

Overall, the results of our study show that the musical factors (melody and lyrics) have a much stronger effect on emotional responses to music than do individual differences. A limitation of this study is that we only used 10 second clips of the music. Although 10 seconds is enough time for participants to identify the music, they must be relying on memory to know that the lyrics are happy or sad. Future studies should use unfamiliar music and perhaps longer segments of music. Using unfamiliar music would force participants to base their ratings on the music they are currently hearing as opposed to their memory of a familiar song.

References

Ali, S. O., & Peynircioğlu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners?. *Psychology Of Music*, *34*(4), 511-534. doi:10.1177/0305735606067168

Balconi, M., Falbo, L., & Conte, V. A. (2012). BIS and BAS correlates with psychophysiological and cortical response systems during aversive and appetitive emotional stimuli processing. *Motivation And Emotion*, *36*(2), 218-231. doi:10.1007/s11031-011-9244-7

Barrett, F. S., & Janata, P. (2016). Neural responses to nostalgia-evoking music modeled by elements of dynamic musical structure and individual differences in affective traits. *Neuropsychologia*, *91*234-246. doi:10.1016/j.neuropsychologia.2016.08.012

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497-529. doi:10.1037/0033-2909.117.3.497

Carver, C. S., & White, T. L. (1994). BIS/BAS Scales. *Psyctests*, doi:10.1037/t00751-000

Dunbar, R. I. M. (2012).On the evolutionary function of song and dance. In N.Bannan (Ed.), Music, language and human evolution (pp. 201-214). Oxford, England: Oxford University Press.

Elvers, P. (2016). Songs for the ego: Theorizing musical self-e nhancement. *Frontiers In Psychology*, *7*

Getz, L. M., Marks, S., & Roy, M. (2014). The influence of stress, optimism, and music training on music uses and preferences. *Psychology Of Music*, *42*(1), 71-85. doi:10.1177/0305735612456727

Gupta, U., & Gupta, B. S. (2015). Psychophysiological reactions to music in male coronary patients and healthy controls. *Psychology Of Music*, *43*(5), 736-755. doi:10.1177/0305735614536754

Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human Nature*, *14*(1), 21-51. doi:10.1007/s12110-003-1015-z

Karreman, A., Laceulle, O. M., Hanser, W. E., & Vingerhoets, A. M. (2017). Effects of emotion regulation strategies on music-elicited emotions: An experimental study explaining individual differences. *Personality And Individual Differences*, *114*36-41. doi:10.1016/j.paid.2017.03.059

Ladinig, O., & Schellenberg, E. G. (2012). Liking unfamiliar music: Effects of felt emotion and individual differences. *Psychology Of Aesthetics, Creativity, And The Arts*, *6*(2), 146-154. doi:10.1037/a0024671

Leary, Mark R., Kelly, Kristine M., Cottrell, Catherine A., & Schreindorfer, Lisa S. (2013). Construct validity of the Need To Belong Scale: Mapping the nomological network. Journal of Personality Assessment, Vol 95(6), 610-624. doi: 10.1080/00223891.2013.819511

Loersch, C., & Arbuckle, N. L. (2013). Unraveling the mystery of music: Music as an evolved group process. *Journal Of Personality And Social Psychology*, *105*(5), 777-798. doi:10.1037/a0033691

Loxton, N. J., Mitchell, R., Dingle, G. A., & Sharman, L. S. (2016). How to tame your BAS: Reward sensitivity and music involvement. *Personality And Individual Differences*, *97*35-39. doi:10.1016/j.paid.2016.03.018

Mori K, Iwanaga M. General reward sensitivity predicts intensity of music-evoked chills. *Music Perception* [serial online]. June 2015;32(5):484-492.

Noguchi, K., Gohm, C. L., & Dalsky, D. J. (2006). Cognitive tendencies of focusing on positive and negative information. *Journal Of Research In Personality*, *40*(6), 891-910. doi:10.1016/j.jrp.2005.09.008

Ravaja, N., & Kallinen, K. (2004). Emotional effects of startling background music during reading news reports: The moderating influence of dispositional BIS and BAS sensitivities. *Scandinavian Journal Of Psychology*, *45*(3), 231-238. doi:10.1111/j.1467-9450.2004.00399.x

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Revised Life Orientation Test. *Psyctests*, doi:10.1037/t09267-000

Saito, Y., Sakuma, N., Ishii, K., & Mizusawa, H. (2009). The role of lyrics and melody in song recognition: Why is song recognition faster?. *Japanese Journal Of Psychology*, *80*(5), 405-413. doi:10.4992/jjpsy.80.405

Vella, E. J., & Mills, G. (2017). Personality, uses of music, and music preference: The influence of openness to experience and extraversion. *Psychology Of Music*, *45*(3), 338-354. doi:10.1177/0305735616658957

Vuoskoski, J. K., & Eerola, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, *15*(2), 159-173. doi:10.1177/1029864911403367

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

A large scale study on the participants of the "Jugend musiziert" music competition: Starting points and questions

Heiner Gembris[1] and Claudia Bullerjahn[2]

[1]*Institute for Research on Musical Ability (IBFM), Paderborn University, Germany*
[2]*Institute for Musicology and Music Education, Justus-Liebig-University Giessen, Germany*

[1]heiner.gembris@uni-paderborn.de, [2]claudia.bullerjahn@musik.uni-giessen.de

Abstract

This study surveyed highly gifted young people who participated at the national level in the annual "Jugend musiziert" (youth making music) contest in Germany. Some previous large-scale studies from the 1990ies have dealt with the participants of the "Jugend musiziert" contest, their socio-cultural background, motivation, their experiences with music etc. Since then, only very few studies have been published. The aim of the present survey is to gain a deeper up-to-date insight into the socio-cultural contexts, motivation, interests, music preferences, personalities, leisure activities etc. of the participants at the national level of the contest "Jugend musiziert". Furthermore, aspects like health and wellbeing, preferences, stage-fright are included as well. The evaluation of data is not finished yet, the present paper focuses on general information about the social background and the familiar atmosphere of the young musicians. A 16-pages standardized paper-pencil questionnaire was administered to ca. 2,260 participants at the national level "Jugend musiziert" contest in 2017. A number of 1,143 valid questionnaires was returned. The young musicians (aged 9 to 24; M = 15.1; SD = 2.14) came from families with clear above-average academic education, with parents strongly interested in music. They had an above-average number of brothers and sisters, who mostly also played instruments. The vast majority of parents had no music-related occupation. The parents and the family as "persons in the shadow" provided a rich music-oriented sociotope, which values and intensively supports music activities, so that children are enabled to unfold their musical giftedness to excellence. These preliminary findings suggest that future research should study the role and contributions of the "persons in the shadow" for a better understanding not only of the development of musical excellence, but also of possible negative effects that can occur through too much pressure in the promotion of particularly gifted children and adolescents.

Background

The music competition "Jugend musiziert" (youth making music) is the largest and most important music contest for young musicians in Germany since more than 50 years. The main aim of "Jugend musiziert" is to encourage music activities and the advancement of talents in music. This annual competition is organized into three levels: the regional level, the federal state level, and the national level. Winners of the regional level can participate in the federal state level, winners of the federal state level can enter the national level. In 2017, approx. 20,529 young people attended the regional level, 8,300 attended the federal state level, and a number of ca. 2,732 finally participated in the national level, among them 90 students from German schools abroad (Deutscher Musikrat, 2017, pp. 70; 31). The participants on the national level competition are commonly regarded as highly gifted young musicians, and the winners of the national level competition surely belong to the national elite of young musicians. In 2017, the "Jugend musiziert" competition was announced for the following instruments / categories: piano, harp, voice, drum set (pop), guitar (pop), string ensemble, wind ensemble, chamber music for accordions, and "Neue Musik" (New Music; i.e. avant-garde music of the 20th century / contemporary music) (Deutscher Musikrat, 2017, p. 1)

Despite the "Jugend musiziert" contest exists since more than 50 years, research on this contest and its participants is scarce. About 30 years ago, Bastian (1987; 1989; 1991) carried out his comprehensive research on the participants of "Jugend musiziert" at the federal state level and national level, on their socio-cultural background, motivation, their experiences with music etc. Some years later, Linzenkirchner, & Eger-Harsch (1995) published their research on participants and instrumental teachers, who had been involved in the "Jugend musiziert" competition. In different data collections they focused especially on participants who attended the regional level and on a comparison between participants of the regional, the federal state, and the national level. Since then, only very little research has been published. Mund (2007) compared successful participants of the national competitions "Jugend musiziert" and "Jugend forscht" (youth researching) concerning cognitive characteristics and personality traits. Bullerjahn, Hantschel, & Hirchenhein (2017) recently studied the motivation of participants at the regional level of the "Jugend musiziert" competition. The present paper deals with the participants at the national level contest, carried out from 1st to the 8th of June 2017 in Paderborn, Germany.

Aims

The general aim is to gain a deeper insight into the socio-cultural contexts, motivation, interests, musical preferences, personalities, leisure activities etc. of participants of the "Jugend musiziert" competition at the national level. Furthermore, aspects like health and wellbeing, playing related disorders, stage-fright etc. are included as well. In the present paper, we focus on general information about the social background and the familiar atmosphere of the young musicians.

Method

We developed a standardized paper-pencil questionnaire (including some open questions), which covers a broad range of aspects, e.g. instrument, personal experiences with the "Jugend musiziert" contest, socio-cultural variables, personality, musical training and practice, motivation, musical preferences, support by the family, etc. On the one hand, the questionnaire contains questions, which allow direct comparisons with earlier research carried out 30 years ago. On the other hand, we used questions related to up to date issues like the use of social media and the connections between making music, health, and wellbeing. In some parts we integrated items from existing standardized instruments, e.g. to measure the motivation for participation in the contest (Bullerjahn, Hantschel, & Hirchenhein, 2017), playing-related disorders (Gembris, & Ebinger, 2017), stage fright (Nusseck, Zander, & Spahn, 2015), the use of media (Krupp-Schleußner, 2016), and the big-five-personality inventory (Rammstedt et al., 2013). The complete questionnaire enfolded 16 pages.

Some weeks before the contest started, all participants at the national level and their parents received a letter which informed about the research project. We contacted the participants personally when they enrolled at the central registration desk for the national contest in Paderborn. A number of ca. 2,260 participants agreed to participate in the survey. They were asked to fill in the questionnaire immediately, if possible, or to return the completed questionnaire in the following days into special boxes available at every of the 20 places where the contest was carried out. The evaluation of data is ongoing and not yet completed. Quantitative data are evaluated with SPSS 25, for the evaluation of the qualitative data (open questions) we are using the method of qualitative content analysis (Kuckartz, 2014). As a first step of data evaluation, we will focus on quantitative descriptions in the following sections.

Results

A number of 1,143 valid questionnaires has been returned (rate of return = ca. 50%).

The age of the young musicians ranged from 9 to 24 years (M = 15.1; SD = 2.14), 62% (n = 692) were female, 38% male (n = 425). Most of them (69%; n = 775) competed in the ensemble contests (strings, winds, accordion, “Neue Musik”), the smaller part (27%) participated in the solo contests (piano, harp, voice, drum set, guitar), 2% competed in both, the ensemble contest and the solo contest, 2% were accompanists.

The vast majority of the participants (95%) was born in Germany as their mothers (82%) and fathers (85%), too.

Residential area and school

Since larger cities usually provide more music schools, instrumental teachers and a richer musical life than small cities, it is surprising that 41% of the young musicians indicated to live in a small city or in the countryside. A share of 35% lived in a middle-sized city, and only 23% in a major city. Most of the participants (85%) attended a grammar school (Gymnasium). Although it is possible that spending much time with playing an instrument may detract attention and time necessary for school, the school achievements of the young musicians are rather good: the grade point average was 1.8 (SD = 0.55;) with a little, but significant advantage for the girls (M = 1.8, SD = 0.63) compared to the boys (M = 1.9; SD = 0.69; T-test, p = .001; d = 0.22). (The best possible grade is 1, the poorest 6.)

Educational background and occupation of the parents

The occupations of the parents were asked with open questions. A total number of 1,069 of the adolescents indicated their father’s job; a little less (N = 1,043) specified the profession of the mother. When indicating the occupations by the adolescents, it should be noted that not all adolescents may know or be able to specify the exact job title of their parents. For a first description of the general level of parental education, the professions of parents had been categorized according to occupations requiring academic studies, and occupations requiring no academic studies. More than a half of the fathers (57%) worked in academic professions, including 10.9% musicians. The smaller share of 20.4% had non-academic professions. In 21.5% of the cases, the specified professions could not be clearly assigned to one of the two categories. Only a very small part (n = 10 or 0.9%) of the fathers had not been working. For a further differentiation of the parent’s occupation, we classified the mentioned professions into several main categories. The major category (31.9%) comprises occupations in the areas of administration, services and commerce (e.g., civil servants, craftsmen, merchants, insurance employees). The second largest group (19.6%) were scientific professions (chemist, physicist, computer scientist, engineer). In 15.1% of cases, music-related occupations were specified (musicians, music educators, musicologists). Other more frequently mentioned professions were doctors (7.9%) and teachers (6.2%).

Of the mothers, 52% had academic professions, including 16.1% musicians. The relatively largest group (25.8%) of the mother’s occupations included jobs in administration, services, and commerce. Music-related vocations of the mother were mentioned by 19.2% of the adolescents. With 11.5%, the teachers followed in the third place. Furthermore, 10.2% mentioned non-academic health professionals (naturopaths, nurses), female doctors (7.6%) and social professions (educators, day care; 6.0%). A relatively small proportion of 5.3% said their mothers were not working at the time the survey took place. There were no details of the extent of employment (full-time, part-time). In summary, more than half of the fathers and mothers of the participants in the national competition have received an academic training. The share of academic professions of the contestant’s parents in this sample is almost three times as high as the population average, which is around 20% in the age group from 45 to 65 years (Autorengruppe Bildungsberichterstattung, 2016, p. 44)

Parental home and musical activities in the family

In order to get information about the socioeconomic background of the parental home, questions about the housing situation and home ownership were used as more indirect indicators. We assumed that the direct question of the parents’ income could hardly have been reliably answered by the adolescents. In addition to the fact that home ownership tends to indicate a higher income than a rental home, home ownership (or a rented house) generally offers better opportunities for practice than rented accommodation.

The majority of our adolescents lived in an owner-occupied house (61%), an owner-occupied flat (14%) or a rented house (5%). Only a relatively small proportion (16%) lived in a rented apartment.

These results correspond substantially to the results found by Bastian (1991). 81% of the “Jugend musiziert” participants he surveyed lived in their own homes, 13% in rented apartments and 5% in rented houses (Bastian, 1991, p. 59).

Brothers and sisters

The vast majority (86%) of our sample comes from families with multiple children. Only 14% are single children. Most common are one (42%) or two siblings (28%). From families with four children comes another 11%. A small part (6%) lived in homes with five or more children.

Coincidentally, the data from Bastian (1991) and the data from our study show that most of the participants in the national competition come from families with several children, with families with three, four and more children being strongly represented. In both studies children are more likely to stem from large families with three, four or more children than from families with only one child.

Compared to the federal statistics (Federal average 2016; Statistisches Bundesamt, 2017), the large families are much more common in the "Jugend musiziert" participants. The number of families with four and more children at the national level "Jugend musiziert" participants is more than twice as high compared with the national average (17% vs. 8%). The number of families with three children (28%) is about ten percent points higher than the national average (19%). At the same time, there were about half as many single-child families (14%) among the "Jugend musiziert" participants in 2017 as in the national average (26%). In Bastian's (1991) study, the number of single-child families on the national level competition was only one third of the national average. There is also evidence from other data that families engaged in music have a greater number of children compared to the national average (Bayerischer Rundfunk, 2014, p. 86; SOMM – Society Of Music Merchants e.V., 2013, p. 37)

Significance of music in the family

If there are several siblings in the family, most of them also play an instrument. Of the interviewed "Jugend musiziert" participants with siblings, 85% state that at least one sibling plays an instrument or sings. 64% of the mothers play an instrument and / or sing in a choir. Of the fathers, there are 51% who sing in a choir or play an instrument. In the wider family environment, 40% of grandmothers or grandfathers were musically active (instrumental playing and / or choir) and 38% of the uncles and aunts.

The importance of music in the family was recorded using six different items, each to be responded on a five-point rating scale. For reasons of simplification, the consenting items are summarized as approval and the rejecting items as rejection. The vast majority of respondents (93%) said that parents were interested in music, and 82% said their family heard a lot of music. The statement "My parents wanted me to learn an instrument" was approved by 82%. Two-thirds (65%) said that their family spoke a lot about music. These statements suggest that music is an important topic in these families. One third of the families also make music together. 36% agreed to the item "In our family, we often play together". Half (51%) of the adolescents did not find these statements applicable. Due to the question formulation, it is not clear whether music is made with parents and / or siblings. Since 85% of the interviewees said that their siblings also make music, while only 64% (mother) or (51% father) parents play an instrument and / or sing in chorus, it is conceivable that the siblings are often the family members with which music is made. The fact that for half (51%) of the respondents frequent music making with family was not appropriate, does not exclude the possibility of occasionally making music together. Similar is the case regarding singing. The item "In my family we often sing together" found 28% approval, 57% disagreed.

Evidence for the high value of music in the family and the related involvement of the parents in supporting the contestants comes from the following results: Virtually all (99%) of respondents agreed to the item "My parents are ready to invest money and time for my music lessons". Almost as many (94%) said that the family came to auditions and concerts to listen. Almost one half (47%) of them agreed with the item "My parents urge me to practice / have asked me to practice". On the other hand, 40% disagreed.

A considerable proportion (44%) stated that parents assist in practicing when there are problems, but for an almost equal proportion (45%) this is not the case. Furthermore, a third (33%) said that parents control whether or not to practice regularly. The last three findings show that in addition to the time commitment and financial effort required to support the participants in the competition, a considerable proportion of the parents also called for, supported and partially controlled the practice. However, these findings are age-dependent. Adolescents whose parents help with practice, encourage practice, and control practice are significantly younger (14.38 to 14.61 years; *SD* 2.01 to 2.14) than those adolescents, whose parents do not (15.55 to 15.78; *SD* 2.10 to 2.18; *t*-Tests, p =.000; Eta^2 = .047 to .073). But since the effect of age is relatively week, there may be considerable inter-individual differences.

Conclusion

In order to fulfill the very high performance requirements for a participation in the national competition "Jugend musiziert", clearly musical giftedness and early, long-term support from the parents are required. With Ziegler, & Stöger (2016), one can regard the home environment as a sociotope, providing objective and normative spaces of action in which children's musical giftedness can unfold. In the case of the national competition participants, most familiar sociotopes are characterized by a far above-average academic education of the parents and material-economic conditions that enable them to live in their own homes.

Although up to a fifth of the parents have a musical education, the vast majority of parents does not carry on any musical profession. Nevertheless, the music has a high, normative value in the family. More than half of the parents sing in a choir or play an instrument. Music listening and talking about music belongs to the home environment. About a third of the adolescents reports that in their homes music is played together with parents and siblings. With the high rate of academics and pronounced musical activities and interests, the parents' homes have a high "educational capital" that provides favorable conditions for the development of excellence (see Lehmann, & Kristensen, 2014, p. 60). The parents take a strong part in the musical development of the adolescents. Almost all parents accompany the young people to auditions and concerts. Especially for the younger ones, practicing is often demanded and controlled by the parents. These domestic environmental conditions, the educational capital of the parents, and their commitment form an infrastructural and thematically favorable learning-sociotope, which promotes the development of musical giftedness. It is also noteworthy that these families, compared with the population average, have more often two, three or more children, whereby the siblings usually also make music. This requires a high personal, time-consuming engagement of the parents, which cannot be provided without concessions to their own needs, wishes, etc. The significant contribution of these "persons in the shadow" (Gruber, Lehtinen, Palonen, & Degner, 2008) in the promotion of talents and careers has been little considered by research (Lehmann, & Kristensen, 2014, p. 59). For a better understanding of the contribution of "persons in the shadow" to the development of musical excellence and careers as well as of the negative effects that may occur through too much pressure in the promotion of particularly gifted children and adolescents (Heye, in prep.), more research is required. Our results strongly confirm results from studies carried out 30 years ago, showing that the participants on the national level of the "Jugend musiziert" contest came predominantly from families with a clearly above average educational and socio-economic background (Bastian, 1989; 1991). It should be examined, why highly gifted young musicians from other socio-economic classes are absent and what can be done to find and to support those children.

Acknowledgements. We would like to thank the participants of the Bundeswettbewerb "Jugend musiziert" for their participation in the study and for completing the extensive questionnaire. Furthermore our thanks include the students of both the Paderborn University and the Justus-Liebig-University Giessen who helped us with the distribution of the questionnaires and by motivating the participants of the "Jugend musiziert" music contest to take part on our survey.

References

Autorengruppe Bildungsberichterstattung (Ed.) (2016). *Bildung in Deutschland 2016. Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration.* Bielefeld: Bertelsmann.

Bastian, H. G. (1987). *Jugend musiziert. Der Wettbewerb aus der Sicht von Teilnehmern und Verantwortlichen.* Mainz: Schott.

Bastian, H. G. (1989). *Leben für Musik: eine Biographie-Studie über musikalische (Hoch-) Begabungen*. Mainz: Schott.

Bastian, H. G. (1991). *Jugend am Instrument: Eine Repräsentativstudie*. Mainz: Schott.

Bayerischer Rundfunk (Ed.) (2014). *Symphonieorchester des Bayerischen Rundfunks 2014 – 2015*. München: Bayerischer Rundfunk.

Bullerjahn, C., Hantschel, F., & Hirchenhein, T. (2017). Anreize für die Teilnahme am mittelhessischen Regionalwettbewerb „Jugend musiziert" – eine Fragebogenstudie. In W. Auhagen, C. Bullerjahn, & C. Louven (Eds.), *Musikpsychologie – Akustik und musikalische Hörwahrnehmung* (pp. 153-186). Göttingen: Hogrefe.

Deutscher Musikrat (2017). *54. Bundeswettbewerb Jugend musiziert: 1. bis 8. Juni 2017 in Paderborn.* Bonn: Deutscher Musikrat.

Gembris, H., & Ebinger, F. (2017). Musizierbedingte körperliche Beschwerden bei Kindern und Jugendlichen. In H. Gembris (Ed.)*, Musik studieren und Abitur machen. Doppelbelastung im gesundheitlich-gesellschaftlichen Kontext* (pp. 129-158). Münster: LIT.

Gruber, H., Lehtinen, E., Palonen, T., & Degner, S. (2008). Persons in the shadow: Assessing the social context of high ability. *Psychology Science Quarterly, 50*(2), 237-258.

Heye, A. (in prep.). *Mehrfachbelastung in der Ausbildung musikalisch besonders begabter Jugendlicher*. Münster: LIT.

Krupp-Schleußner, V. (2016). *Jedem Kind ein Instrument? Teilhabe an Musikkultur vor dem Hintergrund des capability approach*. Münster: Waxmann.

Kuckartz, U. (2014). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (2nd rev. ed.). Weinheim: Beltz Juventa.

Lehmann, A., & Kristensen, F. (2014). "Persons in the Shadow" Brought to Light: Parents, Teachers, and Mentors – How Guidance Works in the Acquisition of Musical Skills. *Talent Development & Excellence, 6*(1), 57-70.

Linzenkirchner, P., & Eger-Harsch, G. (1995). *Gute Noten mit kritischen Anmerkungen: Wirkungsanalyse der Wettbewerbe „Jugend musiziert" 1984 bis 1993*. Bonn: Deutscher Musikrat.

Mund, W. (2007). *Jugend forscht und Jugend musiziert: Kognitive Fähigkeiten und Persönlichkeitsmerkmale erfolgreicher Teilnehmer.* Dissertation, Philipps-Universität Marburg. Retrieved from http://archiv.ub.uni-marburg.de/diss/z2008/0750/pdf/dwm.pdf

Nusseck, M., Zander, M., & Spahn, C. (2015). Music Performance Anxiety in Young Musicians: Comparison of Playing Classical or Popular Music. *Medical Problems of Performing Artists, 30*(1), 30-37.

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A Short Scale for Assessing the Big Five Dimensions of Personality: 10 Item Big Five Inventory (BFI-10). *methoden, daten, analysen*, *7*(2), 233-249.

SOMM – Society Of Music Merchants e.V. (Ed.) (2013). *Jahrbuch der Musikinstrumenten- und Musikequipmentbranche 2011/2012.* Berlin: SOMM – Society Of Music Merchants e.V.

Statistisches Bundesamt (Ed.) (2017). *Statistisches Jahrbuch. Deutschland und Internationales 2017.* Retrieved from https://www.destatis.de/DE/Publikationen/StatistischesJahrbuch/StatistischesJahrbuch2017.pdf;jsessionid=FBEE47E3B57DD5DDD3802D4F4F7E3E04.InternetLive2?__blob=publicationFile

Ziegler, A., & Stöger, H. (2016). Der Soziotop-Ansatz in der Begabungsförderung: Theoretischer Hintergrund und Anwendungsperspektiven. *journal für begabtenförderung, 16*(2), 4-14.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Psychometric Features to Assess Absolute Pitch: Looking for Construct Validity Evidences Regarding Isolated Pitch Tasks in Undergraduate Brazilian Music Students

Nayana Di Giuseppe Germano,[1] Hugo Cogo-Moreira,[2] and Graziela Bortz[3]

[1, 3] *Department of Music, UNESP, Sao Paulo Brazil*

[2] *Department of Psychiatry, Federal University of Sao Paulo, Sao Paulo Brazil*

[1]nayanager@hotmail.com, [2]hugocogobr@gmail.com, [3]g_bortz@hotmail.com

Abstract

A bibliographical review of absolute pitch (AP) displayed that its main characteristic lies in the ability of identifying tones using verbal labels without any type of external reference (Germano, 2015). However, among its different definitions, we also found several non-consensual criteria describing such latent phenomenon, leading to different and non-directly comparable models: e.g., AP as a trait (continuous measure) versus AP as a categorical phenomenon (categorized in groups). For the study of any latent psychological measurement, it is essential to identify a set of observable indicators. Such criteria must have content validity based on evidence from empirical observation and theoretical foundation. Subsequently, it is crucial to test the conceptual model (formed by tasks based on underlying latent entities) fit to real data. Under a dimensional approach, besides the goodness of fit, it is possible to evaluate tasks parameters such as discrimination, difficulty, and probability of guessing. The aim of this research is to provide evidences regarding psychometric features (i.e., model goodness of fit and item parameters) for a set containing ten isolated pitch dichotomous tasks (correct versus incorrect). A total of 783 Brazilian undergraduate music students from 7 universities took part in the test. The data were analyzed based on two approaches: dimensional factor analysis, also known as item response theory (IRT) and categorical latent class analysis (LCA). IRT results evidence a good adjustment in assessing the ability to label isolated pitches without reference. For LCA, two latent classes showed the best class solution: the first with high probability of correct recognition and the second with low probability. Nevertheless, the dimensional solution fits better. Our results also showed empirically that the prevailing procedure adopted on AP research where it is credited the same score to the correct recognition for all pitches (i.e., ordinary summing of correct answers) might not be reliable due to considerable differences in items' difficulty and discrimination.

Introduction

The Absolute Pitch (AP) phenomenon is a cognitive ability that has been widely researched during the last century. Although the earliest scientific description of AP appeared in a volume on psycho-acoustics by Stumpf (1883), previous records have already been describing the ability since the era of Mozart. As AP ability is usually considered a mystery to be unveiled, it has intrigued musicians, psychologists and neuroscientists over the years. Due to this fact, since Stumpf, various aspects of AP have been investigated by various researchers (e.g., Abraham, 1901; Mull, 1925; Wellek, 1938; Takeuchi & Hulse, 1993; Neu 1947; Ward, 1999; Miyazaki & Ogawa, 2006; Germano, 2015).

Although there is a substantial number of AP researches published, after conducting an extensive bibliographical review, we have faced an issue when trying to design a construct for AP's trait (Germano, 2015):

- Different definitions for the same ability;
- Different classificatory tests for the same ability;
- Different score for right responses and/or semitones errors responses;
- Different cut-off points to be considered an AP possessor.

The AP trait is not equally defined amongst the academic community. The existence of several definitions for the same cognitive phenomenon inevitably leads to significant variations among AP researches and, more importantly, in their findings. The major problem we found was not the general perspective of the AP phenomenon, but significant non-consensual criteria which are almost never considered in AP definition.

The main characteristic of this cognitive ability, an agreement among researchers, is the capacity that AP possessors have of tone identification using verbal labels (what Levitin, 1994, named as Pitch Labeling) without reference (e.g., without a diapason). According to this reasoning, the AP phenomenon is usually defined as a rare ability that refers to a long-term internal representation for pitches. It is typically manifested behaviourally by the ability to identify, by the name of the musical note, the pitch of any sound without reference to another sound or by producing a given musical tone on demand, without external reference (Baggaley, 1974; Zatorre et. al., 1998; Ward, 1999; Parncutt & Levitin, 2001; Deutsch, 2002). However, the definition adopted by several researchers excludes many important non-consensual criteria, such as the time demanded to identify a tone and the degree of precision in tone identification (questioning directly the established common sense that a subject with problems on tone identification due to certain musical parameters, like register and timbre, is sometimes considered not to be an AP possessor). Many authors have been researching these non-consensual criteria, but their findings are almost never discussed within the pre-definition context, i.e., defending modifications on the base AP definition.

Main AP non-consensual criteria are:

- On which timbres is an AP possessor able to identify pitches?
- On which registers is an AP possessor able to recognize pitches?
- If a subject cannot sing a demanded pitch without external reference, should he be considered an AP possessor?
- How many semitone errors an AP possessor usually commits?
- What percentage of correct answers must a subject achieve to be considered an AP possessor?

If one adopts a very restrict definition (e.g., AP possessors must identify precisely every note on every instrument and every register), there would be only a few subjects successfully classified as AP possessors. On the other hand, a broader definition (e.g., if AP possessors can take as much time as needed to identify a tone or if they are able to identify it in only one instrument) can make it very difficult, or even impossible, to distinguish AP possessors from non-possessors, as a wide range of answers are considered indicatives of the cognitive ability.

After further consideration, we decided to propose a new research approach that measures musical perception and AP using Psychometrics and the *MPlus* software (version 8.0). Our main objective is to measure music perception and AP in the same way that medical researches measure diseases such as depression and dementia. The resulting theoretical model could then be tested using procedures from a field of statistics known as structural equation modelling (SEM), focused specifically on testing theoretical models. In order to do so, we considered AP a latent trait (represented by a cycle) and the stimuli as indicators (represented by rectangles) (see figure 1). The crucial questions involved are: What are the criteria necessary to define AP? Considering a designed set of criteria, are they capable of accurately explaining the latent psychological trait? From a scientific perspective, it is important that the designed criteria can be tested to provide evidences supporting or falsifying the core theoretical model adopted, allowing its enhancement (Germano et al., 2016).

Model evaluation

The identification of a set of structured, consistent, observable criteria based on evidence is essential to study any latent psychological phenomenon (i.e., one that cannot be measured directly, such as AP).

Most AP researches adopt experimental testing as their core methodology to measure different patterns of subject response in relation to parameter variation (such as register, timbre, the time necessary for pitch identification or the proportion of correct/wrong answers). However, as highlighted in the previous section, the lack of consideration regarding correspondences between theoretical model and empirical data makes the process of knowledge acquisition extremely difficult.

If it is unknown how well the proposed set of criteria for the definition of the AP phenomenon fits to the reality of the AP possessor and its abilities, so solving this matter should be the first phase in experimental research. In fact, the borderline that separates AP possessors from non-possessors has not been consensually defined, considering that both groups exhibit limitations on pitch identification. Considering that many researches adopt as a starting point of their methodology the segregation of AP possessors from non-possessors, this problem is shown to be of great importance.

As presented in Germano et al. (2016), a possible solution is the creation of a reproducible model for AP categorization using standardization of criteria and the construct validity evaluation as conducted in the medical area, throughout guidelines stated in sources such as *Diagnostic and Statistical Manual of Mental Disorders* or *Composite International Diagnostic Interview*. The resulting theoretical model could then be tested by using procedures from a field of statistics known as structural equation modelling, focused specifically on testing theoretical models.

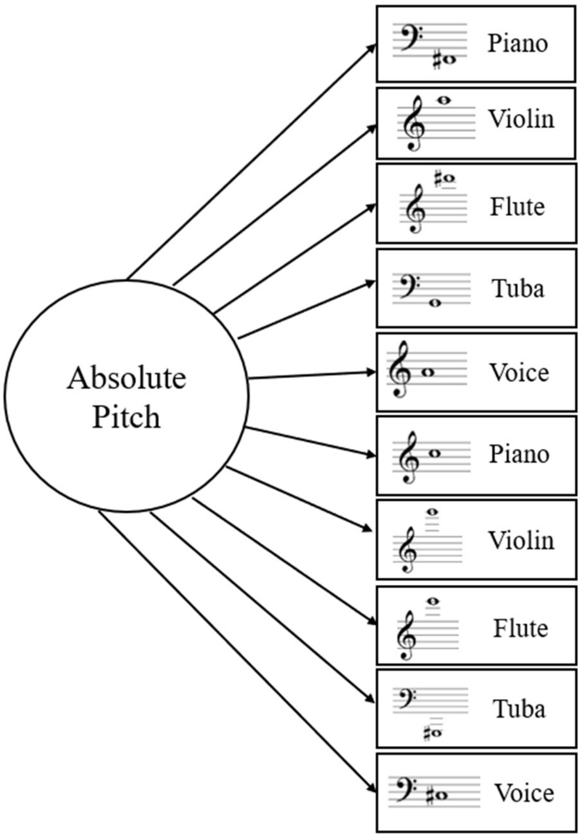


Figure 1. Theoretical Model for Absolute Pitch trait. The circle represents the latent trait (that cannot be measured directly) and the rectangles represent the items (that can be tested and measured).

Material and method

This test is part of a set of tests conducted to analyze AP and RP. At this point, we will analyze only the first battery of tests (isolated pitch for AP) and, in future research, we are going to analyze other parts of the test, conducted especially for the RP ability, such as intervals and triads.

A total of 783 undergraduate music students (n=512 male; 65.4%) from first to tenth semester of study at 7 different Brazilian universities, 5 located in São Paulo city and 2 in Curitiba city (Ethics Committee's Approval CAAE: 60855816.3.0000.5477) took part on this research. The participants' mean age was 24.7 (range= 17 to 72) and they had an average of 10.29 years of music practice (SD=, ranging from 1 to 65 years). All subjects self-reported their auditory music perception, i.e., if they consider themselves as AP possessors, Relative Pitch (RP) possessors, both or none. Subjects also answered questions about their music experience (time of music study, main instrument and starting age on musical training).

The test consisted on a battery of 10 isolated pitches in 5 different timbres recorded from real instruments (see figure 1). Each stimulus was played once, with 3-second duration and 15-second pause in between. No reference pitch was given. The subjects were instructed to mark the pitch they though as correct in a piano-keyboard drawn on a paper. The octave parameter was not considered in this task. Our focus was only in pitch discrimination and, because of that, there were only one octave drawn in the piano-keyboard response sheet.

For the analysis presented on this paper, only exact pitch responses were considered as correct answers (different criteria and items' responses will be considered on future publications). The data were analyzed based on two approaches: dimensional factor analysis, a.k.a. item response theory (IRT) and categorical approach latent class analysis (LCA).

Results

IRT results regarding Root Mean Square Error of Approximation (RMSEA), 90 Percent Confidence Interval (CI), Comparative Fit Index (CFI), Tucker Lewis Index (TLI) and Weighted Root Mean Square Residual (WRMR) evidenced a good adjustment (RMSEA=0.033, 90% CI=0.020-0.045, CFI=0.973, TLI=0.965, WRMR=1.129). The items' discrimination and difficulty were very high, high and moderate:

Table 1. IRT Parameterization: Item Discrimination and Difficulty. Isolated Pitches.

	Discrimination	Difficulty
Isolated Pich a	1.927	2.050
Isolated Pich b	1.209	0.448
Isolated Pich c	1.391	1.937
Isolated Pich d	1.250	0.949
Isolated Pich e	1.129	0.277
Isolated Pich f	1.850	0.240
Isolated Pich g	1.929	0.767
Isolated Pich h	1.830	0.724
Isolated Pich i	1.303	2.419
Isolated Pich j	1.356	2.141

As can be seen, the most discriminatory item is Isolated Pitch *g* (G on violin) followed by item *a* (F# on piano), *f* (C on piano) and *h* (E on flute), as they present very high discrimination value. The remaining items have high and moderate discrimination. The item with the highest difficulty value is Isolated Pitch *i* (G# on tuba), followed by item *j* (C# on voice) and item *a* (F# on piano).

The result of two latent classes showed the best class solution (entropy 0.914), being that the most entropic item was Isolated Pitch *g* (entropy 0.586). Comparing the Bayesian Information Criterion (BIC), Aika Information Criterion (AIC) and Sample-size Adjusted Bayesian Information Criterion (SSABIC) for dimensional solution (BIC=7082.150, SSABIC=7018.640, AIC=6988.887) with those for categorical solution (BIC=7105.199, SSABIC=7038.513, AIC=7007,273), we have evidence that dimensional solution fits better due to its lower values.

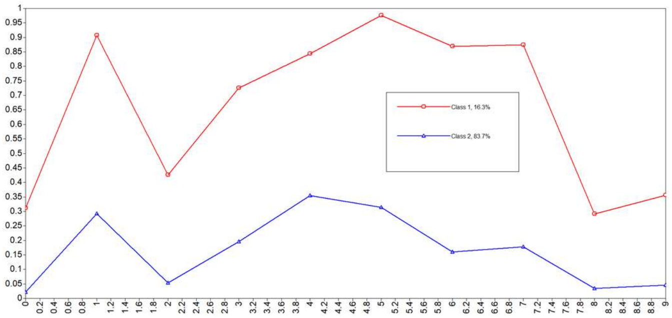


Figure 2. Sample Proportions and Estimated Probabilities. Two Classes. Class 1 (red line) 16,3%; Class 2 (blue line) 83,7%. The horizontal line indicates the stimuli (e.g. 0 = isolated pitch a, 1 = isolated pitch b, and so on). The vertical line indicates the probability of success.

Discussion

Although bibliography considers AP phenomenon as instantaneous pitch recognition (e.g., Takeuchi & Hulse, 1993) and a large number of researches choose to limit the time for responses in their tasks (e.g. Miyazaki, 1995; Barharloo et al., 1998; Athos et al., 2007), the response time cannot be measured exactly (if a subject response pitch is immediate, in 1 second, in 2 seconds or even in 3 seconds). To measure the response time exactly, it is necessary to use a reaction time; without this feature, it is impossible to state if the subject's answer was instantaneous or not. Furthermore, the association between AP phenomenon and instantaneous response is not consensual in the bibliography and in one of our previous studies (Germano et al., 2011), as some self-reported AP possessors described the need for a certain time to identify tones. Based on these considerations, we decided to give 15-seconds between the stimuli for the participant to mark the answer and to not consider reaction time as a criterion. With a longer time for response, the subject had enough time to look at the material, choose their answer, confirm the right musical note on the drawn keyboard and finally mark their response, decreasing, thus, errors that do not come from pitch discrimination.

We were very careful in trying to hamper the use of RP in this task by not giving a reference pitch, playing each stimulus only once and varying timbre and register between stimuli. Even so, we cannot assert that RP ability was eliminated, as each subject has their personal strategy to identify pitches and it is possible that some non-AP possessors had a pitch retained in their memory and had, consequently, established relations between this retained pitch and the listened pitch. It is also important to mention that AP ability does not necessarily excludes RP ability (or vice-versa): a subject can possess both abilities, although with different degrees of precision in each one.

The perspective that giving a short response time restricts the use of RP cannot be considered totally accurate for two reasons: 1) the speed of RP response time is not the same for all RP possessors and more experienced subjects can be able to identify the relation between pitches almost immediately; 2) a short time between stimuli can make it easier for the subject to memorize a previous pitch and to relate it with the present pitch; and on the other hand, a long time between stimuli can cause the subject to forget the previous pitch.

This discussion about the possible use of RP ability in our task is based on the possibility that some subjects can have some pitches retained in their memory and we argue that it is impossible to totally avoid the use of RP, not only in our task, but in all absolute pitch tasks. Besides, it is possible to find several studies that segregate AP possessors from RP possessors (e.g. Burns and Campbell, 1994; Schulze et al., 2009), but most do not apply a RP task, so the label "RP possessor" is applied for all subjects who did not perform well in the AP task. This methodology leads to the false impression that all AP possessors do not have RP, and that all non-AP possessors are automatically RP possessors. This type of methodology should be regarded with caution.

Conclusion

IRT results imply that the test has a good fit in assessing the ability to label isolated pitches without reference. They also showed that the proposed stimuli varied according to difficulty and discriminatory levels. These findings showed empirically that the common procedure adopted on AP research, where it is attributed the same score to the correct recognition for all pitches (i.e., ordinary summing of corrected answers), might be not reliable. In other words, results based on simple score sum without taking into consideration the difficulty and discrimination level of each item should be seen with caution. LCA results indicate that there are two classes of perception ability regarding isolated pitches for undergraduate music students: the first with high probability of correct recognition and the second with low probability. Comparing the dimensional solution with the categorical one, we have evidence that dimensional solution fits better. This result suggests that the frequent assumption that there is a clear distinction between AP-possessors and non-AP-possessors under the task of labeling isolated pitches might not be entirely correct. This hypothesis will be developed and evaluated in future publications.

Acknowledgements. We thank all students who volunteered in this research, the professors and universities for allowing the application of this test. We also thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for the provided funding.

References

Abraham, O. (1901). Das absolute Tonbewusstsein. Psychologisch-musikalische Studie. *Sammelbände der internationalen Musikgesellschaft*, 3, 1-86.

Asparouhov, T. & Muthén, B. (2016). *IRT in Mplus (Version 2). Technical report*. Retrieved from www.statmodel.com.

Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N., & Gitschier, J. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences*, *104* (37), 14795–14800.

Baggaley, J. (1974). Measurement of absolute pitch. *Psychology of Music*, *2* (2), 11-17.

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., & Freimer, N.B. (1998). Absolute Pitch: An Approach for Identification of Genetic and Nongenetic Components. *The American Journal of Human Genetics*, *62* (2), 224–231.

Burns, E., & Campbell, S. (1994). Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception. *Acoustical Society of America*, *96* (5), 2704-2719.

Deutsch, D. (2002). The Puzzle of Absolute Pitch. *Current Directions in Psychological Science*, *11* (6), 200-204.

Germano, N. G. (2011). Categorização de Ouvido Absoluto em Estudantes de Música de Nível Universitário nas cidades de São Paulo e Brasília. (Bachelor's Dissertation). Instituto de Artes da UNESP, São Paulo, Brazil.

Germano, N. G. (2015). Em busca de uma definição para o Fenômeno do Ouvido Absoluto. (Master's Dissertation). Instituto de Artes da UNESP, São Paulo, Brazil.

Germano, N. G., Cogo-Moreira, H., Bortz, G. (2016). Absolute Pitch: In Search of a Testable Model. *International Conference on Music Perception and Cognition 14*, ISBN: 1-876346-65-5. San Francisco, USA, 710-713.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned memories. *Perception and Psychophysics*, *56* (4), 414–423.

Marcoulides, G. & Schumacker, R. (Eds.). (2001). *New Developments and Techniques in Structural Equation Modeling.* Psychology Press.

Miyazaki, K. (1995) Perception of relative pitch with different references: Some absolute-pitch listeners can't tell musical interval names. *Perception & Psychophysics*, *57* (7), 962-970.

Miyazaki, K., & Ogawa, Y. (2006). Learning Absolute Pitch by Children. *Music Perception*, *24* (1), 63-78.

Mull, H. K. (1925). The Acquisition of Absolute Pitch. *The American Journal of Psychology*, *36* (4), 469-493.

Neu, D. M. (1947). A Critical Review of the Literature on 'Absolute Pitch'. *Psychological Bulletin*, *44* (3), 249-266.

Parncutt, R., & Levitin, D. J. (2001) Absolute pitch. In S. Sadie (Ed.), *New Grove Dictionary of Music and Musicians* (p. 37-39). New York: St. Martins Press.

Stumpf, C. (1883). *Tonpsychologie (v. 1)*. Leipzig: Verlag von S. Hirzel.

Schulze, K., Gaab, N., & Schlaug, G. (2009). Perceiving pitch absolutely: Comparing absolute and relative pitch possessors in a pitch memory task. *BMC Neuroscience*, *10* (1), 10-106.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute Pitch. *Psychological Bulletin*, *113* (2), 345–361.

Ward, W. D. (1999). Absolute Pitch. In D. Deutsch (Ed.), *The Psychology of Music (2nd Edition)* (pp. 265-298). San Diego: Academic Press.

Wellek, A. (1938). Das absolute Gehör und seine Typen [Absolute pitch and its types]. *Zeitschrift für angewandte Psychologie und Charakterkunde-Beihefte*, 83, 1–368.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences of the United States of America*, *95* (6), 3172-3177.

Musicking as Emergent Ecological Behavior: Linking Cognition, Culture and Neuroscience

Michael D. Golden

Soka University of America, USA, and
Min-On Music Research Institute

golden@soka.edu

Abstract

For musicians and students of human musicking, advances in scientific and philosophical research and scholarship in recent years offer an opportunity to address some enduring questions about this universal human behavior and its role in our daily lives and the development of our species. This paper will suggest an integrated perspective on musicking, specifically by exploring and linking together insights from a range of disciplines. While ethnomusicology has demonstrated the diversity of specific functions attributed to music around the world, a meta-analysis of this work reveals a common thread across diverse cultures; we express the sense that musicking connects us to our environments – social, physical, and/or metaphysical. To explore the significance of this seemingly abstract sense, I will draw on work in neuroscience and cognition, beginning with the Santiago theory of Maturana and Varela. One of the several components of the Santiago theory important to this paper is the idea that with a sufficiently complex nervous system, we "bring forth" both inner and outer worlds, and connect them through structural coupling. Put another way, we have as a foundation a biological correlate for the discourse concerning the value of music found in cultures around the world. Musicking can be understood as essentially an emergent connective or ecological behavior, a view consonant with current work in evolutionary musicology, "4E cognition" and auditory neuroscience. In turn, this view suggests areas for further research, including the potential of musical activities aimed at improving the state of our relationships with our various environments. These might include, for example, education, peace-building, or fostering ecological awareness. The text that follows is a synopsis of the paper presented at the conference.

Introduction

Composers and philosophers have often expressed variations of the idea that music in some way connects us to our environments (e.g., Dewey, 1934; Dunn, in Lampert, 1989; Ikeda, 2010; Scaletti, 2015), and ethnomusicologists report similar expressions from cultures around the world (for representative citations, see Golden, 2011, 2016), but there remain some prominent questions concerning our musical behaviors in the face of these expressions. What exactly is the nature or meaning of "connects" or "connection" in this context? Is there any basis in our biology for these expressions? What, if anything, is the significance or value of musical behavior to our lives and to our development as a species? Recent work in fields as diverse as cell biology, philosophy of cognition, neuroscience, ecology, and evolutionary musicology can now be brought to bear on these questions, and so my principal aim in the present paper is to contribute, at least at a theoretical level, to a synthesis of ideas that may address them.

After defining some terms and issues, I will briefly discuss the contribution made by ethnomusicologists to this study. Following that I will highlight some of the concepts introduced by Maturana and Varela (the Santiago theory) that are most germane to the ecological conception of musicking. The theory of the biological bases for cognition developed by the Santiago authors has contributed greatly to the recent development of "4E cognition" – the view that cognition is embodied, embedded, enactive, and extended – and I will next explore the relationships between these ideas and musicking-as-ecology, along with related findings in auditory neuroscience. Finally, this will lead to some discussion of recent work in evolutionary musicology, where many of these threads are coming together.

In concluding, I will touch on some possible implications of this theoretical approach, both for further research in music perception and cognition and for application in activities such as contributing to social and ecological peace-building.

Background Issues and Definitions

For ease of reference, I will re-present here my definition/conception of musicking as ecological behavior (Golden, 2016):

> Musicking is an activity of human beings involving sound and time, the function of which is to facilitate and enhance our connection with our environment. Environment here includes three mutually-related realms or domains: the social realm, the natural/physical world, and, as understood in many cultures, the metaphysical or spiritual realm. (p. 268)

The "three realms" here are really not separate from each other in essence, but with that understanding, it may still be useful to draw the distinctions, if for no other reason than that we tend to behave somewhat differently in interacting with each one. Traditionally, the thinking in musicology has been that music is primarily an intra-species (social) phenomenon, but Small (1998), who re-activated the word "musicking," notes that the meaning of musicking for us lies in the "relationships between person and person, between individual and society, between humanity and the natural world and even perhaps the supernatural world" (p. 13). I will return shortly to the notion of "connection" in these realms in the context of the Santiago theory.

A note about the terms "ecology" and "ecomusicology" may also be useful. Most work in the fairly recent field of ecomusicology involves application of music to raising ecological awareness (Pedelty, 2012), ethnomusicological case studies of particular cultures and their relationships to their natural environments (Post, 2006), or ecocriticism (Allen & Dawe, 2016). "Ecology" is used in many ways today, sometimes meaning study of any complex system, but more often as a substitute for "nature" or "environment." My approach, focusing on musicking as ecological behavior, is based on the use of "ecology" in its original sense – the study, usually scientific, of the relationships and interactions between organisms and their environments. Musicking, I will assert, is an active mode of "study" of the environment, one that has evolved in this ecosystem, and so can be called ecological.

Starting from Ethnomusicology

My claim is that the ubiquity across cultures of expressions of the *sense* that musicking connects us to our various environments is sufficient basis to propose the above inclusive definition, but also, and more importantly here, that it can be read as support and motivation for pursuit of an explanation, or at least a correlate, in terms of our biology.

There are certainly distinctions in the precise words used to describe these "connections," but the difficulties, at least at this stage, of studying (in the lab) "real" or everyday musicking experiences, in which context is crucial to understanding, suggest the need for a synthesis of a broad set of methodologies.

I have published elsewhere (Golden, 2011, 2016) collections of passages from the literature in which researchers describe conceptions found in the cultures being studied that feature the connective role of music; there are many more that could be cited. Having surveyed numerous publications including *The Garland Encyclopedia of World Music* (Nettl, Stone, Porter & Rice, 1998-2002), I can summarize these findings fairly simply. While evidence for the sense of connective function in at least one of the three realms from my definition is found in all the cultures studied, it is not the case that all three domains are discussed in all the studies. I can assert a scholarly basis in support of what is surely common knowledge; one function of human musicking is to connect people to each other. I have found no studies which do not mention at least one social-bonding function of musicking, although, again, there is a great diversity of such functions. While it may not yet be possible to "prove" the physical reality of connection to these realms of the environment, it is clear from the ethnomusicologists' work that people around the globe *feel* that music has this function.

Key Concepts from the Santiago Theory

The work of Chilean neuroscientists Maturana and Varela (1980, 1998) is particularly useful in considering musical behavior because it provides a framework for understanding human cognition, including the range of capabilities referred to as musicality, in the context of fundamental processes common to all living things. I am not the first to apply elements of the Santiago theory to musicking (Dunn, 1983; Becker, 2005, 2009; Döbereiner, 2014; van der Schyff, 2015), and I have elsewhere presented an overview of the theory (2016, 2017), so here I will focus on just a few key concepts.

Cognition, for the Santiago authors, is an essential process common to all living things, in which the organism *"brings forth"* a world according to its own *autopoietic* needs and abilities (1998, pp. 26-30). Recurrent interactions between organisms can lead to *structural coupling*, in which they modify their own structures in adapting to the other (1998, pp. 74-75). In organisms with sufficiently complex and interconnected nervous systems, the organism brings forth both inner and outer worlds, "opening new dimensions of structural coupling for the organism, by making possible in the organism the association of many different internal states with the different interactions in which the organism is involved" (1998, p. 175).

In other words, this theory suggests that there is a biological basis or correlate for the sensation reported in cultures around the world that musicking connects our inner selves with our environments.

Embodied, Embedded, Enacted and Extended: Musicking and Contemporary Cognitive Science

Edwin Hutchins (2010) traces the development of what is commonly called "4E cognition," the movement towards understanding the phenomenon of cognition as both biological and ecological, which has brought together ideas from, among others, Bateson (1972), Gibson (1986), Varela, Thompson and Rosch (1991), Lakoff and Johnson (1999) and the Santiago authors.

Hutchins states that the central question in studying cognition is determining "the right boundaries for a unit of analysis" (2010, p. 705), and that the answer is increasingly understood to be a brain-body-world ecosystem, i.e., a dynamic interaction among numerous elements and processes; applied to the current topic, it could be argued that focusing narrowly on a smaller unit, such as *only* neural activity or *only* social interactions, cannot fully explain cognitive behavior of any kind, including musicking.

My definition of musicking aligns with these ideas in that musicking is behavior (enacted), it engages sound and thus the brain and body (embodied), and it is both inseparable from and extensive to the environment (embedded and extended.) Musicking is thus a mode of cognition, something we *do* in order to better *know* (and perhaps be known by) the environment, and so can be understood as ecological behavior, in the sense described earlier.

Firing and Wiring: Musicking and Neuroscience

If neural phenomena alone cannot completely explain the ecosystem of cognition, they clearly constitute an essential component of it. Edelman's theory of neuronal group selection (Edelman & Tononi, 2000) suggests that the structure of the brain is ultimately shaped by the neurons' own activity, which in turn is "selected" by behavioral experience. In terms of the Santiago theory, the extraordinary reciprocal and recursive interconnections among groups of neurons, which Edelman calls "reentrant organization," enable the "new dimensions of structural coupling," which in turn enable the synchronized, coordinated activity that ultimately forms the basis for the "collectively-brought-forth" inner world.

Recent research in neuroscience (Brown, Martinez & Parsons, 2006; Meyer, Elmer, Baumann & Janke, 2007; Turner & Ioannides, 2009; Groussard et al., 2010; Kraus & Chandrasekaran, 2010; Wan & Schlaug, 2010; Rauscher, 2011; Schlaug, 2011; Chaudry, Nag, Jain & Wadhwa, 2013) supports the notion that musicking serves a connective function within the environment of the brain itself. Although some controversy remains, it appears there is no single "music module" in our brains; rather, musicking engages multiple brain systems – auditory, temporal, motor, emotional, and so on – and in particular, inter-hemispheric connections across the corpus callosum.

In terms of the social realm, Bharucha, Curtis, and Paroo

(2012) write that "the underlying function of musical communication is not to communicate anything per se ... but rather to align our brain states and foster social cohesion," (p. 139), and Abrams et al. (2013) write that "music synchronizes brain responses across listeners" (p. 1458). These observations also dovetail with the understanding of communication in the Santiago theory, i.e., that it is actually the coordination of behavior and not the transmission of information (1998, pp. 193-196).

Of special interest to some researchers (Dissanayake, 2000, 2009; Cross, 2003, Parncutt, 2009, among others) is the impact of early "proto-musical" interactions between infants and caregivers, especially mothers. These behaviors may prove essential to the later development not just of musicality, but of many of the attributes and abilities we consider most human.

Emergence and Evolutionary Musicology

A full review of the recent developments in this field is beyond the scope of this synopsis, but I refer the reader to Cross (2003, 2012), Tomlinson (2015) and van der Schyff and Schiavio (2017) among others, whose work is the basis of this section. Of greatest relevance to this paper are the ideas centering on biocultural coevolution and emergence. In this view, evolution is a complex system process involving feedback cycles among organisms and environments, but with the addition of cultural epicycles, evolutionary processes in themselves (see Tomlinson 2015, pp. 41-47). Tomlinson considers the emergence of modern musicking as the gradual coalescing of the range of neural and cognitive capacities required for this complex behavior; from the perspective of the Santiago theory and 4E cognition, these can be understood as results of recursive interactions (structural coupling) within and among living systems. Put another way, musicking is an evolving behavior that both takes advantage of and recursively contributes to our evolving biological capabilities.

Conclusion

I have outlined here what I hope can be seen as a way of integrating some of the diverse approaches to understanding human musicking.

To recapitulate, the sense of connection through music, as reported in the cultural studies, is a mode of cognition, as understood in the Santiago theory. Being enacted and embodied, musicking is also embedded and extended and thus ecological behavior, both in terms of the relationships between inner and outer worlds and of having emerged in conjunction with the properties of life in this ecosystem.

Thinking of musicking in these ways may open new approaches for application and research in various arenas. Van der Schyff (2015), van der Schyff and Schiavio (2017) and I (Golden, 2016), among others, have suggested ways in which these ideas might contribute to education. Organizations such as Musicians without Borders and the Min-On Music Research Institute and others are working to develop ways to apply ideas about music's connective and coordinating functions to peace-building activities, and there are a plethora of individuals and organizations working on environmental issues and promoting ecological awareness who might fruitfully apply these ideas in their work as well. For cognitive scientists and neuroscientists, investigating the neural phenomena involved in linking brought-forth inner and outer worlds during musical activity and correlating the results with phenomenological or anthropological approaches might also prove an interesting avenue for research.

References

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., & Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal of Neuroscience, 37*(9), 1458-1469.

Allen, A. S., & Dawe, K. (Eds.). (2016). *Current directions in ecomusicology: Music, culture, nature*. New York, NY: Routledge.

Bateson, G. (1972). *Steps to an ecology of mind*. Chicago, IL: Univ. of Chicago Press.

Becker, J. (2005). *Deep listeners: Music, emotion, and trancing*. Bloomington, IN: Indiana Univ. Press.

Becker, J. (2009). Ethnomusicology and empiricism in the twenty-first century. *Ethnomusicology, 53*(3), 478-501.

Bharucha, J., Curtis, M., & Paroo, K. (2012). Musical communication as alignment of brain states. In P. Rebuschat, M. Rohmeier, J. Hawkins & Cross, I. (Eds.), *Language and music as cognitive systems* (pp. 139-155). Oxford, UK: Oxford Univ. Press.

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience, 27*, 2791-2803.

Chaudry, S., Nag, T. C., Jain, S. & Wadwha, S. (2013). Role of sound stimulation in reprogramming brain connectivity. *Journal of Biosciences, 38*(3), 605-614.

Cross, I. (2003). Music and evolution: Consequences and causes. *Contemporary Music Review, 22*(3), 79-89.

Cross, I. (2012). Music and biocultural evolution. In M. Clayton, T. Herbert & R. Middleton (Eds.), *The cultural study of music: A critical introduction* (pp. 17-27). New York, NY: Routledge.

Dewey, J. (1934). *Art as experience*. New York, NY: Penguin Group.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In N. Wallin, B. Merker & S. Brown (Eds.), *The origins of music*. (pp. 389-410). Cambridge, MA: The MIT Press.

Dissanayake, E. (2009). Root, leaf, blossom, or bole: concerning the origin and adaptive function of music. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 17-30). New York, NY: Oxford Univ. Press.

Döbereiner, L. (2014). Resonances of subjectivity: Hegel's concept of sound and Di Scipio's audible ecosystems. *Contemporary Music Review, 33*(1), 19-30.

Dunn, D. (1983). Speculations: On the evolutionary continuity of music and animal communication behavior. *Perspectives of New Music, 22*(1/2), 87-102.

Edelman, G. M. & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York, NY: Basic Books.

Gibson, J. J. (1986). *The ecological approach to visual perception*. New York, NY: Psychology Press.

Golden, M. (2011). On music, interconnection, and consciousness. In F. Laurence & O. Urbain (Eds.), *Music and solidarity: Questions of universality, consciousness and connection* (pp. 31-46). New Brunswick, NJ: Transaction Publishers.

Golden, M. (2016). Musicking as education for social and ecological peace. *Journal of Peace Education, 13*(6), 266-282. http://dx.doi.org/10.1080/17400201.2016.1234612.

Golden, M. (2017). Music emergent: Autopoiesis and connected worlds. *Proceedings of A Body of Knowledge Conference*. Irvine, CA: Univ. of California Press. https://escholarship.org/uc/item/0qr0z45g.

Groussard, M., La Joie, R., Rauchs, G., Landeau, B., Chételat, G., Viader, F., Desgranges, B., Eustache, F. & Platel, H. (2010). When music and long-term memory interact: Effects of musical expertise on functional and structural plasticity in the hippocampus. *PLoS ONE 5*(10): e13225. doi: 10.1371/journal.pone.0013225.

Hutchins, E. (2010). Cognitive ecology. *Topics in Cognitive Science*, 2. 705-715.

Ikeda, D. (2010). Creative life. In *A new humanism: The university addresses of Daisaku Ikeda* (pp. 3-11). London,UK and New York, NY: I.B. Tauris.

Kraus, N. & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews: Neuroscience, 11*, 599-605.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought.* New York, NY: Basic Books.

Lampert, M. R. (1989). "Environment, consciousness and magic: An interview with David Dunn." *Perspectives in New Music, 27*(1), 94-105

Maturana, H. R. & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Dordrecht, Holland: D. Reidel

Maturana, H. R. & Varela, F. J. (1998). *The tree of knowledge: The biological roots of human understanding* (Rev. Ed.). Boston, MA and London, UK: Shambhala Publications.

Meyer, M., Elmer, S., Baumann, S. & Jancke, L. (2007). Short-term plasticity in the auditory system: Differential neural responses to perception and imagery of speech and music. *Restorative Neurology and Neuroscience, 25*, 411-431.

Nettl, B., Stone, R. M., Porter, J. & Rice, T. (Eds.). (1998-2002). *The Garland encyclopedia of world music* (Vols. 1-10). New York, NY: Routledge.

Pedelty, M. (2012). *Ecomusicology: Rock, folk, and the environment.* Philadelphia, PA: Temple Univ. Press.

Post, J. C. (Ed.). (2006). *Ethnomusicology: A contemporary reader.* New York, NY: Routledge.

Rauscher, F. H. (2011). The impact of music instruction on other skills. In S. Hallam, I. Cross & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 244-252). Oxford, UK: Oxford Univ. Press.

Scaletti, C. (2015). *Sonification of the world*. Retrieved from https://eartotheearth.org/2015/01/sonification-of-the-world.

Schlaug, G. (2011). Music, musicians, and brain plasticity. In S. Hallam, I. Cross & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 197-207). Oxford, UK: Oxford Univ. Press.

Small, C. (1998). Musicking: *The meanings of performing and listening*. Middletown, CT: Wesleyan Univ. Press.

Tomlinson, G. (2015). *A million years of music: The emergence of human modernity*. Brooklyn, NY: Zone Books.

Turner, R. & Ioannides, A. A. (2009). Brain, music and musicality: Inferences from neuroimaging. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 147-181). New York, NY: Oxford Univ. Press.

van der Schyff, D. (2015). Music as a manifestation of life: exploring enactivism and the "eastern perspective" for music education. *Frontiers in Psychology, 6*. doi: 10.3389/fpsyg.2015.00345

van der Schyff, D. & Schiavio, A. (2017). Evolutionary musicology meets embodied cognition: Biocultural coevolution and the enactive origins of human musicality. *Frontiers of Neuroscience*, *11:*519. dDoi: 10.3389/fnins.2017.00519.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, MA: The MIT Press.

Wan, C. Y. & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *The Neuroscientist, 16*(5), 566-577.

Analysis of Objective Descriptors for Music Performance Assessment

Siddharth Gururani, Kumar Ashis Pati, Chih-Wei Wu and Alexander Lerch

Center for Music Technology, Georgia Institute of Technology, USA

{sgururani3, apati3, cwu307, alexander.lerch}@gatech.edu

Abstract

The assessment of musical performances in, e.g., student competitions or auditions, is a largely subjective evaluation of a performer's technical skills and expressivity. Objective descriptors extracted from the audio signal have been proposed for automatic performance assessment in such a context. Such descriptors represent different aspects of pitch, dynamics and timing of a performance and have been shown to be reasonably successful in modeling human assessments of student performances through regression. This study aims to identify the influence of individual descriptors on models of human assessment in 4 categories: musicality, note accuracy, rhythmic accuracy, and tone quality. To evaluate the influence of the individual descriptors, the descriptors highly correlated with the human assessments are identified. Subsequently, various subsets are chosen using different selection criteria and the adjusted R-squared metric is computed to evaluate the degree to which these subsets explain the variance in the assessments. In addition, sequential forward selection is performed to identify the most meaningful descriptors. The goal of this study is to gain insights into which objective descriptors contribute most to the human assessments as well as to identify a subset of well-performing descriptors. The results indicate that a small subset of the designed descriptors can perform at a similar accuracy as the full set of descriptors. Sequential forward selection shows how around 33% of the descriptors do not add new information to the linear regression models, pointing towards redundancy in the descriptors.

I. Introduction

A musical performance by a human requires the interpretation of musical ideas, typically from a score, the planning of retrieved musical units and transforming thoughts into motion. This makes the task one of the most complex serial actions performed by a human (Palmer, 1997). Therefore, students learning to perform music require regular feedback and attention from trained teachers. This feedback may be in the form of qualitative and quantitative assessment of the student performances on various criteria such as rhythmic and pitch accuracy.

A common issue with such feedback, however, is its highly subjective nature (Wesolowski, 2012). This may include bias and lead to inconsistencies counter-productive to the students' learning experience. Thus, computational methods for music performance assessment are sought after because they are able to provide objective, consistent, and reproducible assessments.

Tools and techniques for automatically extracting musical information from audio signals matured as a result of advances in music information retrieval (MIR). Relevant tasks are pitch and beat tracking, transcription and source separation in audio signals. These MIR tasks may be treated as fundamental building blocks for automatic music performance assessment. Several academic works leverage these techniques to develop automatic music performance assessment systems (Dittmar, Cano, Abeßer, & Grollmisch, 2012). In addition, companies such as Smart Music (http://www.smartmusic.com Last access: 2018/05/26) and Yousician (https://get.yousician.com Last access: 2018/05/26) have developed commercial software for performance assessment.

Typical automatic performance assessment systems comprise of algorithms that extract descriptors or features from the audio signal. These descriptors are, in turn, used to train statistical models using expert assessment data. The models are then applied to predict the assessment scores. In this paper, we focus on the importance of these descriptors for modelling student performances.

The structure of the paper is as follows: Section 2 provides a brief overview of related work. Section 3 introduces the descriptors and the dataset used for the analysis. The methodology and experimental setup are described in Section 4. The results of the experiments and conclusion follow in Sections 5 and 6 respectively.

II. Related Work

Objective descriptors, also known as audio features, are quantities that represent various characteristics of an audio signal. These are typically computed on short blocks of the audio signal to capture short-term characteristics and then summarized using statistical measures such as the mean and standard deviation (Lerch, 2012). In the context of music performance assessment, different descriptors may be used to capture low-level information pertaining to signal energy and timbre and subsequently linked to high-level semantic concepts through computational models.

According to work done by Vidwans et al. (Vidwans et al., 2017), objective descriptors for music performance may be broadly categorized into two major categories: (i) Score-independent and (ii) Score-dependent descriptors. Score-independent descriptors are derived without using additional information about the musical score being performed. The benefit of using these descriptors is that they only require the audio file and do not rely on the availability of the score. The intuition behind using this approach is that humans are able to assess performances even without the score by observing pitch and rhythm stability, among others. Some examples of work involving score independent features are (Abeßer, Hasselhorn, Dittmar, Lehmann, & Grollmisch, 2013), (Han & Lee, 2014) and (Wu et al., 2016).

Score-dependent descriptors are derived by leveraging the information provided by the score being performed. The advantage of using this approach is that direct comparison between the performance and the score is possible and therefore, more accurate descriptors of the performance may be extracted. These are applicable in the audition setting where assessors have access to the score that is supposed to be performed. Work investigating score-dependent features are (Vidwans et al., 2017), (Devaney, Mandel, & Fujinaga, 2012) and (Mayor, Bonada, & Loscos, 2009).

In addition to the aforementioned categories, objective descriptors can also be automatically inferred from the data using machine learning techniques such as sparse coding (Wu & Lerch, 2018) or neural networks (Pati, Gururani, & Lerch, 2018). Learned features allow the extraction of relevant information that might be overlooked by human engineers, however, these features tend to be abstract and have no specific physical meaning. Thus, they are outside the scope of this paper.

In previous work by (Wu et al., 2016) and (Vidwans et al., 2017), score-independent and score-dependent descriptors were designed and shown to be useful for the task of automatic music performance assessment. These studies trained regression models using the large set of descriptors to achieve the best performance. However, an analysis of the importance or contribution of the descriptors is not performed. To increase the interpretability of such approaches and gain more insights about the system, we aim to analyze these descriptors in detail using various methods and identify the well-performing set of descriptors from among the larger set of descriptors.

III. Dataset and Descriptors

Dataset

The dataset used in this paper is obtained from the Florida Bandmasters Association (FBA). It consists of audio recordings of All-State auditions of middle and high school students. Each recording consists of exercises such as etudes, scales, and sight reading and is accompanied by expert assessments in the four following categories: musicality, note accuracy, rhythmic accuracy and tone quality. For more details about the dataset, we refer readers to our previous work, (Wu et al., 2016) and (Vidwans et al., 2017). We consider middle school students performing alto saxophone (n = 392). Only the technical etude is considered for these experiments.

Descriptors

The descriptors investigated here have shown their meaningfulness in previous studies. They were designed to model different facets of a student performance. We provide a brief overview of the descriptors used in this paper and refer readers to work done by (Vidwans et al., 2017) for a detailed description of all the descriptors designed for this task.

As described in Section 2, the descriptors chosen are broadly categorized into two classes:

1. Score-independent
2. Score-dependent

The score-independent descriptors may be further divided into 3 categories:

Pitch: Descriptors extracted from the pitch contour of the performance fall under this category. They include measures for note steadiness, accuracy and intonation. They are computed on a note-by-note basis and aggregated for an entire performance using the mean, standard deviation, maximum and minimum value.

Rhythm: Descriptors extracted from the inter-onset-interval (IOI) histogram computer from note onset times. They measure the timing accuracy of the note. Standard statistical measures are extracted from the histogram such as crest, skewness, rolloff, etc.

Table 1. Score-Independent Descriptors

Index	Descriptor	Description
N1	Pitch 1	Average note accuracy
N2-5	Pitch 2	St. dev. of pitch values (mean, st. dev., min, max)
N6-9	Pitch 3	% of pitch values deviating more than one st. dev. (mean, st. dev., min, max)
N10	Intonation	% of notes in tune
N11-14	Dynamics 1	Amplitude deviation (mean, st. dev., min, max)
N15-18	Dynamics 2	Amplitude envelope spikes (mean, st. dev., min, max)
N19-24	Rhythm	Crest, bin resolution, skewness, kurtosis, roll-off, power ratio of the IOI histogram

Dynamics: Descriptors extracted from the amplitude of each note. These include note-level descriptors that measure amplitude steadiness or spikes. Similar to pitch descriptors, they are aggregated using the mean, standard deviation, maximum and minimum across all notes.

Score-dependent descriptors are computed after aligning the pitch sequence or contour of the performance with the score of the performance. Alignment gives a more accurate segmentation of the notes in the performance. The score-dependent descriptors are also of 3 types:

Pitch: Most of the pitch descriptors are similar to the score-independent descriptors and differ in the fact that note boundaries are computed using score alignment. In addition, we compute descriptors measuring the deviation of the played note from the note in the score.

Rhythm: Similar descriptors as the score-independent are computed with score-aligned note onsets.

Alignment: The score alignment is performed using dynamic time warping (Müller, 2007) of the pitch contour with the sequence of notes in the score. Descriptors are extracted from the alignment path such as the length and deviation of the slope from a straight line. In addition, the cost of alignment and a measure of extra notes or unplayed notes are computed.

There are 46 descriptors that are analyzed in this paper. 24 of these are score-independent and 22 are score-dependent. We index these descriptors prefixed by 'N' for score-independent and 'S' for score-dependent descriptors. Tables 1 and 2 enumerate all the descriptors and their index.

IV. Experiments

In this section, we describe the two experiments carried out to analyze the importance of the extracted descriptors. The first experiment involves studying the direct correlation between the descriptors and the assessments. The second involves different methods for selecting descriptor subsets and subsequently using these subsets to train linear regression models to predict the assessments.

Correlation Analysis

In this experiment, we aim to study how correlated or decorrelated each of the descriptors is with the human assessments. Since these descriptors have been used to model human assessments, we investigate whether a relation can be

Table 2. Score-dependent Descriptors

Index	Descriptor	Description
S1	Note insertion ratio	Notes inserted incorrectly / Total # of notes
S2-5	Pitch 1	Mean of difference between played pitch and score (mean, st. dev., min, max)
S6-9	Pitch 2	St. dev. of difference between playe1d pitch and score (mean, st. dev., min, max)
S10-13	Pitch 3	% of values deviating from score more than 1 st. dev. (mean, st. dev., min, max)
S14	DTW 1	Cost of DTW alignment
S15	DTW 2	Average deviation of alignment path from straight line
S16-21	Rhythm	Crest, bin resolution, skewness, kurtosis, roll-off, and power ratio of IOI histogram
S22	Note deletion ratio	Notes removed incorrectly / Total # of notes

found using the spearman correlation. Note that these descriptors are used in machine learning models which are able to parameterize the assessments as linear or non-linear combinations of the descriptors and hence there may or may not be a direct monotonic relationship between them.

We compute the spearman correlation coefficient of each of the 46 descriptors with each of the 4 human assessments and use the results in further experiments.

Descriptor Selection

In this experiment, we aim to identify the set of descriptors that are best able to explain the variance in each of the human assessments. This is computed by constructing linear regression models using various combinations or subsets of the descriptors at hand. Since the search space for the subsets is very large we first narrow down the search space by applying the following two criteria:

- Top 10 descriptors based on spearman correlation
- |Spearman correlation| > 0.25

We report the adjusted R-squared for the regression models and compare it to a model trained using all the descriptors.

In addition to this, we perform a sequential forward selection of the descriptors. In this experiment, we start with the best descriptor (the one that achieves highest R-squared) and iteratively check for the combination of descriptors with the highest adjusted R-squared and add it to our set of descriptors until the adjusted R-squared stops increasing.

Note that we do not perform validation using methods such as 10-fold cross-validation as used in previous work since our goal here is to understand how well each descriptor explains the variance in the entire dataset. Based on the identified descriptors, predictive models can be trained using a cross-validation scheme. However, the evaluation of such models is out of the scope of this study.

V. Results

The results for feature selection based on spearman correlation coefficient of each feature with the different human assessments are shown in Table 3. Each of the correlation values were statistically significant with $p < 10^{-4}$.

Table 3. Subsets of descriptors chosen based on spearman correlation coefficient of individual descriptors with the human assessments. The descriptors are arranged in decreasing order of correlation. Underlined feature indices indicate positive correlation.

	Top 10	|r| > 0.25
Musicality	N20, S17, S22, S10, N3, N15, N17, N2, S7, S1	N20, S17, S22, S10, N3, N15
Note Accuracy	S17, S14, S7, S2, S8, S20, N20, S6, S3, S10	S17, S14, S7, S2, S8, S20, N20
Rhythmic Accuracy	S17, N20, S14, S10, S3, S2, S20, S8, N15, S4	S17, N20, S14, S10, S3, S2, S20, S8, N15
Tone Quality	S17, N20, S1, S10, S2, S14, S3, S8, S7, S5	S17, N20, S1, S10, S2, S14, S3, S8, S7, S5

We can make the following observations:

- Most of the descriptors used are negatively correlated with the assessments. This makes sense because these descriptors are trying to summarize the mistakes made by the student performer.
- Descriptor S17 ranks very high for all assessment categories. The average spearman correlation with the assessments is -0.44. This feature is a measure of the IOI histogram's bin resolution. A high value indicates larger variance in tempo of the performance which is undesirable for the technical etude.
- Most of the top ranked descriptors are score-dependent descriptors. This is likely due to the descriptors being dependent on correct note segmentation of the performance. The additional score information allows, as expected, for a more robust note description and thus a more accurate extraction of performance parameters.

In Figure 1 we compare different regression models trained using 3 different subsets of descriptors: All descriptors, descriptors with correlation > 0.25 and the top 10 descriptors based on their correlation with the assessment. We observe that the smaller subsets are able to account for a large degree of variance explained by the entire set for Musicality, in particular. This is not true for Rhythmic Accuracy, which might be since the best descriptors (S17 and N20) are highly correlated and only 3 of the 12 Rhythmic Accuracy descriptors appear in the top 10. We also observe that the descriptors are poor at predicting Tone Quality. This is most likely due to the fact that we do not have descriptors for timbral characteristics of the performance.

Finally, Figure 2 shows the results for the sequential forward selection. We observe that after around 20 to 30 iterations over the descriptors, the models stop improving. This could be due to the fact that the remaining descriptors are not adding any new information and are not causing any improvement in the models' predictive accuracy. This calls for removal of redundant descriptors and addition of new descriptors. Another possible explanation for this is the curse of dimensionality (Friedman, 1997). This implies that given the number of descriptors, the amount of data is insufficient for the model to take advantage of additional information.

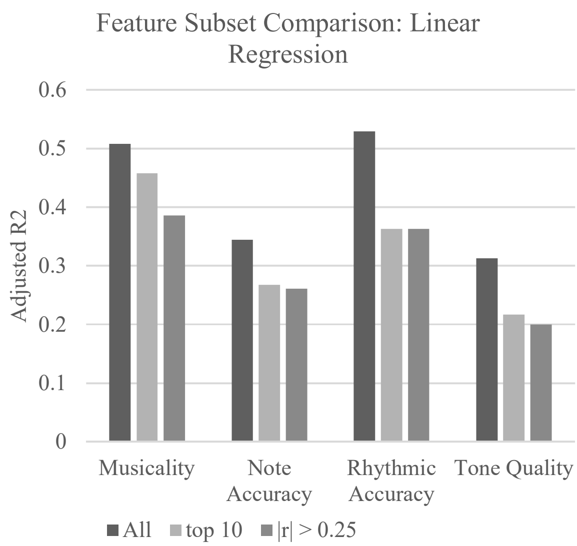


Figure 1. Comparison of linear regression models fitted with different subsets of descriptors.

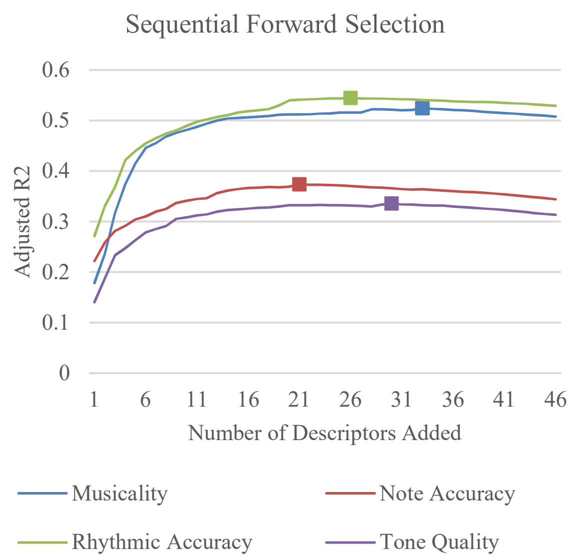


Figure 2. Learning curves for linear regression models trained using sequential forward selection. The boxes indicate the point where the adjusted R-squared starts decreasing.

From this experiment, we were able to identify the set of features that are able to explain the maximum variance in each assessment criteria. These descriptors are the ones that are selected for the model indicated by the box in Figure 2. We can also observe that these descriptors explain the variance to a greater degree than all the subsets of descriptors used in Figure 1. This is due to the fact that the descriptors chosen using correlation as the metric may be correlated amongst themselves while in the forward selection procedure, descriptors are added to the subset based on the increase in R-squared, leading to descriptors that are not highly correlated amongst themselves.

The first iteration of the experiment selects the descriptor that is best able to explain the variance among all descriptors. For Musicality, it is descriptor N20 which is the score-independent IOI histogram bin resolution (adjusted R2 = 0.18). For Note Accuracy, it is descriptor S14 which is the DTW cost (adjusted R2 = 0.22). For Rhythmic Accuracy, it is descriptor S17 which is the score-dependent IOI histogram bin resolution (adjusted R2 = 0.27). For Tone Quality it is descriptor S1 which is the note insertion ratio (adjusted R2 = 0.14).

We observe from Figure 2 that the performance for Musicality increases rapidly with the iterations. This is possibly due to the fact that Musicality is loosely defined, and the assessments are better explained with a combination of different kinds of descriptors. For Note Accuracy, the curve is flatter implying that the variance is captured in early iterations and subsequent descriptors are redundant. The DTW cost is a relevant descriptor for note accuracy since it is computed by aligning the pitch contour and the note sequence. For Rhythmic Accuracy, the best descriptor explains the variance to a greater degree than the other categories which may be attributed to the relevance of IOI histogram bin resolution. In the results for correlation analysis, it was shown to be among the top descriptors. In the case of Tone Quality, the descriptors perform the worst. This reiterates the fact that the current set of descriptors lack in terms of capturing timbral characteristics of the student music performance.

Conclusion

In this paper, we perform an in-depth analysis of 46 hand-crafted descriptors for the assessment of student alto saxophone performances to quantify their relevance. Our experiments show that a subset of the descriptors is well correlated with the human assessments. In addition, we find that the score-dependent descriptors are better correlated with the assessments compared to the score-independent descriptors.

We select subsets of descriptors with relatively high correlations with the assessments and construct linear regression models. We use the adjusted R-squared metric to compare the descriptor subsets with the whole set. The results from this experiment show that some of the top descriptors are able to account for a large degree of the variance explained by the entire set. The experiment for sequential forward selection shows that only around 30 of the 46 descriptors are selected for the models with the highest adjusted R-squared for each individual assessment category. This may be explained by redundancy in the descriptors or that the dimensionality is too high after a point and more data is required for improvement in regression performance.

As future work, we aim to remove the redundant descriptors and add new descriptors which can help capture a higher degree of variance in each assessment category focusing on adding features relevant to timbral characteristics.

Acknowledgements. We would like to thank the Florida Bandmasters Association (FBA) for kindly providing us with the dataset used in this work.

References

Abeßer, J., Hasselhorn, J., Dittmar, C., Lehmann, A., & Grollmisch, S. (2013, October). Automatic quality assessment of vocal and instrumental performances of ninth-grade and tenth-grade pupils. In *Proceedings of the International Symposium on Computer Music Multidisciplinary Research (CMMR), Marseille, France* (pp. 975–988)

Devaney, J., Mandel, M. I., & Fujinaga, I. (2012, October). A Study of Intonation in Three-Part Singing using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT). In *Proceedings. of the International Society for Music Information Retrieval Conference (ISMIR), Porto, Portugal* (pp. 511-516).

Dittmar, C., Cano, E., Abeßer, J., & Grollmisch, S. (2012). Music information retrieval meets music education. In *Dagstuhl Follow-Ups* (Vol. 3). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Friedman, J. H. (1997). On bias, variance, 0/1—loss, and the curse-of-dimensionality. *Data mining and knowledge discovery*, *1*(1), 55-77.

Han, Y., & Lee, K. (2014, October). Hierarchical Approach to Detect Common Mistakes of Beginner Flute Players. In *Proceedings. of the International Society for Music Information Retrieval Conference (ISMIR), Taipei, Taiwan* (pp. 77-82).

Lerch, A. (2012). *An introduction to audio content analysis: Applications in signal processing and music informatics*. John Wiley & Sons.

Mayor, O., Bonada, J., & Loscos, A. (2009, February). Performance analysis and scoring of the singing voice. In *Proceedings 35th AES International. Conference., London, UK* (pp. 1-7).

Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, 69-84.

Palmer, C. (1997). Music performance. *Annual review of psychology, 48(1)*, 115-138.

Pati, K. A., Gururani, S., & Lerch, A. (2018). Assessment of Student Music Performances Using Deep Neural Networks. *Applied Sciences*, *8*(4), 507.

Vidwans, A., Gururani, S., Wu, C. W., Subramanian, V., Swaminathan, R. V., & Lerch, A. (2017, June). Objective descriptors for the assessment of student music performances. In *Proceedings of 2017 Audio Engineering Society (AES) Conference on Semantic Audio.*

Wesolowski, B. C. (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, *98*(3), 36-42.

Wu, C. W., Gururani, S., Laguna, C., Pati, A., Vidwans, A., & Lerch, A. (2016, July). Towards the Objective Assessment of Music Performances. In *Proc. of the International Conference on Music Perception and Cognition (ICMPC)* (pp. 99-103).

Wu, C. W., & Lerch, A. (2018, February). Learned Features for the Assessment of Percussive Music Performances. In *Proceedings of the International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA* (Vol. 31).

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Consonant Length as Expressive Resource in Sung Spanish

Mariano Nicolás Guzmán[1], and Favio Shifres[2]

Laboratorio para el Estudio de la Experiencia Musical, Universidad Nacional de La Plata, Argentina

[1]marianoguzman791@gmail.com, [2]fshifres@fba.unlp.edu.ar

Abstract

In classical singing, the pronunciation of Spanish is subordinated to an aesthetic canon which limits the expressive range of speech. According to this canon, vowels have a leading role since they can be sustained, while consonants must be articulated "clearly" but "marked". Due to these requirements, consonants must be shortened in classical singing, which ignores their variability and identity effects on communication. In order to study the imposition of this aesthetic canon in classical singing and outside of it, we measured the length of the consonants /l m n/ (which can be sustained) in 10 famous singers' recordings (5 classical and 5 folk) of "La Tempranera" by Carlos Guastavino. The correlation between syllable length and consonant length was significant in all cases, which indicates that the consonants /l m n/ keep in proportion with the subsequent vowels. The absolute and relative lengths were higher in folk-style performances (means = 0.109 s 27.61%) than in classical ones (means = 0.090 s 21.86%). Nevertheless, the data showed a high length variability in both singing styles. These results show that in Spanish folk singing the consonants /l m n/ in consonant-vowel syllables tend to be longer than in classical singing. However, although the imposed aesthetic canon seems to have an effect on the classical performances' pronunciation, the evidence suggests that the length of the consonants /l m n/ is used in an expressive way in both singing styles.

Background

A review of 19th-century Spanish literature on vocal pedagogy (Guzmán, Shifres & Carranza, September 2017) found that in classical singing the pronunciation is subordinated to an aesthetic canon which limits the expressive range of speech. According to this canon, vowels have a leading role since they can be sustained, while consonants must be articulated "clearly" but "marked". Due to these requirements, consonants must be shortened in classical singing (Miller, 1996), which ignores the variability of segmental length in spoken Spanish (Mendoza et al., 2003) and its identity effects on communication (Carter & Wolford, 2016).

Although phonemes have traditionally been considered meaningless linguistic structures, it has been proposed that they get expressive value when placed in context, within a word or a phrase (Alarcos Llorach, 1950; Posadas de Julián, 2008). For example, the sound /s/ in <u>s</u>*igh* can evoke the exhalation in this action through its own hissing or *sibilance*. Due to each phoneme contains a bundle of articulatory features (such as *sibilance*, *precision*, *length*, etc.), it would be possible to take advantage of them to increase the expressive range of singing (Guzmán, 2017). However, the common scission between technique and expression in musical training (Shifres, 1994) reduces the role of pronunciation to technical purposes, ignoring its expressive nature.

Studies on diction for singing (Mahaney, 2006) show this scission. They examine how speech sounds are written and transform this knowledge into a system of *phonetic rules* that musicians can apply to the study of vocal pieces in a foreign language. In a certain way, these rules contribute to the development of a *proper* vocal performance and give it intelligibility. Nevertheless, we note that putting pronunciation at the service of technical problems tends to homogenise it and deprives singers of expressive resources that are possible in speech (such as the consonant length).

Although recent studies compare the expressiveness in speaking and singing (Scherer, Sundberg, Tamarit & Salomão, 2015), the scope of consonant length as expressive resource in sung Spanish is still unknown.

Aims

Study the imposition of the aesthetic canon of classical singing on the length of 3 sustainable consonants in sung Spanish and how they are articulated outside of that canon in a more spontaneous way.

Methods

Ten famous singers' recordings (5 classical and 5 folk) of "La Tempranera" (a classical chamber song by Carlos Guastavino) were phonemically segmented and labelled by Praat 6.0.26 through auditory and visual evaluations of their acoustic signals. Since this song is composed in a zamba rhythm (an Argentinian folk dance), it is widely performed by both classical and folk singers. The length of the consonants /l m n/ (which can be sustained) was measured in all available consonant-vowel (CV) syllables, as well as the full syllables that contain them (e.g., *e<u>le</u>gía*, *<u>mí</u>a*, *tempra<u>ne</u>ra*, etc.). Thus, it was possible to study the length of the consonants /l m n/ at the syllable onsets, where these sounds should be shortened in classical singing to give way to the vowel.

Results

The correlation between syllable length and consonant length was significant in all cases, which indicates that the consonants /l m n/ keep in proportion with the subsequent vowels. The absolute and relative lengths were higher in folk-style performances (means = .109 s 27.61%) (Table 2) than in classical ones (means = .090 s 21.86%) (Table 1). Differences are statistically significant ($F_{[1\text{-}54]} = 44.459$; $p < .000$). However, the differences between consonants are not significant, which means that classical singers tend to shorten the relative length of these sounds in all cases.

Table 1. The descriptive statistics of classical performances.

	N	Min.	Mean	Max.	SD
CV syllable length (s)	285	0.133	0.648	4.561	0.591
Absolute C length (s)	285	0.015	**0.090**	0.475	0.053
Relative C length (%)	285	1.42	**21.86**	60.79	15.59

Table 2. The descriptive statistics of folk performances.

	N	Min.	Mean	Max.	SD
CV syllable length (s)	285	0.144	0.615	5.533	0.688
Absolute C length (s)	285	0.020	**0.109**	0.444	0.067
Relative C length (%)	285	2.27	**27.61**	68.13	16.29

In addition, the relative minimum length was 37.23% lower in classical performances, which indicates that –even in short onsets– folk singers give more presence to the consonants /l m n/. However, although the standard deviations of the relative lengths were also higher in folk performances, we observed a high variability in consonant length of both singing styles (mean = 9.18%) (Figure 1).

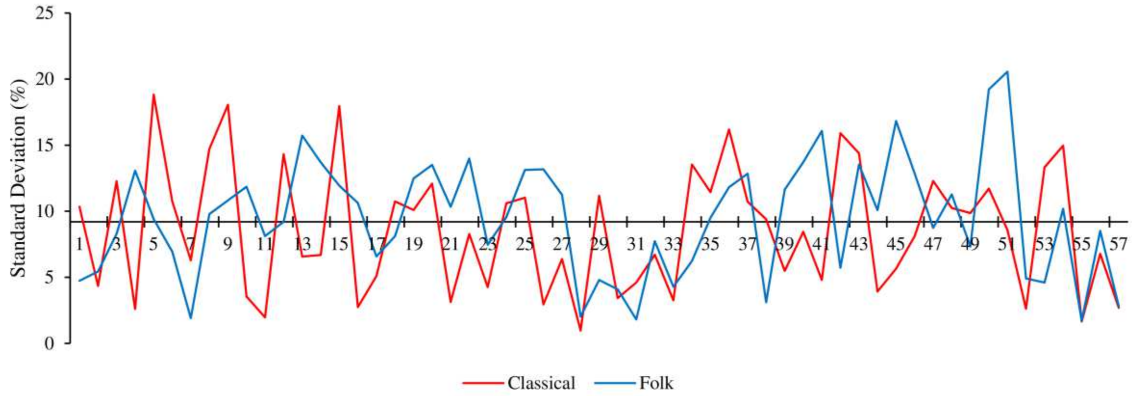


Figure 1. Average standard deviations of consonant lengths in classical and folk performances.

Moreover, each singer develops an idiosyncratic way of pronouncing (Figure 2) with significant differences between them ($F_{[4\text{-}54]} = 4.063$; $p = .003$). Also, the interaction between factors *singing style* (folk vs. classical), and *singer* was moderately significant ($F_{[4\text{-}8]} = 4{,}560$; $p = .001$). Similarly, the interaction between singing styles, singer and consonant was significant ($F_{[4\text{-}8]} = 3.820$; $p < .000$). These data support the high variability found in both singing styles.

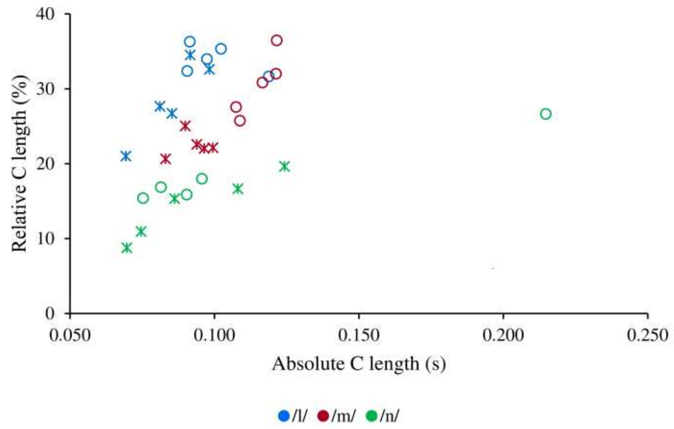


Figure 2. Average /l m n/ lengths in classical (asterisks) and folk (circles) performances.

Conclusion

The results show that in Spanish folk singing the consonants /l m n/ in CV syllables tend to be longer than in classical singing. However, although the imposed aesthetic canon seems to have an effect on the pronunciation of the classical professional' performances, the evidence suggests that the length of the consonants /l m n/ –in particular, its variability– is used in an expressive way in both singing styles. Future studies will be done in order to determine the incidence of that canon in training periods of singers, studying the length of the consonants /l m n/ in students' performances.

References

Alarcos Llorach, E. (1950). Fonología expresiva y poesía. *Revista de letras, XI*(3), 179–197. Universidad de Oviedo. Spain.

Carter, P. M., & Wolford, T. (2016). Cross-generational prosodic convergence in South Texas Spanish. *Spanish in Context, 13*(1), 29–52.

Guzmán, M. N. (2017). Pronunciación y Expresión. Un estudio preliminar sobre las posibilidades articulatorias de las consonantes del español como recursos musicales expresivos. *Epistemus. Revista de estudios en Música, Cognición y Cultura, 5*(2), 9–26.

Guzmán, M. N., Shifres, F., & Carranza, R. (September 2017). Pronunciación en el canto en español y aisthesis decolonial. In *V Jornadas Internacionales de Fonética y Fonología*. Universidad Nacional de La Plata. Argentina.

Mahaney, C.L. (2006). *Diction for Singers: A comprehensive assessment of books and sources* [PhD thesis]. The Ohio State University. USA.

Mendoza, E., Carballo, G., Cruz, A., Fresneda, M. D., Muñoz, J., & Marrero, V. (2003). Temporal variability in speech segments of Spanish: Context and speaker related differences. *Speech Communication, 40*(4), 431–447.

Miller, R. (1996). *On the Art of Singing*. New York: Oxford University Press.

Posadas de Julián, P. (2008). Los fonemas como recurso expresivo en el canto lírico. *Language Design: Journal of theoretical and experimental linguistics,* 107–118.

Scherer, K. R., Sundberg, J., Tamarit, L., & Salomão, G. L. (2015). Comparing the acoustic expression of emotion in the speaking and the singing voice. *Computer Speech & Language, 29*(1), 218–235.

Shifres, F. (1994). La Ejecución Musical en Términos Interpretativos. Reflexiones en torno al estado actual de su enseñanza en instituciones especializadas de nuestro país. In *IV Encuentro de Educadores Musicales del Interior del País*. Collegium. Argentina.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

The influence of music-based interventions on the cognitive abilities of people with dementia as measured by the MMSE

Sylvia I. Haering

Department of Educational Sciences, University Roma Tre, Italy

sylvia-ingeborg.haering@uniroma3.it

Abstract

The purpose of this work is to examine the effect of music interventions on cognitive abilities in patients with dementia, and to find differences in studies with significant and nonsignificant outcomes on the Mini-mental State Examination MMSE (Folstein, Folstein, & McHugh, 1975), which is a test used frequently to measure the cognitive abilities in individuals with dementia. Based on the observation, that the effects of music-based interventions – including instrumental lessons, music therapy, music programs, and listening to music – measured by the MMSE vary immensely in their outcomes, a literature review is conducted to find cues toward underlying factors that might influence the test results. A literature review is conducted on the databases Cochrane, DIMDI, PubMed, Science Direct, Open Grey and Google Scholar, using the keywords "Dementia," "Alzheimer's," "Music Therapy," "Cognitive," and "MMSE" as well as their German translations. Eighteen studies which use music-based interventions, work with a population of elderly with dementia and apply the MMSE were included. A connection between the MMSE outcome and the frequency of interventions could be found: significant results are most commonly reported in studies with one intervention per week, whereas non-significant results are only found in studies with 2 to 3 interventions per week. The results might help understand how music can help delay the decline of cognitive skills throughout the neurodegenerative process of dementia.

Introduction

Music therapy is gaining more and more importance as a treatment method for dementia. Recent research shows potential benefits in reducing anxiety, depression and agitated behaviour displayed by elderly people with dementia and a general improvement in quality of life. The improvement of cognitive functioning is particularly interesting, since cognitive decline is one of the leading characteristics of dementia (McDermott et al., 2013; Blackburn & Bradshaw, 2014). It can be measured using the Mini-Mental State Examination MMSE (Folstein, Folstein, & McHugh, 1975), a test frequently used to measure the cognitive abilities in individuals with dementia. Results measured by the MMSE when conducting music-based interventions – like instrumental lessons, music therapy, music programs, and listening to music – vary immensely in their outcomes. Therefore, a literature review is conducted to find cues toward underlying factors that might influence the test results. Consequently, the purpose of this study is to examine the effect of music interventions on cognitive abilities in patients with dementia and to find differences in studies with significant and nonsignificant outcomes on the MMSE.

Background

Dementia is a psychopathological syndrome arising from an acquired dysfunction of memory functions and at least one other cognitive impairment in either thinking, orientation, perception, arithmetic, learning, language or judgment. Additionally, changes in emotional control, social behavior and motivation are common (Jahn & Wehrheid, 2015). There are several types of dementia whose symptoms can differ a lot. In Alzheimer's Disease, for example, weaknesses in memory functions are common. Vascular dementia is caused by several small, territorial ischemic insults. Depending on the location of the vascular events, their effects on cognitive abilities can differ. Parkinson's symptoms with rigor or tremor in combination with problems in attention, vigilance and memory are common in dementia with Levy Bodies, but the severity of symptoms can vary considerably during the course of the disease (Prosiegel & Böttger, 2007). Frontotemporal dementia can lead to language impairment and/or behavioral change (Jahn & Werheid, 2015). Dementia is therefore not a single disease, but an umbrella term for several types of neurodegenerative diseases.

There are several kinds of music-based interventions that use the element of music in different ways. *Music therapy* is among the best-known music-based approaches. It is a therapeutical approach in which music is used to achieve a non-musical goal, like enhancing language skills after a stroke (Thaut, Hoemberg, & Von Wild, 2014). Music therapeutical techniques can follow either an active or receptive approach. In *active music therapy* the patient participates for example by singing or playing instruments. It is primarily a productive and creative process. *Receptive music therapy* on the other hand focusses on the perception of music. The patient listens to the music which is provided by the therapist. *Music therapeutic listening interventions* are a subtype of receptive music therapy, in which the listening intervention is guided by the therapist. This includes reflecting over the music and the connected emotional processes. In contrast, *listening interventions* do not include the involvement of a therapist. In contrast, the aim of *instrumental training* (or *music education*) is to teach musical, instrumental or vocal skills. Its purpose is to master the art of playing an instrument, while the objective in music therapy lies within the improvement or sustain of non-musical goals. A *music program* is a music intervention carried out by musicians. Examples are a moderated concert or a singing circle. The musicians do not take on a therapeutic or teaching role but commit rather to entertaining. The type of music intervention does therefore influence the way cognition is targeted, which might have influences on the results obtained.

The term cognition describes all kinds of manifestations connected to thinking and knowledge. Cognition consists of both content and processes. The content includes facts, rules and memory, while processes are procedures in charge of processing information. Intelligence, language, thinking, problem solving, memory, attention and perception are all fields of cognitive psychology (Gerrig, 2015).

The Mini-Mental State Examination (MMSE) is a short test used to determine cognitive performance losses in patients with dementia. The MMSE assesses cognitive performance in only 5-10-minutes and consists of two parts. The first part measures the parameters orientation, attention and memory based on the patents' oral answers. The second part evaluates the ability to name objects, both in verbal and in written form. The patients are tested for their ability to follow instructions, to write a sentence and to create a drawing. The maximum points achievable on the MMSE are 30 (Folstein, Folstein & McHugh, 1975). Depending on the number of points, the degree of dementia is estimated. 27 to 30 points equal no cognitive impairment. 20 to 26 points are considered mild dementia. Moderate dementia starts at 10 and goes up to 19 points. At 9 points or less, severe dementia is assumed (Stechl et al., 2012).

Methods

Literature research

A literature review is conducted on the databases Cochrane, DIMDI, PubMed, Science Direct, Open Grey and Google Scholar, using the keywords "Dementia," "Alzheimer's," "Music Therapy," "Cognitive," and "MMSE" as well as their German translations. Eighteen studies (CTs, RCTs, exploratory studies and quasi-experimental trials) which use music-based interventions, work with a population of elderly with dementia and apply the MMSE, were included. The results of the studies rage from "significant improvement," "nonsignificant tendencies," "no significant results," to "significant decline" on the MMSE. These results are then compared with other aspects, like the type of music intervention, the severity of dementia, the musical material, and the duration/frequency of the interventions.

The studies included must meet the following criteria:

- A music-based intervention must be applied.
- Cognition must be measured by the MMSE. It can either be administered as the main or as a secondary measurement.
- The participants in the studies are elderly with dementia. There are no restrictions regarding the type and severity of dementia.
- The studies must have been published in either German or English.
- Articles, doctoral theses and master's theses are included, if they meet the criteria mentioned above.

The eighteen studies included in the review are listed in table 1. The most common intervention type is music therapy (12). Two studies use instrumental lessons, three listening interventions and two utilize a music program. One study compares a listening intervention to instrumental lessons. Therefore, the total number of music-based interventions results in nineteen. Nine studies apply a randomized controlled trial (RCT) and six studies use a controlled trial (CT).

Table 1. Classification of the studies according to study design and intervention types

Study	Study design	Intervention types
Arroyo-Annló, Díaz & Gil (2013)	RCT	LI
Brotons & Koger (2000)	RCT	Active MT; conversation
Bruer, Spitznagel & Cloninger (2007)	RCT	Active MT; watching movies
Camic, Williams & Meeten (2014)	MM	MP
Choi et al. (2009)	CT	Active MT; NA
Cooke et al. (2010)	RCT	MP
Fischer-Terworth (2010)	CT	Active MT; occupational therapy
Groene (1993)	RCT	MT; reading
Guétin et al. (2009)	RCT	LI; NA
Hartl (2010)	MM	MT; LI
Hong & Choi (2011)	RCT	Active MT; NA
Landsiedel (2003)	CT	Active MT; reminiscence
Li et al. (2015)	QET	LI; NA
Liesk, Hartogh & Kalbe (2015)	RCT	ME; cognitive stimulation
Raglio et al. (2008)	CT	Active MT; entertainment
Särkämö et a. (2013, 2016)	RCT	mtLI; ME
Smith (1986)	CT	Active MT; reminiscence; singing
Suzuki et al. (2007)	CT	Active MT; NA

RCT = Randomized Controlled Trial; CT = Controlled Trial; MM = Mixed Methods; QET = Quasi-experimental Trial; mtLI = music therapeutical listening intervention; LI = listening intervention; MP = music program; ME = music education; MT = music therapy; NA = normal activity

Results

Among the studies included in this work six interventions show "significant improvement" on the MMSE, while six others do not indicate any improvement and one even measures a significant decline in cognitive ability. Six studies fall under the category "non-significant tendencies" and show light improvement in cognitive skills, which are however too small to be considered "significant improvement". It seems substantial to include this category, because a small improvement might already be valuable in a population predisposed to cognitive decline. In some cases, significant improvements could be documented on an individual level, that do not appear when the average is calculated (Liesk, Harthogh, & Kalbe, 2015). In general, the MMSE results fall equally under all four result intervention categories (table 2).

Table 2. Relationship between MMSE and music intervention

Results / Inter-vention	significant improvement (6)	non-significant tendencies (6)	no significant results (6)	significant decline (1)
Music therapy (12)	4	4	4	-
Instrumental lessons (2)	1	-	1	-
Listening to music (3)	1	2	-	-
Music program (2)	-	-	1	1

There is a slight tendency towards the use of *bibliographic songs* in the studies with significant improvement (table 3). *Bibliographic songs* are the pieces of music that are important in the biography of the recipient, like a song played during a wedding, or favorite music during teenage years. *Preferred songs* are songs that have been chosen – usually by the participant or their caretakers – out of a list of preselected music. *Classic music* refers to mostly instrumental music from historic periods in Europe. In two cases the kind of music applied is not specified.

The division into *preferred* and *biographic songs* might be a bit vague because on an individual level they can coincide: *Biographic songs* might be also *preferred songs*. However, not all *preferred songs* are *biographic* as well.

Table 3. Relationship between MMSE and musical material

Results (MMSE):	**significant improvement (6)**	**non-significant tendencies (6)**	**no significant results (6)**	**significant decline (1)**
Musical material:	Preferred songs (2) Biographic songs (4) Classic music (0) Not specified (0)	Preferred songs (3) Biographic songs (1) Classic music (1) Not specified (1)	Preferred songs (2) Biographic songs (2) Classic music (0) Not specified (2)	Preferred songs (1) Biographic songs (0) Classic music (0) Not specified (0)

When comparing the MMSE results to the severity of dementia, the outcomes are difficult to compare. Many studies work on rather broad information: seven studies target mild to moderate, three on moderate to severe, and two on mild to severe cases (table 4). It is therefore difficult to evaluate the significance of the severity of dementia on the study outcome. Future studies should possibly work on only one stage of dementia, and preferably provide data regarding the type of dementia treated as well.

Table 4. Relationship between MMSE and severity of dementia

	significant improvement (6)	**non-significant tendencies (6)**	**no significant results (6)**	**significant decline (1)**
Mild dementia (1)	-	1	-	-
Mild to moderate dementia (7)	3	2	2	-
Mild to severe (2)	-	-	1	1
Moderate dementia (5)	1	2	2	-
Moderate to severe dementia (3)	1	1	1	-
Not specified (1)	1	-	-	-

By comparing the MMSE results to the frequency of interventions, a trend can be found: most "significant improvements" have been documented when music interventions were applied once per week, while the most results without significant outcome have been reported when interventions were applied 2 or 3 times a week (table 5). However, because of methodical weaknesses within some of the studies, the significance of the results must be discussed. The results suggest that the frequency of interventions might have a major effect on the cognitive abilities as measured by the MMSE and in consequence on the benefit that people with dementia might receive from music therapy and other music interventions. Tendencies in the impact of different types of music can be carefully concluded. However, further research is needed.

Table 5. Relationship between MMSE and frequency of interventions

	significant improvement (6)	**non-significant tendencies (6)**	**no significant results (6)**	**significant decline (1)**
Daily (3)	1	2	-	-
4-6 x/week (0)	-	-	-	-
2-3 x/week (7)	-	1	6	-
1 x/week (9)	5	3	-	1

Discussion

Studies with significant improvement show a slight tendency towards the use of *bibliographic songs*. However, the categorization into *preferred* and *biographic songs* might be a bit vague because they might coincide on the individual level: *biographic music* can be *preferred music*, too.

Considering that active music-making seems to have an influence on various areas of executive function – like working memory (Haering, 2018) – and the functionality of working memory influences the severity of symptoms in dementia (Boyle at al., 2008), active participation in music-making might train executive function in people with dementia helping the individual to compensate the cognitive decline to a certain extend. Also, it is possible that active and receptive techniques tackle cognitive impairment in different ways. Receptive approaches might help patients to relax, which could ease challenging situations and lead to better cognitive functioning. Active approaches on the other hand could contribute to training brain regions correlated with cognitive skills. Therefore, active and receptive approaches might lead to similar results on tests, while their working mechanisms are different.

Due to a lack of information, the types of dementia targeted in the 18 studies could not be included into the study. This might be problematic since diseases on the dementia spectrum can vary significantly in appearance and progression, as shown in the meta-analysis by Zakzanis, Leach and Kaplan (1999) that compares the performance domains of Alzheimer's patients to those of patients with FTD and PPA. Moreover, the evaluation of the severity of symptoms in frontotemporal dementia is difficult because patients tend to have unremarkable findings in short cognitive tests such as MMSE (Gregory & Hodges, 1996). Therefore, the test results might vary according to type of dementia. Furthermore, the work of Särkämö et al. (2016) indicates that the effects of musical interventions might differ depending on severity and type of dementia. Consequentially, it might be useful for future studies to include information on dementia types.

The MMSE is only a screening instrument and does not provide enough details to offer a complete picture of cognitive abilities in a person. Differences within the MMSE results might be partly caused by this fact. To investigate the effect of music interventions on cognition in dementia, it might be useful to either use a broader testing battery in future studies. And further research on the effects of different music types and frequency of interventions is needed.

Conclusion

The results of this work might help understand how music therapy can maintain cognitive abilities or delay the decline of cognitive skills throughout the neurodegenerative process of dementia. The results of the analysis suggest that there might be a correlation between frequency of music-based intervention and MMSE outcome. However, there are methodical weaknesses within the studies, therefore further research is required.

On the long run, the outcome of this review might contribute to improving the quality of music-based interventions for people with dementia. It may be useful especially for music therapists in order to evaluate the impact of different intervention styles, and to consequently pick the one that fits the needs of the patients best.

References

Arroyo-Anlló, E., Díaz, J., & Gil, R. (2013). Familiar Music as an Enhancer of Self-Consciousness in Patients with Alzheimer's Disease. *BioMed Research International*, 1-10.

Blackburn, R., & Bradshaw, T. (2014). Music therapy for service users with dementia: A critical review of the literature. *Journal of Psychiatric and Mental Health Nursing, 21(10)*, 879-888.

Boyle, P., et al. (2008). Processing resources reduce the effect of Alzheimer pathology on other cognitive systems. *Neurology, 70(17)*, 1534-1542.

Brotons, M., & Koger, S. (2000). The impact of music therapy on language functioning in dementia. *Journal of Music Therapy, 37*, 183-195.

Bruer, R., Spitznagel, E., & Cloninger, C. (2007). The Temporal Limits of Cognitive Change from Music Therapy in Elderly Persons with Dementia or Dementia-Like Cognitive Impairment: A Randomized Controlled Trial. *Journal of Music Therapy, 44(4)*, 308-328.

Camic, P., Williams, C., & Meeten, F. (2014). Does a 'singing together group' improve the quality of life of people with a dementia and their carers? A pilot evaluation study. *Dementia, 12(2)*, 157-176.

Choi, A., et al. (2009). Effects of Group Music Intervention on Behavioral and Psychological Symptoms in Patients with Dementia: a pilot-controlled trial. *International Journal of Neuroscience, 119*, 471-481.

Cooke, M., et al. (2010). A randomized controlled trial exploring the effect of music on agitated behaviours and anxiety in older people with dementia. *Aging and Mental Health, 14(8)*, 905-916.

Fischer-Terworth, C. (2010). Evaluation einer TEACCH-basierten psychologischen Intervention bei leichter bis mittlerer Demenz: Eine kontrollierte Studie [Evaluation of a TEACCH-based psychologiscal intervention in mild to moderate Dementia: a controlled study]. (Unpublished doctoral dissertation). University of Hamburg, Hamburg, Germany.

Folstein, M., Folstein, S., & McHugh, P. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12(3)*, 189-198.

Fusar-Poli, L., et al. (2017). The effect of music therapy on cognitive functions in patients with dementia: A systematic review and meta-analysis. *Aging & Mental Health*, 1-10.

Gerrig, R. (2015). *Psychologie* [Psychology], Hallbergmoos: Pearson.

Gregory, C., & Hodges J. (1996). Clinical features of frontal lobe dementia in comparison to Alzheimer's disease. *Journal of Neural Transmission (Supplementum), 47*, 103-123.

Groene, R. (1993). Effectiveness of music therapy 1:1 intervention with individuals having senile dementia of the Alzheimer's type. *Journal of Music Therapy, 30(3)*, 138-157.

Guétin, S., et al. (2009). Effect of music therapy on anxiety and depression in patients with Alzheimer's type dementia: randomized, controlled study. *Dementia and Geriatric Cognitive Disorders, 28(1)*, 36-46.

Haering, S. (2018). Are music-based interventions a way to improve executive function in autism? *International Journal of Learning and Teaching, 10(1)*, 27-37.

Hartl, L. (2010). Musiktherapie in der Behandlung von Demenz: leibphilosophische und psycho-physiologische Annäherungen [Music Therapy in the Treatment of Dementia: physical-philosophical and psycho-physiological approaches]. (Master's Thesis). Institut für Ethnomusiktherapie, Wien, Austria.

Hong, I. S., & Choi, M. J. (2011). Songwriting oriented activities improve the cognitive functions of the aged with dementia. *The Arts in Psychotherapy, 38(4)*, 221-228.

Jahn, T., & Werheid, K. (2015). *Demenzen: Fortschritte der Neuropsychologie* [Dementia: Progress in Neuropsychology]. Göttingen: Hogrefe.

Landsiedel-Anders, S. (2003). Musiktherapie bei Demenzerkrankungen: eine klinisch-experimentelle Studie im Rahmen einer Gedächtnissprechstunde [Music Therapy for Dementia: a clinical-experimental study within a memory consultation intervention]. (Unpublished Master's Thesis) Johann-Wolfgang-Goethe-Universität, Frankfurt/Main, Germany.

Li, C., et al. (2015). Adjunct effect of music therapy on cognition in Alzheimer's disease in Taiwan: a pilot study. *Neuropsychiatric Disease and Treatment, 11*, 291-296.

Liesk, J., Hartogh, T., & Kalbe, E. (2015). Kognitive Stimulation und Musikintervention bei stationär versorgten Menschen mit Demenz [Cognitive Stimulation and Music Intervention in Inpatients with Dementia]. *Zeitschrift für Gerontologie und Geriatrie, 48(3)*, 275-281.

McDermott, O., Crellin, N., Ridder, H., & Orrell, M. (2013). Music therapy in dementia: A narrative synthesis systematic review. *International Journal of Geriatric Psychiatry, 28(8)*, 781-794.

Prosiegel M., & Böttger S. (2007). *Neurologische Störungen und ihre Rehabilitation* [Neurological Disorders and their Rehabilitation], München: Pflaum.

Raglio, A., et al. (2008). Efficacy of music therapy in the treatment of behavioral and psychiatric symptoms of dementia. *Alzheimer Disease and Associated Disorders, 22*, 158-62.

Särkämö, T., et al. (2013). Cognitive, Emotional, and Social Benefits of Regular Musical Activities in Early Dementia: randomized controlled study. *The Gerontologist, 54(4)*, 634-650.

Särkämö, T., et al. (2016). Clinical and Demographic Factors Associated with the Cognitive and Emotional Efficacy of Regular Musical Activities in Dementia. randomized controlled study. *Journal for Alzheimer Disease, 49(3)*, 767-781.

Smith, G. H. (1986). A comparison of the effects of three treatment interventions on cognitive functioning of Alzheimer patients. *Music therapy, 6(1)*, 41-56.

Stechl, E. (2012). *Praxishandbuch Demenz: Erkennen, Verstehen, Behandeln* [Practical Handbook of Dementia: Recognize, Understand, Treat]. Frankfurt/Main: Mabuse.

Suzuki, M., et al. (2007). Music therapy-induced changes in behavioral evaluations, and salvia chromogranin A and immunoglobulin A concentrations in elderly patients with senile dementia. *Geriatrics and Gerontology, 7*, 61-71.

Thaut, M., Hoemberg, V., & Von Wild, K. (2014). *Handbook of neurologic music therapy.* Oxford: Oxford University Press.

Zakzanis, K., Leach, L., & Kaplan, E. (1999). *Neuropsychological differential diagnosis*. Lisse: Swets & Zeitlinger.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Do Dotted Rhythms Increase Performance Precision: Why Marches Have Dotted Rhythms

Niels Chr. Hansen,[1] Nicholas J. Shea,[2] and David Huron[3]

[1,2,3]*Cognitive and Systematic Musicology Laboratory, School of Music, Ohio State University, USA*

[1]Ni3lsChrHansen@gmail.com, [2]shea.185@buckeyemail.osu.edu, [3]huron.1@osu.edu

Abstract

Among the most impressive human behaviors are closely synchronized actions involving large groups of people—as seen in military displays, synchronized swimming, and much dance. Since coordinated movement promotes cooperation among participants and increases the extent to which observers attribute group rapport and cohesion, there are good *a priori* reasons to aspire to such behaviors on an individual and group level. Interestingly, synchronized displays are often accompanied by music. This raises the question of which musical features might facilitate precise movement coordination.

Musicians recognize that beat subdivision (either imagined or actual) facilitates synchronization. Compared with isochronous rhythms exhibiting the same event density, dotted rhythms necessitate beat subdivision as a criterion for accurate performance. Accordingly, one might predict that dotted rhythms improve group synchronization.

A corpus analysis is reported whose purpose was to test the conjecture that dotted rhythms appear more often in music associated with group synchronization (specifically marches) than in other types of music. Two hundred marches were randomly sampled along with a matched sample of 200 control pieces written by the same composer, employing the same instrumental genre, and using the same metric class. The four pairs of notes preceding the first four sounded downbeats were examined. Surprisingly, the results indicate that dotted rhythms are not significantly more common in marches. Double-dotted rhythms were also not more common for slower than for faster tempos. If indeed dotted rhythms contribute to the impressiveness of march displays by facilitating synchronization, Westerns march music does not significantly seem to capitalize on this phenomenon.

Introduction[1]

On a visit to the Great Smoky Mountains, Paulo Urbano was struck by the impressive display of thousands of fireflies flashing in perfect synchrony (Urbano, 2018). The synchronous flashes of fireflies notwithstanding, the capacity to entrain to an isochronous beat is biologically rare (Merker, 2009). Most animals, including our nearest primate relatives have little or no ability to synchronize rhythmically (Patel, Iversen, Bregman, & Schulz, 2009).

The human aptitude for synchronization is evident in an enormous range of group behaviors. These include nearly all forms of dancing (from Western ballet to West-African *agbadza*), synchronized athletics (gymnastics, figure skating, synchronized swimming), marching bands, drum lines, and many other cultural expressions, from Maori *hakas* and Japanese precision walking to cheer leading and flash mobs. Olympic opening ceremonies commonly feature several displays of precise synchronous human actions.

The worldwide web offers a cornucopia of videos featuring synchronized displays from changing-of-the-guard ceremonies to air force Thunderbirds aerobatics. Video titles and viewer commentaries commonly include adjectives such as “amazing,” “unbelievable,” “impressive,” or “awesome” suggesting that viewers experience synchronized human behavior as compelling. In short, not only do humans have an aptitude for synchronization—and a penchant to engage in synchronous behaviors—but in most cases, we find such displays engaging or compelling. Research also shows that coordinated movement promotes cooperation among participants (Valdesolo, Ouyang & DeSteno, 2010) and increases the extent to which observers attribute group rapport and cohesion (Lakens & Stel, 2011).

Historically, there is a strong link between synchronous action and military displays. In particular, the seemingly simple phenomenon of coordinated walking (i.e., marching) is widespread across many cultures and over a very long history. Since the early twentieth century, trains, motorized vehicles and aircraft have significantly diminished the need for military transport by foot. Nevertheless, armies all over the world continue to drill soldiers in the art of synchronous marching. If anything, marching has become more precise over the past century (Monelle, 2006). This suggests that the value of marching has little to do with transportation. Within military leaderships, marching is widely thought to promote a sense of unity and self-disinterest among soldiers, while impressing or even intimidating observers (McNeil, 1997).

There are, of course, degrees of synchronization, and highly precise synchronization is prized more than loose synchronization. For example, the International Swimming Federation provides explicit criteria for judging synchronized swimming routines. The criterion for scoring a perfect 10 specifies: “Totally synchronised with the music and each other. Absolute precision throughout.” By contrast, the lowest score is assigned when there is “Very little or no attempt to synchronise with music or each other” (Fédération Internationale de la Natation, 2015, pp. 96, 102). For a wide range of artistic, athletic, and other displays, the result is experienced as more or less impressive in direct proportion to the degree of synchrony.

In the case of music, for many (though not all) musical genres, musicians similarly value precise temporal coordination. Research on finger tapping has established a number of factors that influence beat precision (Repp, 2005). An important factor is tempo. The greatest accuracy in beat synchronization is known to occur at an inter-onset interval of around 600 ms (Fraisse, 1982). As the tempo slows,

[1] Please note that the analysis reported here was expanded after the submission deadline in ways that modify the results themselves as well as change the interpretation of our findings. We encourage the reader to seek out future peer-reviewed publications resulting from this work.

synchronization degrades noticeably. Bamberger (1978) noted that poor synchronization is associated with a switch from "clock" (beat-based) coding to "figural" (duration-based) coding.

Another factor known to influence the precision of synchronization is the cognitive strategy of mentally subdividing the beat. Especially when performing at a slow tempo, beat subdivision can significantly improve synchronization. Such mental subdivision is especially important as a crucial moment of synchronization approaches. For example, in a 4/4 meter, a complete measure of rest prior to a critical downbeat might cause a musician to mentally count each of the first three quarter durations and then switch to counting the sixteenths in the final quarter ("1 ..., 2 ..., 3 ..., 1234") prior to the ensuing downbeat. This same procedure underlies so-called drum-fills, which also likely improves synchronization. The importance of temporal preparation is evident in conducting behaviors. When beginning a musical work, a preparatory or up-beat gesture is essential as a way of increasing the uniformity of attack for an ensemble. In short, musicians recognize that a preparatory moment prior to a moment of synchronization tends to have an especially facilitating effect on the precision of synchronization.

Rather than leaving this strategy to silent imagination, another way to encourage such preparatory beat subdivision is to make the preparatory moment part of the music. A single preparatory event is likely to lead to a dotted or double-dotted rhythm. (Later, we will refine this notion.)

In topic theory, the presence of dotted rhythms has long been proposed as a characteristic feature of the "march" or "military" topoi (Hatten, 2004; and others). For example, in describing the "march" topic, Leonard Ratner (1980) explicitly included the dotted rhythm as a defining feature: "its natural habitats were the parade ground and battlefield, where its moderately quick duple meter, dotted rhythms, and bold manner quickened the spirit" (p.16).

Motivation

The motivation for the current study is the following suggestion. We conjecture that one of the main motivations for marching is to impress observers through the synchronous movement of large numbers of people; that such synchronization is facilitated by music with a predictable beat, that dotted rhythms in particular facilitate precise synchronization, and consequently dotted rhythms have become a characteristic feature of the march genre or topic.

In the current study we do not propose to test the entire theory. For example, we do not propose to examine evidence whether marching is intended to impress observers. In addition, despite the anecdotal evidence that people are impressed by synchronous movement, we have found no formal test of this idea—nor do we intend to test this idea in the current study. Instead, we propose to test a single facet of our interpretive story. Since dotted rhythms are commonplace in music, we will test the assumption that marches favor dotted rhythms.

Corpus Analysis

As noted, various writers have informally drawn attention to the tendency for marches to exhibit dotted rhythms. Our proposed account critically relies on this intuitive observation. Accordingly, we propose to conduct a formal corpus study that tests the following hypothesis:

> ***H1.*** *In march-like works the second of the two notes preceding a sounded downbeat will tend to be shorter than the first of the pair.*

With regard to Hypothesis 1, for the purposes of this study we will refer to the target rhythm as a *cretic rhythm.* (In Greek poetic theory, a cretic rhythm is a long-short-long pattern.) We will define a cretic rhythm as a three-note pattern where the final note coincides with a sounded downbeat (specifically, the beginning of a measure), and the two preceding notes exhibit a long-short durational pattern. In defining a duration, ensuing rests are deemed to belong to and extend the duration of the preceding note onset. Hence an eighth rest following after a quarter note will be deemed to transform the quarter note into a dotted quarter duration. Examples of cretic and non-cretic rhythms are illustrated in Figure 1.

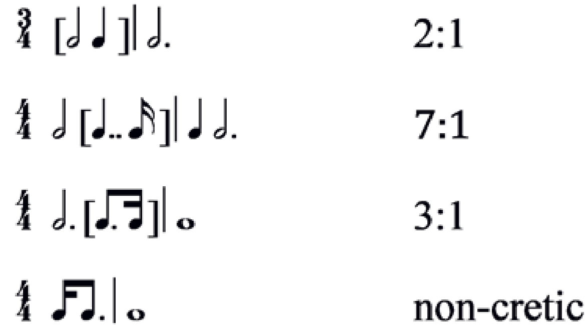


Figure 1. Examples of cretic and non-cretic rhythms.

If dotted rhythms encourage performers to subdivide the beat in ways that facilitate rhythmic synchronization, then we might expect that the degree of subdivision will interact with the tempo of the music. Consequently, we might expect that slower marches will be more likely to employ double-dotted rhythms than faster marches. Accordingly, we might propose a second hypothesis:

> ***H2.*** *For march-like works, the ratio of the durations of the first two notes of cretic rhythms will be larger for slower tempos than for faster tempos.*

Method

In brief, our corpus study involves tallying the number of cretic rhythms in a representative sample of march-like works and in matched control pieces.

Sample

In order to test our first hypothesis, we need to contrast occurrences of cretic rhythms in target march works with control works. For this study we made use of the International Music Scores Library Project (IMSLP). A search for the keyword "march" produced a list of 1,622 scores written by 779 unique composers that are categorized as marches. Examples include the *Children's March* by Percy Grainger, *March in D Major* by G.F. Handel, and *Triumph of Time* by John Philip Sousa. Not all titles in this list include the word "march," although they are all categorized by IMSLP as marches. Consequently, we will refer to these works as "march-like."

We randomly ordered this long list and selected the first 200 marches for which we were able to identify a matching control work, taking care to exclude anonymous works (see below). Matching control works must be by the same composer, but may not contain the word "march" (or non-English equivalent, e.g. *marcia*) in the title, and must also not be classified as "march" in the IMSLP "work type" designation. In addition, control works were selected from the same IMSLP "instrumentation" designation, such as "for piano," "for piano four hands," "for orchestra," "for orchestra with vocalist," etc. An additional sampling constraint is that the control work must also be matched for meter classification. That is, if the target (march) work has a simple-duple classification (e.g., 2/4, 2/8, 2/2), the control work must also exhibit the same classification. By way of summary, control works were non-march works by the same composer, classified as the same instrumental genre, and exhibiting the same metric category.

Notice that Hypothesis 2 does not require matching control works. Consequently, we were able to include 94 further encoded marches for which no control was found. Hypothesis 2 does require that we code tempo for the target marches. Our method of tempo coding involves rather strict criteria (see below). Many of the marches included no suitable tempo indication; of the 294 sampled marches, 151 provided useable tempo information for testing our second hypothesis.

A final sampling issue is that many works begin with an introduction with the main exposition or melody appearing only after several measures. It is not uncommon for introductions to exhibit different tempos or meter signatures from the main expository material. In order to minimize the possibility of experimenter bias, we elected to avoid conducting analyses intended to determine the main expository passages. Instead, we simply resolved to code material at the outset of both target and control works.

Data Coding

Having identified suitable target and control samples, we analyzed the works as follows. We identified the first four sounded downbeats that were each preceded by two notes in the preceding measure. Downbeats were simply defined as notes initiating a measure. Notice that the first measure in a work beginning with a sounded downbeat would fail to have two preceding notes, and so the first candidate downbeat might coincide with the onset of the second measure. Referring back to Figure 1, examples of target cretic rhythms are indicated using square brackets. For each collection of four sounded downbeats, we tallied the total number of cretic rhythms (which necessarily will range between 0 and 4).

In addition to counting the number of cretic rhythms, we coded the durations of the two notes preceding each of potentially four sounded downbeats. This was done for both target and control passages. From these durations we calculated the ratio of the durations of the first and second notes. Hence, for example, a dotted-eighth/sixteenth pattern (preceding the downbeat) would result in a ratio of 3:1. Once again, appropriate ratios are shown in Figure 1 following the square brackets identifying cretic rhythms.

Finally, we also coded whether the work begins with a downbeat or employs some sort of pickup gesture. This information was collected, not in order to test an *a priori* hypothesis, but to provide an opportunity to conduct a *posthoc* test of a possible anticipated confound that will be addressed in the Results section.

In coding these values, an essential question was which musical part or parts were selected for rhythmic encoding. Few marches are written for a single monophonic instrument. Several possibilities arise. One might choose to code the highest musical instrument (more likely to contain the melody), the lowest musical instrument (more likely to contain the bass line), the first instrument to enter, or the aggregate rhythm from all instruments sounding at the beginning. One might also endeavor to select that instrument which could be considered to convey the principal musical line or melody. With the exception of homorhythmic *tutti* passages, none of the above coding strategies is without issues.

In light of these considerations we established, *a priori*, the following procedure:

(1) We identified the first four sounded downbeats that were preceded by a minimum of two sounded events in the same part.

(2) For multi-part works (including polyphonic keyboard textures), we then identified the instrument or part positioned highest in the score system that coincided with the candidate downbeat.

(3) We then determined whether the identified instrument or instrument-part also exhibited two or more sounded notes in the preceding measure. If not, the next lower instrument was identified and steps (2) and (3) were repeated recursively. If this recursive chain failed to lead to an encodable condition, then the candidate downbeat was discarded and replaced by the next candidate downbeat.

(4) Having identified a candidate downbeat that fulfilled the above conditions, the rhythmic sequence was encoded to include the duration of the two notes preceding the downbeat, the tempo, meter signature, and whether the work started with a downbeat or with a pickup gesture (see below).

Note that grace notes were treated as a pickup gesture if occurring at the beginning of a piece, but were ignored when encoding durations for the note-pairs preceding downbeats.

Tempo. As noted earlier, Hypothesis 2 requires that we code the tempi of the target marches. When coding tempo, a number of considerations arise. Only some works include explicit metronome markings, and moreover, metronome markings do not necessarily provide a good indication of the speed or pace of a work. In coding tempo, we made use of an operationalization used by Horn and Huron (2015), as modified by Hansen and Huron (2018). This operationalized method relies on an existing Wikipedia table of tempo terms ("tempo") ordered from slow to fast (accessed on January 3, 2017). In the Wikipedia entry, each term is accompanied by a corresponding range of metronome markings. In Table 1, each of the ranges has been distilled to a single average metronome value. Musicians would view these values with considerable skepticism, nevertheless, they provide useful though crude estimates. In effect, Table 1 provides an ordinal (rank-ordered) tempo scale for 20 common Italian tempo terms.

Two methods were used to code the tempos for sampled works: one based on Italian tempo terms, and a second method based on metronome indications (when present). Coding of Italian tempo terms relied on the terms shown in Table 1. In many cases, the Italian terms included tempo adjectives or modifiers such as *Allegro assai* (very much), or *Allegro ma non* *troppo* (not too much). In order to avoid excluding large numbers of sampled works, we elected to include all tempo designations given in Table 1, while ignoring those modifiers not already present in the table. Hence, *Allegro assai* and *Allegro ma non troppo* would both be coded as equivalent to *Allegro*. Although this procedure discards subtle information that may be important for performers, it still retains the idea that a particular passage is relatively fast (slow, moderate, etc.) in overall tempo.

Table 1. Tempo terms and estimated average metronome value.

Rank	Tempo Term	Median BPM
1	*Larghissimo*	≤24
2	*Grave*	35
3	*Largo*	50
4	*Lento*	52.5
5	*Larghetto*	61.5
6	*Adagio*	71
7	*Adagietto*	74
8	*Andante*	92
9	*Andantino*	94
10	*Marcia Moderato**	84**
11	*Andante moderato*	102
12	*Moderato*	114
13	*Allegretto*	116
14	*Allegro moderato*	118
15	*Allegro*	144
16	*Vivace*	172
17	*Vivacissimo*	174
18	*Allegrissimo/Allegro vivace*	174
19	*Presto*	184
20	*Prestissimo*	≥220

* Excluded from our earlier papers. *Tempo di Marcia*, *Alla Marcia*, *March Tempo*, *Martial*, *Marchmässig*, and *Mouvement de marche* were all counted as belonging to this category.
**Due to inconsistency with rank order, this metronome marking was ignored.

A number of marches (n = 155) offered no tempo term coinciding with the list shown in Table 1. Of these 155 works, 13 reported metronome indications. Accordingly, these metronome indications were recoded according to the 20 ranked categories displayed in Table 1. For example, a metronome marking of quarter=88 beats per minute would be categorized as equivalent to *Andante*. We were unable to code tempo for 143 works. In testing Hypothesis 2 these were simply treated as missing data.

Results

First, we performed visual inspection of the differences between the use of cretic and non-cretic rhythms preceding downbeats in marches and control works. Figure 2 plots the mean number of cretic rhythms per passage for the sampled marches (open bar) compared with the control passages (shaded bar). Numbers above each bar report the total number of cretic rhythms for their respective repertoires. Figure 3 provides more detail about the types of rhythms preceding downbeats. Specifically, Figure 3 plots the total number of rhythms for the durational ratios that occurred five or more times in any of the two corpora: 1:8, 1:4, 1:3, 1:2, 1:1, 2:1, 3:1, and 7:1. Both cretic and non-cretic rhythms are included. Once again, open bars plot the results for the target marches whereas the shaded bars plot the results for the control works. Notice that 1:1 ratios are slightly more common for the control works indicating that the control works tend to employ slightly more isochronous rhythmic patterns when approaching a sounded downbeat. Conversely, the marches employ more 3:1 ratios, consistent with a greater use of dotted (cretic) rhythms when approaching a sounded downbeat.

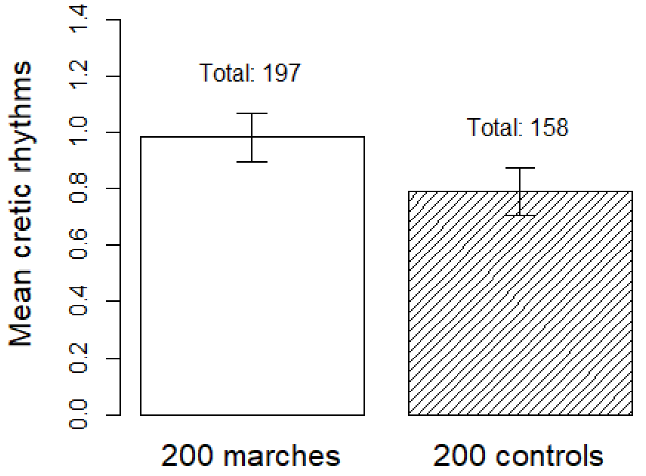


Figure 2. Mean and total number of cretic rhythms in the corpora of nominal marches and control works. Cretic rhythms designate rhythms where the two note durations preceding a downbeat follow a long-short pattern. Error bars depict the standard error of the mean.

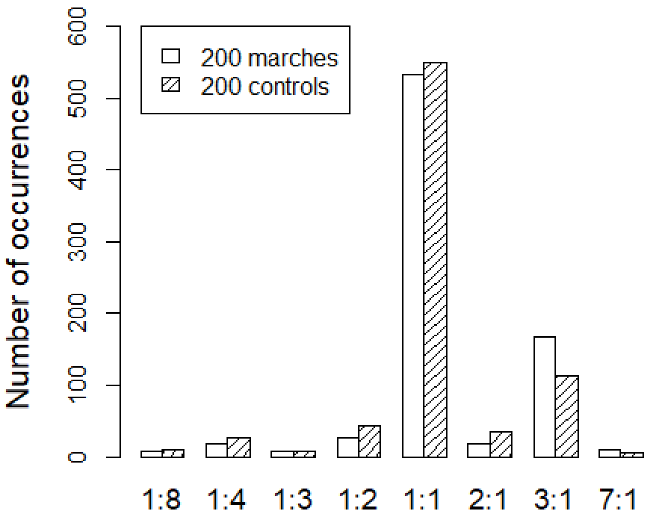


Figure 3. Number of each two-note durational patterns preceding downbeats in the corpora of nominal marches and control works. Bars to the right of 1:1 are cretic rhythms (i.e., 2:1, 3:1, and 7:1).

Next, we conducted inferential statistical analysis to test our main hypotheses. Due to non-normal distribution of the differences between marches and controls, a nonparametric test was applied. Although the absolute number of cretic rhythms was greater for the marches than for the non-marches (cf. Fig. 2), a Wilcoxon signed-ranks test with continuity correction (matching target and control passages, using the same metric type, and written by the same composer) revealed that the marches did not exhibit significantly more cretic rhythms compared with the control corpus (Z = -1.85, p = .064). These results are not consistent with Hypothesis 1.

In our study of the musical scores from march pieces we realized that some marches use pickup gestures other than dotted rhythms. For example, a dotted quarter note followed by an eighth note preceding a downbeat could easily be replaced by a dotted quarter note followed by four thirty-second notes. This pickup gesture could in principle enhance performance precision just like the underlying dotted (or cretic) rhythm that it was replacing. Therefore, we carried out an extra *posthoc* test of the conjecture giving rise to Hypothesis 1. Recall that we also recorded whether works belonging to the march and non-march categories begun with a pickup gesture or with a downbeat. From the sample of 200 marches and 200 matched control works, 64 marches and 61 non-marches began with a pickup gesture whereas 136 marches and 139 non-marches began with a downbeat. A chi-squared test with Yates' continuity correction showed that there was no significant difference between marches and non-marches as to whether they began with a downbeat or a pickup gesture, $\chi^2 = 0.047$, $df = 1$, p = .829. In other words, the fact that our initial results were not consistent with Hypothesis 1 could not simply be explained by the use of isochronous pickup gestures in marches.

Hypothesis 2 predicted that, in the case of cretic rhythms in marches, double-dotted rhythms should be observed more frequently in slower tempos than in faster tempos. Recall that for each target march, we have data for note-pairs preceding four downbeats. This means that up to four downbeats may be approached using cretic rhythms. In conducting any statistical analysis, we should aim to have a single independent observation for each work. Accordingly, whether a work contains one, two, three or four cretic rhythms, we need to distill these values into a single representative ratio.

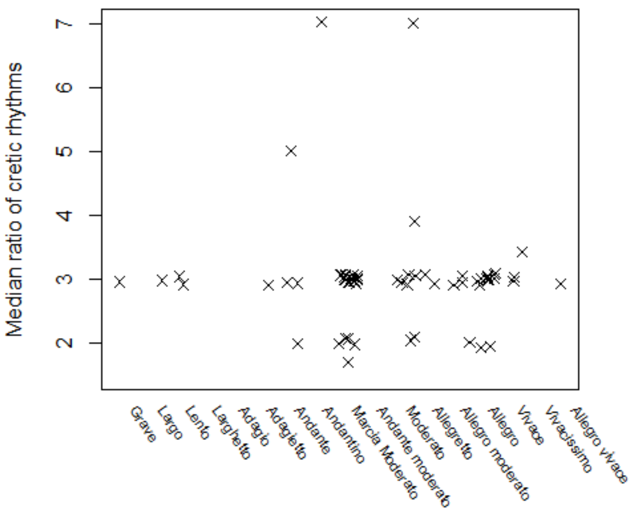


Figure 4. Scatterplot of tempo category and median rhythm proportion of cretic rhythms for each of the 65 marches that contained both tempo information and one or more cretic rhythms. Jitter has been added to make coinciding data points stand out more clearly.

Since ratios are not linear, the appropriate summary statistic is to calculate the median. Consequently, for each candidate march we calculated the median ratio for all cretic rhythms. Those marches that did not include any cretic rhythms were excluded from this analysis.

Note that two values are now available for each of the marches containing cretic rhythms as well as available tempo information: the rank-ordered tempo value and the median cretic ratio. Hypothesis 2 would predict a significant negative correlation–indicating that slower tempos are associated with higher median duration ratios for the cretic rhythms. Of the 294 marches, 65 marches provided both appropriate tempo data and included at least one cretic rhythm. Calculating Spearman's rank-order correlation, the results were not consistent with our second hypothesis, *rho* = .007, df = 63, p = .954). Figure 4 shows a scatterplot of the median ratio of cretic rhythms plotted against the nominal tempo of each of the 65 marches included in this analysis.

Conclusion

The motivation for this study was the informal observation that coordinated group movement is impressive to observers. In comparison with silence, any music with a steady beat might be expected to facilitate coordinated group actions. However, some musical tempos and rhythms might offer better opportunities for synchronization than others.

Musicians have long noted that subdivision of the beat (either imagined or performed) facilitates synchronization. For the same event density (note onsets per second), one would predict that isochronous rhythms would be less effective in promoting synchronization than cretic rhythms (such as dotted rhythms) since the latter rhythms effectively force performers to subdivide the beat. This intuition notwithstanding, it would be appropriate to conduct a synchronization experiment to test whether dotted rhythms do indeed facilitate synchronization compared with isochronous rhythms of equal event density.

In this preliminary study we have tested two hypotheses that link our proposed account to actual musical practice. Specifically, we have tested whether a genre of music associated with group synchronization (namely marches) exhibit the presumed preference for dotted (cretic) rhythms. Surprisingly, our results were not consistent with our first hypothesis. Cretic rhythms were not significantly more common in marches than in matched control pieces. The absence of a significant effect despite our sample size of 200 marches and 200 matched control pieces by the same composer using the same metric type suggests that informal analytic descriptions found in the musicological literature do not necessarily hold true. A further *posthoc* analysis allowed us to conclude that the lacking effect could not simply be ascribed to non-cretic pickup gestures using isochronous note durations.

Our results were also not consistent with the second hypothesis that slower tempos are associated with more durational contrast between the first and second notes in cretic approaches to downbeats. Our study, in other words, produced no evidence that composers strive towards facilitating more refined beat subdivisions for slower tempos.

In conclusion, while the perceptual-motor-social theory proposed here seemed to offer a coherent explanation for the musical practice of employing dotted rhythms in marches, our empirical results failed to support this theory. Therefore, the conjectured propensity for marches to employ dotted rhythms which has been noted by historical musicologists does not seem to manifest on a quantitative scale and is likely to have a cultural rather than a perceptual-motor origin.

Acknowledgements. We would like to thank Adam Verbsky for his invaluable contribution to data coding for the research reported here. Although we deem that this contribution qualifies for authorship, editorial procedures did not allow us to include Adam Verbsky as a co-author.

References

Bamberger, J. (1978). Intuitive and formal musical knowing: parables of cognitive dissonance. In S. S. Madeja (Ed.), *The Arts, Cognition and Basic Skills*. New Brunswick, N.J.: Transactions Books.

Fédération Internationale de la Natation (2015). *FINA Synchronised Swimming Manual For Judges, Coaches & Referees 2013-2017*. Lausanne, Switzerland: FINA. Retrieved from http://sportdocbox.com/Swimming/75027413-Fina-synchronised-swimming-manual-for-judges-coaches-referees.html

Fraisse, P. (1982). Rhythm and tempo. In D. Deustch (Ed.), *The Psychology of Music* (pp. 149-180). London, UK: Academic Press Inc.

Hansen, N. C., & Huron, D. (2018). The lone instrument: musical solos and sadness-related features. *Music Perception*, *35*(5), 540-560.

Hatten, R. S. (2004). *Interpreting musical gestures, topics, and tropes: Mozart, Beethoven, Schubert*. Bloomington, IN: Indiana University Press.

Horn, K., & Huron, D. (2015). On the Changing Use of the Major and Minor Modes 1750-1900. *Music Theory Online*, *21*(1).

McNeill, W. H. (1997). *Keeping together in time.* Cambridge, MA: Harvard University Press.

Merker, B. (2009). On the role and origin of isochrony in human rhythmic entrainment. *Cortex*, *45*(1): 4-17.

Monelle, R. (2006). *The musical topic: hunt, military and pastoral.* Bloomington, Indiana: Indiana University Press.

Patel, A. D., Iversen, J. R., Bregman, M. R. & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, 19(10): 827-830.

Ratner, L. (1980). *Classic Music: Expression, Form, and Style*. New York, NY: Schrimer.

Repp, R. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, 12(6): 969–992.

Urbano, P. (2018). The Light Show: Flashing fireflies gathering and flying over digital images. In *International Conference on Computational Intelligence in Music, Sound, Art and Design* (pp. 283-298). Springer, Cham.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Taking the Synchrony Out of Singing

Kathleen Harring, Laura Edelman, Alyssa Scartozzi, and Rosemary Corcoran

Department of Psychology, Muhlenberg College, USA

Kathyharring@muhlenberg.edu, Lauraedelman@muhlenberg.edu , as251063@muhlenberg.edu, rosemarycorcoran@muhlenberg.edu

Abstract

Previous research found that synchronous music and synchronous movement increased perceptions of bonding (Edelman & Harring, 2016). Using a 4-group design, the current study investigated the effects of different types of singing on social bonding. Three confederates participated in each condition along with one participant. Confederates and participants sang all in unison, by pairs in rounds, only one confederate in unison with the participants, and none in unison with the participants. At the end of the manipulation, the participant and confederates completed a questionnaire to measure entitativity, rapport, mood and manipulation checks. We found that the more asynchronous each condition became, the more socially bonded the participants felt to the confederate who performed in unison with them in the condition. However, in the asynchronous condition, the participant was not significantly bonded with any of the confederates. Results will be discussed in terms of factors that affect group cohesion and underlying neural mechanism that support social bonding.

Introduction

Previous research conducted by Livingstone and Thompson (2009) hypothesizes that the evolution of music is dependent on biological factors such as mirror neurons. In their article, Livingstone and Thompson argue that music has developed from a "Theory of Mind" general adaptation. This theory states that music is capable of simulating an emotional experience for its listeners by activating the brain's mirror neuron system. A significant impact of mirror neuron stimulation is its capability to promote feelings of group cohesion by enacting a system of empathy among the group members (Livingstone & Thomspon, 2009). Loersche & Arbuckle (2013) claim that the motive for human inclination toward music is due to the fact that music has evolved as a way of bringing groups together through the synchronization of mirror neurons.

Campbell (1958) called a group that comes together as cohesive. Group cohesion can be measured by the reported amount of two different variables: rapport (how much people understand each other's feelings) and entitativity (how much someone perceives a group as a unit). Four different group characteristics can promote one's perception of the group as a single entity; these characteristics are proximity, similarity, common fate and pragnance (meaningfulness). Essentially, these variables suggest that if the individuals participating in an experiment are near one another, have similarities, move together in the same direction with a similar end point, and form a meaningful pattern, they are likely to be perceived as a single unit.

In addition to promoting group cohesion by means of the four gestalt grouping principles, music has also been shown to be capable of promoting group cohesion through the creation of positive emotions within a group (Tarr, Launay, & Dunbar, 2014). It is evident that synchrony can create positive feelings within a group of people, which leads to social bonding. The claim that synchrony is capable of producing positive feelings is made based on research on mirror neurons, oxytocin and a series of other neural hormonal components. Tarr, Launay, & Dunbar (2014) suggest that the synchronization of music has the potential to coordinate these biological factors between individuals, resulting in a feeling of social bonding and togetherness. Wilermuth & Heath (2009) proposed that the creation of social bonding can be attributed to general synchronization of actions among people. Synchronous movement has been shown to potentially increase positive feelings toward one another. When individuals engage in actions that are similar to the actions of another person, an engagement with neural pathways occurs within the primary motor cortex. Mirror neurons trigger a system of empathy; the neurons fire similarly when watching an action being performed as they do when the individual actually performs the action (Livingstone & Thomspon, 2009). The synchronization of mirror neurons that is created between individuals when viewing and/or performing a similar task is responsible for subsequently creating a bond between individuals. In turn, this bond makes it increasingly difficult to differentiate between self and the rest of a group when acting in synchrony (Tarr, Launay, & Dunbar, 2014).

Angelucci, Ricci, Padua, Sabino & Tonali (2007) found that in addition to its impact on the brain's mirror neuron system, music stimulates specific pathways within the brain that are closely associated to emotional behavior. One of the brain areas that is positively stimulated is the hypothalamus. The hypothalamus works closely with the pituitary gland in order to release a variety of hormones into the body (Angelucci et al., 2007). Oxytocin, a chemical released by the pituitary gland, plays a significant role in social bonding. The main purpose of oxytocin is to regulate stress and anxiety, as well as to act as a mediating factor in social behavior (Heinrichs, von Dawans, & Domes, 2009). Oxytocin works by promoting positive and negative emotions as well as increasing feelings of trust among individuals (Bartz & Hollander., 2006). Keeler et al. (2015) explains the phenomenon that is seen during synchronous singing among strangers in a group. When strangers engage in synchronous singing, oxytocin levels rise which correlate to positive feelings toward one another. This enables individuals in a group to experience high levels of social bonding due to increased levels of oxytocin and various neural hormonal components that are all activated when acting in synchrony.

Social bonding can also be associated to synchrony at a neurological level found through the mirror neuron system. The ideas of the musical experience can be combined with those of the mirror neuron system, creating a model known as SAME – the model of shared affective motor experience (Overy & Molnar-Szakcas, 2009). This model shows that music can be linked to almost every major region of the brain, meaning music is directly connected to both social cues and bonding. These neural patterns allow for an observer to pick up on the same neural system as someone who is listening to music. By connecting the neural systems of the listener to the system of the observer, we can begin to make inferences about forming social bonds. One study connected participants to an fMRI machine and found there is a link between physical action and neural connection (Overy & Monar-Szakcas, 2009). The physical connection found in participants can be tested to study the potential social connections. Further, the brain does not function in isolation. It combines with the efforts of others to create a 'shared system' mechanism (Overy & Monar-Szakcas, 2009). The SAME model uses this finding to propose that a musical sound is not only heard, but processed in terms of the motor acts connected with it as well as the the connection with another's motor neuron system (Overy & Monar-Szakcas, 2009). Imitation can be used as an element of shared behavior, and this shared behavior can further be extended to include shared social bonds and feelings of togetherness. While music does not function as its own language, it is found to tie in with every person's life, and bonds are informally made when music is being shared by people. Bonding is promoted through the concepts of the SAME model. Through the connections that are made between two brains when listening to music together, we can begin to develop hypotheses about the relationship between music and social bonding.

Our past research tested the effects of physical movement and singing on social bonding. Within each trial, two confederates, male and female, were on either side of the participant. Participants participated in one of the six conditions: singing and moving, moving and no singing, no singing and no moving, singing and no moving, singing and moving asynchronously and no singing and moving asynchronously. Participants in the singing conditions sang the song *"Row, Row, Row, Your Boat"*. We hypothesized that the strongest bonding would be reported in the singing and moving condition. However, according to the results singing alone, moving alone, and singing and moving all produced similar levels of bonding (Edelman, Harring, et al., 2016). Thus, in the conditions of moving and singing the ratings of rapport and entitativity were the same. There were no additive effects of bonding when participants sang and moved together. Lastly, in the asynchronous moving trial reported the lowest levels of bonding. In this study music synchrony occured in each condition, but the synchrony of the movement varied.

The current study aimed to isolate these features to find how much social bonding is created with music making and synchrony independently from one another, excluding the variable of movement. To do this we created an aspect of asynchrony in the music itself, thus examining music alone, and compared its effects to those of the synchronous music. We created four different groups: asynchronous singing, asynchronous singing in rounds, synchronous singing in rounds, and unison singing. We hypothesized that the highest level of entitativity and rapport would be found in groups who sang in unison, a moderate amount in those who sang in synchronous rounds, and the lowest amount of entitativity and rapport in those with who sang in asynchronous rounds.

Methods

Participants

The sample included 71 participants, 53 women and 18 men, all over the age of 18. There were 6 participants removed from the data set due to demand characteristics and failure to believe the cover story for the study.

Design

We conducted a 2 x 3 x 4 between subjects factorial experiment. The 4-group between-subjects independent variable was the singing manipulations .The four conditions were singing *"Row, Row, Row Your Boat"* in unison, singing in synchronous rounds, and singing in asynchronous rounds, and singing in completely asynchrony. We measured the reported amount of bonding between subjects (measured by entitativity and rapport), mood ratings of subjects, and manipulations checks. Additionally, participants rated perceptions of the three confederates in the study.

Materials

After experiencing the singing manipulations, participants completed a questionnaire to measure the amount of rapport and entitativity felt for the other participants. A table of eight statements was presented containing columns for responses regarding feelings toward each of the four participants, however, the participant was instructed not to rate themselves. The statements were rated using a 7-point Likert-type scale, with higher ratings indicating more agreement with the statements, and lower ratings indicating more disagreement with the statements. Four statements assessed perceptions of rapport, including the questions, "I felt that this person and I understood each other..." and "I experienced a feeling of togetherness with...". In addition, four questions assessed perceptions of entitativity, including, "I felt coordinated with..." and "I felt that I was a unit with...". Participants indicated the gender with which they identified, how familiar they were with the other experimental group members, and their musical experience. Furthermore, participants rated their mood on a series of 7-point semantic differential scales with varying emotions. These emotions include: happy/sad, negative/positive, excited/calm, confident/insecure, etc.

Procedure

Undergraduate college students participated in what they thought was an experiment on "mindfulness". Each condition involved one participant and three experimenters acting as confederates. All participants sat in seat 2 of 4 chairs, with confederates in the remaining seats. The experimenter explained that the study was meant to investigate factors that affect mindfulness to promote positive study habits. Participants were randomly assigned into one of the four singing conditions. The conditions were: (1) singing in synchronous unison; (2) singing in synchronous rounds, (3) singing with two confederates singing in a round asynchronously with each other, as well as asynchronously with the participant and other confederate who sang in unison (singing in a asynchronous round); and (4) singing with all of the confederates singing asynchronously with the participant. After the experiment was finished, participants completed the questionnaire measuring rapport, entitativity, mood, and manipulation checks. Lastly, the participants answered post-experimental questionnaire to evaluate their beliefs about the study and to assess demand characteristics.

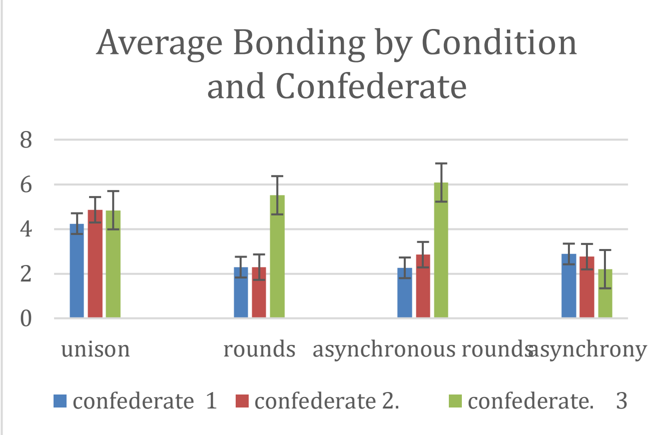
Results

Responses to four questions that measured perception of entitativity were averaged together to get an overall rating of entitativity. Responses to five questions that measured rapport were averaged together to get an overall rating of rapport.

The data were analyzed using a mixed, three-way analysis of variance. The between-subjects variables were the four different singing conditions. The within-subject variables were the separate ratings of each of the three confederates. The difference between entitativity and rapport did not interact with any other factor so we considered them together as a measure of social bonding.

There was a significant interaction between the singing condition, and the individual confederates ,$F(6,134) = 3.52$, $p < .01$, $\eta^2 = .14$. The means in Table 1 indicate that the confederates were rated the same in unison condition, and the more asynchronous the trials became the more bonded the participant felt to the confederate who stayed in unison with the participant. However, in the completely asynchronous trial the participant did not feel more bonded to any confederate. The participants' ratings varied significantly by the condition. LSD post hocs show that the unison condition differed significantly from the synchronous rounds, asynchronous, rounds and synchronous conditions ($p < .01$ for all). Additionally, LSD post hocs indicate the complete asynchronous condition differed significantly from the unison condition, synchronous rounds condition, and synchronous rounds condition ($p < .01$ for all). However, LSD post hocs indicate the synchronous rounds and asynchronously do not vary significantly by ratings ($p = .19$).

In addition, we conducted several manipulations checks. Mood ratings were averaged together to form a single positive mood scale. In general, the participants rated their mood as slightly positive ($M = 5.00$, $SD = 1.82$). There were no significant differences among the conditions for mood or other manipulation checks, $p > .05$.



Discussion

Our results showed that bonding scores increased significantly toward the confederate the participant was in unison with as the asynchrony of the trials increased, with the participant feeling low levels of bonding toward all confederates in the completely asynchronous condition. The introduction of rounds immediately increased the participant's bond with the person they were singing with, and this difference became even greater with asynchronous rounds.

Social psychological theories of group cohesion based on proximity, similarity, common fate and pragnance could be used to explain these results (Campbell, 1958). The results show that the participant felt they were a unit with the confederate they sang in unison with. The factors of singing together synchronously or being physically close to another person, leading to a higher feeling of social bonding, could be explained by the idea of common fate or proximity. The results may explain why the participants felt more bonded to the confederate singing synchronously with them due to the fact that the participants are under the impression that they are working toward a common goal – singing in unison with the confederate. Bonding with the confederate singing in unison with the participant can also be explained by the mirror neuron theory. The synchronization of mirror neurons that is created between individuals when viewing or performing a similar task is responsible for subsequently creating a bond between individuals, which makes it difficult to differentiate between a person and group, in this case the confederates, when acting in synchrony (Tarr et al. 2014). The mirror neuron theory helps to explain the increased scores of bonding toward the confederate the participant sang with, as their mirror neuron systems would successfully align in synchrony and cause them to feel a sense of togetherness.

One limitation of the study is that our manipulation of singing asynchronously did not just remove synchrony but introduced active asynchrony and very discordant sounds. Both the unpleasantness of the sound and the contrasting behaviors may have activated a dislike for the confederates. In future studies we plan to remove the act of making music and investigate the effects of music versus silence on a cooperative and a competitive task. We hypothesize that the cooperative task will provide the participant with a sense of common fate with confederate and therefore trigger bonding. The competitive task should have a negative effect on bonding. Adding the music will allow us to assess the degree to which music increases cohesiveness in an already positive bonding situation (cooperative task) and the degree to which it can induce connection in a negative bonding situation (competitive task).

References

Angelucci, F., Ricci, E., Padua, L., Sabino, A., & Tonali, P. A. (2007). Music exposure differentially alters the levels of brain-derived neurotrophic factor and nerve growth factor in the mouse hypothalamus. *Neuroscience Letters, 429*(2-3), 152-155. doi:10.1016/j.neulet.2007.10.005

Bartz, J. A. and Hollander, E. (2006) The neuroscience of affiliation: forging links between basic and clinical research on neuropeptides and social behavior. *Horm. Behav. 50*, 518–528.

Campbell, D. (1958) Common fate, similarity, and other indices of the status of aggregates of persons as social entities, *Syst. Res. Behavioral Science*, *3*(1), 14-25. doi:10.1002/bs.3830030103

Edelman, L. L. & Harring, K. E. (2014) Music bonds the video walkers but not the participants, Unpublished manuscript.

Hove, M.J. & Risen, J.L. (2009). It's all in the timing: Interpersonal synchrony Increases affiliation, *Social Cognition*, *27*(6), 949-961.

Harring, K., Edelman, L., Ross, J., Ninivaggi, K., Mulvey, K., & Hogan, M. (2016). Singing and Moving Increases Social Bonding. 310-312. Proceedings of the ICMPC14

Heinrichs, M., von Dawans, B., & Domes, G. (2009). Oxytocin, vasopressin, and human social behavior. *Frontiers in Neuroendocrinology*, *30*(4), 548-557. doi.org/10.1016j.yfrne.2009.05.005

Keeler, J. R., Roth, E. A., Neuser, B. L., Spitsbergen, J. M., Waters, D. J. M., & Vianney, J.-M. (2015). The neurochemistry and social flow of singing: bonding and oxytocin. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00518

Livingstone, S. R. & Thompson, W. F. (2009). The emergence of music from the Theory of Mind, Musicae Scientiae, *Special Issue*, 83-115.

Loersch, C., & Arbuckle, N. L. (2013). Unraveling the mystery of music: Music as an evolved group process. *Journal of Personality and Social Psychology, 105*(5), 777-798. doi:10.1037/a0033691

Overy, K. & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system, *Music Perception*, *26(* 5), 489-504. doi:10.1525/mp.2009.26.5.489

Tarr, B., Launay J., & Dunbar, R. M. (2014). Music and social bonding: 'Self- other' merging and neurohormonal mechanisms, *Frontiers In Psychology*, *5*.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science, 20*(1), 1-5. doi:10.1111/j.1467-9280.2008.02253.x

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Dissociating sensory and cognitive theories of harmony perception through computational modeling

Peter M. C. Harrison[1] and Marcus T. Pearce[2]

[1,2]*Cognitive Science Research Group, School of Electronic Engineering and Computer Science, Queen Mary University of London, UK*

[1]p.m.c.harrison@qmul.ac.uk, [2]marcus.pearce@qmul.ac.uk

Abstract

Two approaches exist for explaining harmonic expectation. The *sensory* approach claims that harmonic expectation is a low-level process driven by sensory responses to acoustic properties of musical sounds. Conversely, the *cognitive* approach describes harmonic expectation as a high-level cognitive process driven by the recognition of syntactic structure learned through experience. Many previous studies have sought to distinguish these two hypotheses, largely yielding support for the cognitive hypothesis. However, subsequent re-analysis has shown that most of these results can parsimoniously be explained by a computational model from the sensory tradition, namely Leman's (2000) model of auditory short-term memory (Bigand, Delbé, Poulin-Charronnat, Leman, & Tillmann, 2014). In this research we re-examine the explanatory power of auditory short-term memory models, and compare them to a new model in the Information Dynamics Of Music (IDyOM) tradition, which simulates a cognitive theory of harmony perception based on statistical learning and probabilistic prediction. We test the ability of these models to predict the surprisingness of chords within chord sequences (N = 300), as reported by a sample group of university undergraduates (N = 50). In contrast to previous studies, which typically use artificial stimuli composed in a classical idiom, we use naturalistic chord sequences sampled from a large dataset of popular music. Our results show that the auditory short-term memory models have remarkably low explanatory power in this context. In contrast, the new statistical learning model predicts surprisingness ratings relatively effectively. We conclude that auditory short-term memory is insufficient to explain harmonic expectation, and that cognitive processes of statistical learning and probabilistic prediction provide a viable alternative.

Introduction

Western tonal harmony obeys a complex set of conventions concerning how notes combine to form chords, and how chords combine to form chord sequences. These conventions are termed *harmonic syntax*. Listeners are sensitive to aspects of harmonic syntax, and implicitly judge the syntactic implications of successive chords as they hear a piece of music. This online sensitivity to harmonic syntax is termed *harmonic expectation*.

Theories of the psychological origins of harmonic expectation can be organized along a *sensory-cognitive* spectrum (Collins, Tillmann, Barrett, Delbé, & Janata, 2014). *Sensory* theories hold that harmonic expectation is driven by low-level sensory responses to acoustic properties of musical sounds. Conversely, *cognitive* theories posit that harmonic expectation is driven by high-level cognitive processes similar to those involved in processing linguistic syntax.

Many empirical studies have accumulated in favor of cognitive theories of harmonic expectation. These studies typically report that listeners display sensitivity to syntactic violations that could not be detected by sensory cues alone. However, much of this work was recently undermined by Bigand et al. (2014): the authors reanalyzed data from 18 tonal expectation studies, including 17 harmonic expectation studies, and found that most results could be explained with a sensory model of auditory short-term memory (Leman, 2000). The conflict between sensory and cognitive theories therefore remains largely unresolved.

The present research takes a new approach to the problem of disentangling sensory and cognitive theories of harmonic expectation. We avoid the problematic task of hand-constructing sequences where sensory cues contradict syntactic rules, and avoid the problematically small stimulus sets associated with this approach. Instead, we separate sensory and cognitive theories through computational modeling, giving us the flexibility to use a large dataset of naturalistic stimuli derived from real music.

Many sensory and cognitive theories of harmonic expectation exist in the literature, but here we restrict comparison to two types that have received particular empirical support in recent decades. *Auditory short-term memory* theories explain expectation through the retention and comparison of auditory images in short-term memory. These theories are typically sensory in nature. *Statistical learning* theories, meanwhile, claim that harmonic expectation corresponds to probabilistic predictions made by listeners who have internalized the statistical structure of musical styles through exposure. Statistical learning provides a cognitive alternative to auditory short-term memory theories.

Several auditory short-term memory models exist in the literature. We evaluate three here: Leman's (2000) periodicity-pitch model, Milne, Sethares, Laney & Sharp's (2011) spectral distance model, and Collins et al.'s (2014) tonal expectation model.

Fewer statistical learning models exist in the literature. The Information Dynamics Of Music (IDyOM) model (Pearce, 2005, 2018) is a prominent candidate, but so far it has largely been limited to the melodic domain. We therefore introduce an extension of the model to harmonic expectation and compare it with the auditory short-term memory models.

Our model evaluation involves an explicit behavioral measure of harmonic expectation where listeners rate the surprisingness of particular chords within chord sequences. We evaluate each model in terms of its ability to predict these surprisingness ratings: good models should deliver accurate predictions.

Models

Leman's (2000) Periodicity-Pitch Model

Leman's (2000) model is a sensory auditory short-term memory model. It was recently shown to explain a wide variety of experimental findings previously thought to support a cognitive account of harmony perception (Bigand et al., 2014).

The model takes an audio signal as input and simulates the acoustic filtering of the outer and middle ear, the resonance of the basilar membrane in the inner ear, and the conversion of the resulting signal into neural rate-code patterns. A periodicity analysis produces a series of *pitch images* representing the instantaneous pitch patterns perceived at every point in the audio signal. The model then simulates the persistence of these pitch images in auditory short-term memory to produce *echoic pitch images*. Echoic pitch images are created by *leaky integration*, where the echoic pitch image at each timestep is created by adding the current non-echoic pitch image to the echoic image from the previous time-step. The length of the 'echo' is determined by a time constant defining the echo's half-life.

Two echoic pitch images are created: a *local* image and a *global* image. The *local* pitch image summarizes the pitch content of the immediate temporal context (e.g. the last 0.5 seconds), whereas the *global* pitch image summarizes pitch over a longer temporal context (e.g. 5 seconds). The moment-to-moment similarity between these local and global pitch images is summarized in a *tonal contextuality* profile. Applied to chord sequences, points of high tonal contextuality correspond to chords that are tonally consistent with their recent musical context, whereas low tonal contextuality reflects low tonal consistency.

For this study we defined the model's estimate of a chord's surprisingness as the negative mean tonal contextuality during the time that chord was playing. We used the MATLAB model implementation as created by the original author and available in the IPEM toolbox.[1] Decay constants for the local and global pitch images were set to 0.1 s and 1.5 s, as optimized in Leman (2000). All other parameters were left at their default values.

Milne et al.'s (2011) Spectral Distance Model

This is a second sensory model that also embodies an auditory short-term memory theory of tonal perception. It avoids much of the complexity of Leman's (2000) model: it has no explicit modeling of the peripheral auditory system and does not model the time-course of echoic memory. Nonetheless, the model has demonstrated best-in-class results in modeling certain important results from the psychological literature (Milne & Holland, 2016; Milne, Laney, & Sharp, 2015).

Milne et al.'s (2011) model estimates the perceptual dissimilarity of pairs of pitch or pitch-class sets. It combines each harmonic series implied by every pitch(-class), smooths the resulting spectra to account for perceptual imprecision, and then computes the cosine distance between these spectra. This cosine distance has been shown to predict perceptual judgements of triadic similarity rather effectively (Milne & Holland, 2016).

We applied this model to harmonic expectation by using it to model the perceptual dissimilarity of a chord and its context. The model's dissimilarity estimate was taken as an estimate of chord surprisingness. Context could be defined in several ways; here we defined the context as the immediately preceding chord, but future work should explore alternative definitions.

The model has three important free parameters: the number of harmonics implied by each pitch(-class), the degree of amplitude roll-off as a function of harmonic number, and the degree of spectral smoothing. Additionally, there is freedom to choose either a pitch representation or a pitch-class representation. This study followed the model configuration as psychologically optimized in Milne and Holland (2016): 12 harmonics, roll-off parameter of 0.75, spectral smoothing parameter of 6.83, and a pitch-class representation. The model was given a new Common Lisp implementation, with a selection of results verified against the original author's MATLAB implementation.[2]

Collins et al.'s (2014) Tonal Expectation Model

While the two previous models (Leman, 2000; Milne et al., 2011) describe sensory accounts of tonal perception, this model describes an intermediate sensory-cognitive account of tonal perception. It is still an auditory short-term memory model, but some of its auditory representations involve cognitive abstractions.

The model centers on three representations of the musical input. The first is a *periodicity-pitch* representation corresponding to the pitch images of Leman's (2000) model. A *chroma-vector* representation is derived from the periodicity-pitch representation by collapsing pitches to pitch classes. Lastly, a *tonal space* representation is produced by projecting the periodicity-pitch representation onto a toroidal self-organizing map, after Janata et al. (2002). This last representation allows the model to learn a map of tonal space stored in long-term memory.

Incoming audio activates these representations in a cascade. These activations are blurred by echoic memory, analogous to Leman's (2000) model. For each of the three representations (periodicity pitch, chroma vector, tonal space), local and global images are created which evolve over the course of the stimulus. Local images are created by leaky integration with a short time constant (0.1 s); global images are produced by leaky integration with a longer time constant (4.0 s). These images summarize the recent activation of the respective representational space over the given time period.

Various features are then derived from these blurred activations. These features include both the correlation between local and global images (after Leman, 2000) and the peak activation within the global image. Extending Leman (2000), the model considers both absolute values of the correlations/peak activations and relative changes in these values. Different features are constructed that apply these computations to different time windows within the stimulus.

These features are then combined by linear regression, using coefficients optimized on a set of seven empirical

[1] http://www.ipem.ugent.be/Toolbox

[2] http://www.dynamictonality.com/probe_tone_files/

studies (Collins et al., 2014). The model's output corresponds to predicted deviations in reaction times in priming studies, with slower reaction times being thought to reflect lower expectedness. We therefore used these predicted reaction times as the model's estimate of chord surprisingness.

We used the original authors' implementation as available in the Janata lab music toolbox (JLMT).[3] All parameters were left at their default values.

Information Dynamics Of Music

The Information Dynamics Of Music (IDyOM) model (Pearce, 2005, 2018) is a statistical learning model of musical expectation. It asserts that listeners acquire statistical knowledge of musical syntax through experience, and use this knowledge to generate probabilistic predictions as a piece unfolds.

The model takes a musical score as its input. It generates predictions in an incremental manner, predicting each event based on the preceding events. Its predictions are based on a variety of representations derived from the musical score, termed *viewpoints*. These viewpoints are intended to mirror the psychological representations of music available to listeners (e.g. Shepard, 1982). Predictions are generated by combining the output of a *long-term model*, which captures syntactic structure common to an entire musical style, and a *short-term model*, which captures statistical patterns heard so far in the current musical piece.

The original IDyOM model was restricted to melodies (Pearce, 2005). Here we extended the model to process chord sequences, substantially expanding the range of music to which the model can be applied.

The primary challenge in extending the IDyOM model to harmony is to develop new viewpoints for capturing harmonic structure. Our viewpoint set captures a variety of well-established psychological and music-theoretic principles of harmony, including inversion invariance, transposition invariance, chord roots, scale degrees, and intervallic relationships between successive chords. Chord roots are inferred using the General Chord Type algorithm of Cambouropoulos, Kaliakatsos-Papakostas, & Tsougras (2014), and scale degrees are inferred using the key-finding algorithm of Albrecht & Shanahan (2013). This explicit modeling of root-finding and key-finding improves cognitive realism compared to previous multiple-viewpoint models of harmony, which read key information and chord roots directly from the score (Hedges & Wiggins, 2016; Sears, Pearce, Caplin, & McAdams, 2018; Whorley & Conklin, 2016).

A further change introduced in this research was to replace the IDyOM model's heuristic entropy-weighting scheme with a supervised weighting scheme, where the contribution of each viewpoint is determined by a weight vector derived using maximum-likelihood estimation. The cognitive implications of this change are outside the scope of the present paper, but will be explored in future work.

Consistent with previous IDyOM research, we computed an estimate of surprisingness as the negative logarithm of the chord's conditional probability according to the statistical model. This quantity is termed *information content*.

The resulting model was implemented in the programming languages Common Lisp and R. The implementation extends the publicly available IDyOM codebase.[4]

Methods

Participants

Fifty psychology undergraduates (44 female, six male) participated in exchange for course credit or small financial reward. The mean age was 18.7 years (*SD* = 1.7). Most self-reported as frequent listeners to popular music. The mean musical training score as assessed by the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen, Gingras, Musil, & Stewart, 2014) was 15.1 (*SD* = 8.0), corresponding to the 22nd percentile of the original Gold-MSI sample.

Stimuli

Chord sequences were sourced from the Billboard corpus (Burgoyne, 2012). This dataset comprises a set of transcriptions of popular songs sampled from the Billboard magazine's United States "Hot 100" chart between 1958 and 1991. Three hundred eight-chord sequences were randomly sampled from this dataset, with repeated chords removed, under the constraint that no song appeared twice. Sequences were played with a piano timbre at a tempo of 60 beats per minute without metrical cues. Bass notes were played in the octave below middle C, non-bass notes in the octave above.

Procedure

Participants took the experiment individually in a quiet room at a desktop computer, navigating the experiment using keyboard and mouse. Audio was played over headphones.

The main part of the experiment comprised 150 trials for each participant. In each trial, the participant was played a sequence of eight chords and instructed to rate the sixth chord in this sequence, termed the *target*, for surprisingness. This chord was visually cued by a continuous clock-like animation, and surprisingness ratings were given on a scale from one to nine using the computer keyboard. Participants were given 10-second breaks every 25 trials. Each participant's 150 chord sequences were randomly chosen under the constraints that no participant heard the same sequence twice and that each of the 300 possible chord sequences was presented equally often to all participants over the course of the study.

The main part of the experiment was preceded by a training routine which included three practice trials. After the main part of the experiment, the participant completed a short questionnaire concerning basic demographic details and familiarity with popular music, and then completed the musical training component of the Goldsmiths Musical Sophistication Index (Gold-MSI; Müllensiefen et al., 2014). On average the procedure lasted approximately 40 minutes.

[3] http://atonal.ucdavis.edu/resources/software/jlmt/

[4] https://code.soundsoftware.ac.uk/projects/idyom-project

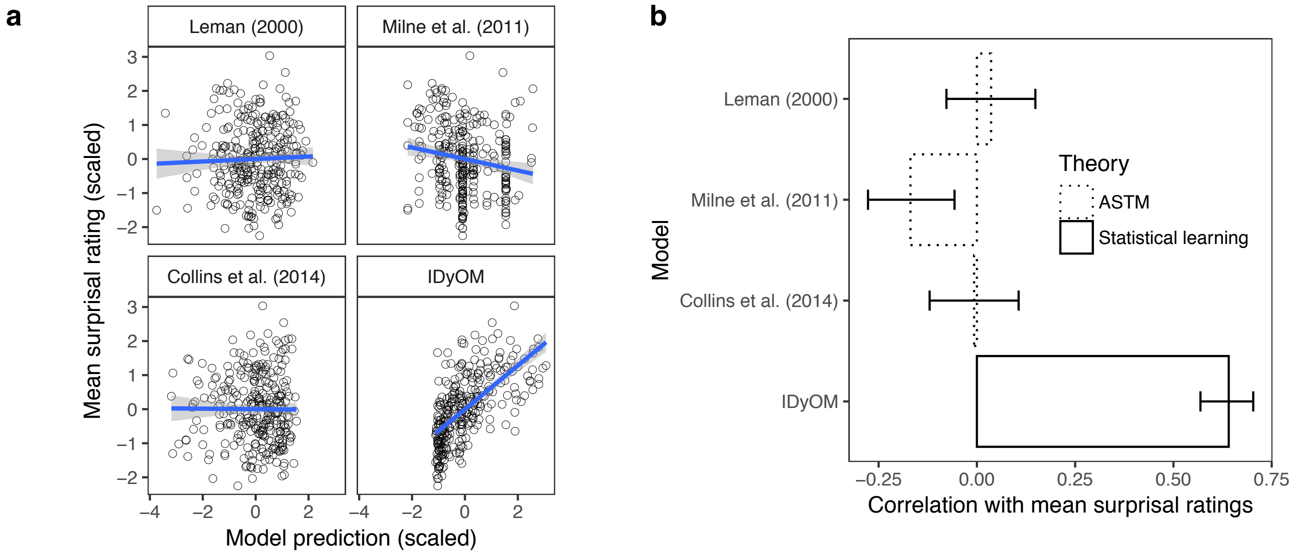


Figure 1. a) Scatterplots of mean surprisal rating (*z*-score) against model prediction (*z*-score) for the four computational models. Linear regression lines are plotted for each model, with the standard errors of these lines shaded in grey. b) Bar chart displaying Pearson correlation coefficients for the relationship between each model's predictions and mean surprisal ratings. The error bars denote 95% confidence intervals. 'ASTM' stands for auditory short-term memory.

Results

Data Preprocessing

One participant was found to give the same response for all 150 stimuli, and so their data were removed from further analyses. The remaining 49 participants had their ratings standardized to *z*-scores to normalize across individual differences in scale usage, and then these *z*-scores were averaged across participants to produce a mean surprisal rating for each chord sequence. These surprisal ratings were then *z*-transformed again across all stimuli, so that the mean surprisal rating would be zero and the standard deviation one.

Each of the four computational models was applied to the target chord in each chord sequence. Model outputs were converted to *z*-scores to facilitate comparison across models.

Model Predictive Performances

The predictive performance of each model was assessed in terms of its Pearson correlation with mean surprisal ratings: a high correlation means that the model predicted perceived surprisal well. These results are summarized in Figure 1 and the last column of Table 1. The three auditory short-term memory models did not display significant positive correlations with mean surprisal ratings. Surprisingly, Milne et al.'s model displayed a significant correlation in the opposite direction to that predicted by theory: greater spectral distance was significantly associated with lower surprisal ratings ($r(298) = -.169$, $p = .003$, 95% CI = [−.277, −.057]). In contrast, IDyOM model outputs exhibited a moderately large positive correlation with surprisal ratings ($r(298) = .641$, $p < .001$, 95% CI = [.569, .703]). A linear regression model predicting surprisal ratings from the four computational models found a significant coefficient for the IDyOM model ($p < .001$) but non-significant coefficients for the remaining models (all *p*-values > .3).

Model Correlations

Table 1 displays pairwise correlations within the set of computational models. Leman's (2000) model outputs and Milne et al.'s (2011) model outputs were fairly well correlated ($r(298) = .591$, $p < .001$, 95% CI = [.512, .660]). Surprisingly, Collins et al.'s (2014) model outputs were not significantly correlated with any of the other model outputs (all *p*-values > .3). IDyOM model outputs were significantly negatively correlated with Milne et al.'s (2011) model outputs ($r(298) = -.203$, $p < .001$, 95% CI = [−.309, −.092]) but not significantly correlated with outputs of the other two models (both *p*-values > .2).

Table 1. Pairwise Pearson correlation coefficients for surprisal ratings and the four computational models.

	Milne et al. (2011)	Collins et al. (2014)	IDyOM	Surprisal ratings
Leman (2000)	.591	−.004	.071	.036
Milne et al. (2011)		.053	−.203	−.169
Collins et al. (2014)			.04	−.007
IDyOM				.641

Musical Examples

Figure 2 displays four specific chord sequences from the dataset; corresponding model outputs are displayed in Table 2. These examples were selected as follows: a) the most and b) the least surprising stimuli according to Leman's (2000) model; c) the most and d) the least surprising stimuli according to the IDyOM model.

Sequence a) comprises solely bare fifths. The progression from ‘EB’ to ‘GD’ on the sixth chord is considered somewhat unsurprising by listeners ($z = -0.101$): in this context, ‘EB’ implies E minor, and ‘GD’ implies G major, so the passage is simply a version of the common progression i-III with the third of each chord removed. However, this passage is considered particularly unexpected by Leman et al.’s (2000) model and Milne et al.’s (2011) model. This is presumably because the missing thirds are the two common tones between these chords; without them, the chords are rather acoustically dissimilar.

Sequence b) alternates between two inversions of the same major triad. The target chord is very spectrally similar with its previous context, yielding low surprisal ratings from Leman et al.’s (2000) model and Milne et al.’s (2011) model. Surprisingly, Collins et al.’s (2014) gives a relatively high surprisal rating for this chord. As expected, the IDyOM model finds the chord relatively predictable, in large part because the same transition occurs several times in the stimulus.

Sequence c) begins with conventional progressions along the circle of fifths. However, the target chord is very tonally distant from its context: it corresponds to a semitone displacement of the previous chord with a tonic pedal in the bass. Correspondingly, most of the models predict high surprisal. The exception again is Collins et al.’s (2014), which predicts only moderate surprisal.

Sequence d) was considered very unsurprising by the participants and by the IDyOM model, but very surprising by the auditory short-term memory models. It corresponds to a major-mode IV-V progression, which is very common in Western popular music. The IDyOM model therefore finds the progression very predictable, because similar progressions occur many times in the corpus and even at the start of the same chord sequence. However, the two chords are not particularly similar spectrally speaking, and so they are considered surprising by the auditory short-term memory models.

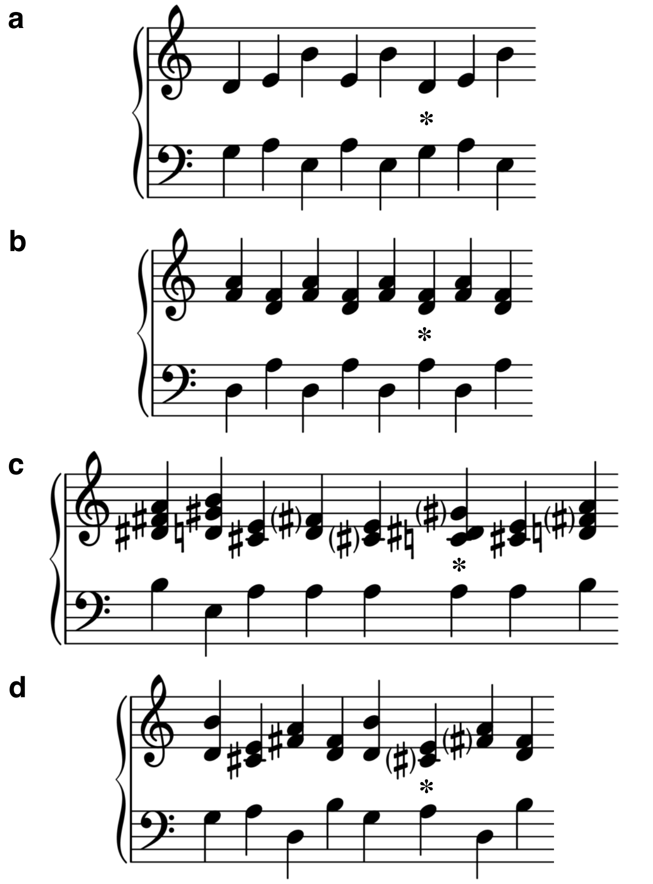


Figure 2. Four example stimuli. Target chords are indicated by asterisks. Sequence a) is the most surprising stimulus in the corpus, according to Leman’s (2000) model. Sequence b) is the least surprising according to Leman’s (2000) model. Sequence c) is the most surprising according to the IDyOM model. Sequence d) is the least surprising according to the IDyOM model. Precise model outputs and participant ratings are given in Table 2.

Table 2. Surprisal ratings as predicted by the models and as reported by the participants. All scores are *z*-scores where higher scores correspond to higher surprisal.

	Leman (2000)	Milne et al. (2011)	Collins et al. (2014)	IDyOM	Participants
a)	2.17	2.51	0.49	0.06	−0.10
b)	−3.74	−2.16	0.35	−0.21	−1.50
c)	1.00	0.56	0.02	3.07	1.28
d)	1.35	1.54	0.79	−1.11	−1.56

Discussion

We tested two competing explanations of harmonic expectation: an auditory short-term memory explanation and a statistical learning explanation. According to the former, harmonic expectation is a low-level process driven by the accumulation of auditory images in short-term memory. According to the latter, harmonic expectation reflects probabilistic predictions of listeners which derive from internalized statistical knowledge about musical styles.

The results were unambiguous. None of the auditory short-term memory models produced statistically significant correlations with surprisal ratings in the direction predicted by theory. In contrast, the statistical learning model predicted surprisal ratings moderately well. The results therefore strongly corroborate the statistical learning account over the auditory short-term memory account.

It is plausible that the statistical learning model might outperform the auditory short-term memory models, but highly surprising that the latter models should not outperform chance. Pre-existing literature gives the impression that harmonic syntax is ineluctably correlated with spectral similarity, with the result that auditory short-term memory models can explain the results of most existing harmonic expectation studies (Bigand et al., 2014). However, we found that these models had no explanatory power for our dataset. Moreover, their predictions did not correlate positively with the predictions of the statistical learning model, suggesting that the correlation between spectral similarity and harmonic syntax was minimal at best.

We suggest several possible reasons for this discrepancy. First, while previous studies typically used stimuli in the style of Western classical music, this study used stimuli from Western popular music, where spectral distance seems to be a worse predictor of harmonic style (Harrison & Pearce, 2018). Second, many previous studies tested perception of final chords in sequences, whereas we tested the perception of non-final chords. Stylistic expectations will differ between these contexts. Third, much of the harmonic expectation literature relies on harmonic priming paradigms with the listener's task being to detect tuning or timbre deviants. Both of these tasks involve detecting deviations in acoustic spectra, and it is clear that this might be facilitated by maximizing spectral similarity between successive chords. Reliance on this paradigm might therefore overemphasize the role played by spectral similarity (and correspondingly auditory short-term memory) in harmonic expectation. Fourth, we have yet to exhaust the potential of each computational model. Leman's (2000) model might perform better with different time constants (e.g. Bigand et al., 2014). Milne et al.'s (2011) model might be improved by incorporating inharmonic partials, different degrees of spectral smoothing, or a continuously decaying echoic memory. The predictors used in Collins et al.'s (2014) model might still have useful explanatory power, even if the regression model doesn't generalize well. Likewise, many computational aspects of the harmonic IDyOM model remain to be psychologically optimized.

We intend to explore these computational models further in ongoing research. However, as the results stand, it seems that auditory short-term memory is insufficient to explain harmonic expectation. We have shown that one viable alternative is statistical learning. However, several other alternatives exist at the sensory end of the sensory-cognitive spectrum. In particular, earlier literature has emphasized the importance of roughness and voice-leading distance in harmonic expectation, both of which correspond to relatively low-level psychological processes (Bigand, Parncutt, & Lerdahl, 1996). Perhaps it is time to re-examine these alternative sensory models.

Acknowledgements. The authors would like to thank Emmanouil Benetos and Matthew Purver for useful feedback and advice regarding this project. PH is supported by a doctoral studentship from the EPSRC and AHRC Centre for Doctoral Training in Media and Arts Technology (EP/L01632X/1).

References

Albrecht, J., & Shanahan, D. (2013). The use of large corpora to train a new type of key-finding algorithm: An improved treatment of the minor mode. *Music Perception*, *31*(1), 59–67.

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00094

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, *58*(1), 124–141.

Burgoyne, J. A. (2012). *Stochastic Processes & Database-Driven Musicology*. (Unpublished doctoral dissertation). McGill University, Montréal, Québec, Canada.

Cambouropoulos, E., Kaliakatsos-Papakostas, M., & Tsougras, C. (2014). An idiom-independent representation of chords for computational music analysis and generation. In *Proceedings of the joint 11th Sound and Music Computing Conference (SMC) and 40th International Computer Music Conference (ICMC)*. Athens, Greece.

Collins, T., Tillmann, B., Barrett, F. S., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review*, *121*(1), 33–65.

Harrison, P. M. C., & Pearce, M. T. (2018). An energy-based generative sequence model for testing sensory theories of Western harmony. In *Proceedings of the 19th International Society for Music Information Retrieval Conference*. Paris, France.

Hedges, T., & Wiggins, G. A. (2016). The prediction of merged attributes with multiple viewpoint systems. *Journal of New Music Research*, *45*(4), 314–332.

Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science*, *298*(5601), 2167–2170.

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17*(4), 481–509.

Milne, A. J., & Holland, S. (2016). Empirically testing Tonnetz, voice-leading, and spectral models of perceived triadic distance. *Journal of Mathematics and Music*, *10*(1), 59–85.

Milne, A. J., Laney, R., & Sharp, D. (2015). A spectral pitch class model of the probe tone data and scalic tonality. *Music Perception*, *32*(4), 364–393.

Milne, A. J., Sethares, W. A., Laney, R., & Sharp, D. B. (2011). Modelling the similarity of pitch collections with expectation tensors. *Journal of Mathematics and Music*, *5*(1), 1–20.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), 1–23.

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. (Unpublished doctoral dissertation). City University, London.

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, 1–18.

Sears, D. R. W., Pearce, M. T., Caplin, W. E., & Mcadams, S. (2018). Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *Journal of New Music Research*, *47*(1), 29–52.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *Psychology of Music* (pp. 343–390). New York, NY: Academic Press.

Whorley, R. P., & Conklin, D. (2016). Music generation from statistical models of harmony. *Journal of New Music Research*, *45*(2), 160–183.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Music preferences in an Asian culture: An examination of the MUSIC Model in a Singaporean context

Jiamin Gladys Heng[*1], Li An Leong[*2], Jie Xin Lim[*3], Moon-Ho Ringo Ho[*4], Shen-Hsing Annabel Chen[*,#,^5]

[*] *Psychology, Nanyang Technological University, Singapore*
[#] *Centre for Research and Development in Learning, Nanyang Technological University, Singapore*
[^] *LKCMedicine, Nanyang Technological University, Singapore*

[1]jheng007@e.ntu.edu.sg, [2]lian0101@e.ntu.edu.sg, [3]jlim138@e.ntu.edu.sg, [4]homh@ntu.edu.sg, [5]annabelchen@ntu.edu.sg

Abstract

Recent studies suggest that music preferences can be conceptualized along five dimensions: *Mellow*, *Unpretentious*, *Sophisticated*, *Intense* and *Contemporary* (MUSIC; Rentfrow et al., 2011, 2012). As majority of these studies are conducted in Western contexts, it is less clear if a similar factor structure would emerge in a non-Western culture. Therefore, the current study aims to investigate the validity of the MUSIC model in a non-Western culture. Using the same musical excerpts as in Rentfrow et al. (2012), 83 Singaporean undergraduates reported their degree of liking for each musical excerpt after listening to them. They also completed questionnaires on acculturation and musical background. Factor analysis with orthogonal target rotation was conducted to examine if the factor structure of the current sample matched that of Rentfrow et al. (2012). Tucker's congruence coefficient was then calculated to assess the degree of similarity between both factor structures. Results suggest that all five factors, except the *Unpretentious* factor, demonstrated fair similarity with the factor structure reported in Rentfrow et al. (2012). However, some musical excerpts showed a substantial factor loading on non-targeted factor. Our findings thus show support for a five-factor structure underlying music preferences and extend the validity of the original MUSIC model to a non-Western context.

Introduction

Music is ubiquitous across cultures and serves various cognitive, emotional and social functions (Schäfer & Sedlmeier, 2009). Earlier studies on music preferences mostly adopted methodological approaches which relied on self-report ratings of preferred music genres (e.g. Rentfrow et al., 2003). While this approach has provided valuable insight to the structure of music preferences, it is constrained by several challenges such as a lack of consensus regarding the definition of music genres, the number and type of music genres to study, difficulties in dealing with musical pieces which fit more than one genre, differences in associations of music genres and social connotations in different societies.

Therefore, recent studies have proposed a need to expand the conceptualization of music preferences beyond music genre, and to include the internal and external properties of music as well (Rentfrow et al., 2011). Internal properties refer to the musical features, such as the speed and general style of the music. External properties refer to the psychological attributes of the music, such as the emotion it evokes. With this conceptualization, Rentfrow et al. (2011, 2012) have documented a robust five-factor structure underlying music preferences from different genres as well as within a genre, which they termed as the MUSIC model: Mellow, Unpretentious, Sophisticated, Intense and Contemporary. Songs in the *Mellow* dimension were perceived as slow, relaxed and romantic, consisting of music genres such as soft rock and R&B. Songs in the *Unpretentious* dimension were unaggressive, uncomplicated and soft sounding, consisting of music genres such as country and folk. Songs in the *Sophisticated* dimension were complex, cultured and intelligent, consisting of music genres such as classical, avant-garde and traditional jazz. Songs in the *Intense* dimension were aggressive, tense and loud, consisting of music genres such as classic rock, punk and heavy metal. Songs in the *Contemporary* dimension were rhythmic and danceable, consisting of music genres such as electronica, Latin, acid jazz and Euro pop.

While several studies have shown support for the MUSIC model in Europe (Colley, 2008; Delsing et al., 2008), there has been a relative paucity of music preference studies in a non-Western context. Drawing from findings of music mood perception studies, differences in music mood perception have been reported in American and Chinese participants (Hu & Lee, 2012), and Korean, Chinese and Western participants (Hu & Lee, 2014). In addition, differences between uses of music were found to predict musical preferences in Indian and German participants (Schäfer et al., 2012). In a recent study, McDermott et al. (2016) showed that a tribe with very limited exposure to Western music did not share the same preference as Caucasians for consonance and dissonance sounds. These differences pose a question if differences in the structure of music preferences would also emerge in a non-Western context, given that the cultural context in which individuals live in influence the varieties of music they are exposed to, the social meanings they attach to them, as well as the ways people use music (Cross, 2001).

The availability of musical excerpts in Rentfrow et al (2011, 2012) facilitates research on cross-cultural investigation of music preferences by enabling researchers to measure participants' affective reactions using the same audio excerpts. As such, common difficulties encountered in cross-cultural research, such as translation issues and lack of consistency in genre classifications can be overcome. In addition, statistical techniques such as factor analysis with target rotation (Browne, 2001) and factor congruence coefficient (Tucker, 1951) can be applied to investigate the cross-cultural invariance property of the musical excerpts.

Therefore, the current study presented participants in a non-Western culture with the same set of musical excerpts as in Rentfrow et al. (2012) and employed factor analysis with orthogonal target rotation to investigate the validity of the MUSIC model in a non-Western culture.

Method

Eighty-three undergraduates (28M 55F, age: *M* = 21.9, *SD* = 2.41) were recruited from Nanyang Technological University, Singapore. 77 (93%) were Chinese, two (2%) were Malay and four (4.8%) had other ethnic backgrounds. Thirty-one (37%) participants received formal musical training and 14 (17%) learnt an instrument informally. Sixteen (19%) previously participated in band, 12 (14%) previously participated in choir and four (5%) participated in both musical activities. Participants with prior involvement in musical activities had a range of musical background: 15 had formal musical training, nine had informal musical training, eight had neither formal nor informal musical training. Overall, participants spent a mean of six days a week (*SD* = 1.70) listening to music. A subsequent survey with 22 participants responding indicated the number of hours per day spent purposefully listening to music ranged from less than 1 to more than 6 hours.

Participants listened to 50 excerpts of commercially unreleased musical excerpts which were presented randomly. These musical excerpts were taken from 21 music genres and subgenres (available for download at http://daniellevitin.com/levitinlab/LabWebsite/expsupport/MUSIC/Rentfrow_MP_Index.html), consisting of: adult contemporary, avant-garde classical, bluegrass, classic rock, classical, country-rock, electronica, Europop, heavy metal, Latin, mainstream country, metal, new country, punk, R&B, rap, rock-n-roll, smooth jazz, soft rock, traditional jazz, world beat (Rentfrow et al., 2012). Each excerpt was approximately 15 seconds. Participants rated their level of preference of the musical excerpt on a 9-point Likert scale (1: *Not at all*, 9: *Very much*).

Next, they completed a music background questionnaire and the Suinn-Lew Asian Self Identify Acculturation Scale (SL-ASIA; Suinn, Rickard-Figueroa, Lew & Vigil, 1987). A score of 1 on the SL-ASIA indicates an Asian-identified sample; a score of 5 indicated a Western-identified sample; and a score of 3 indicated a Bicultural-identified sample.

Factor analysis with orthogonal target rotation was conducted using Rentfrow et al. (2012) results as a reference, such that the target loadings were specified as non-zero for (absolute) primary factor loadings ≥ 0.4 (otherwise 0) as reported in Rentfrow et al. (2012). For instance, "Praying for Time" by Carey Sims had a primary loading (.65) on the *Unpretentious* factor and a secondary loading (.47) on the *Mellow* factor (Rentfrow et al., 2012; see Table 1). The target loadings for this item are specified to be zero for all, except for the *Unpretentious* factor. The target rotation procedure attempts to find a factor structure that is the closest to the specified target loadings. Subsequently, Tucker's congruence coefficient was calculated to quantify the degree of similarity between the factor structure of the current sample and that of Rentfrow et al. (2012). Factor analysis was conducted in Mplus (Version 8; Muthén & Muthén, 1998-2017) and Tucker's congruence coefficient was computed using the *psych* R package (Version 1.8.4; Revelle, 2018).

Results

A mean of 2.15 (*SD* = 1.17) was obtained on the SL-ASIA, indicating that this sample is Asian-identified. Bartlett's test of sphericity was significant at .05 level, indicating the items are not independent of one another. Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .66, indicating sufficient association between items to perform factor analysis.

Table 1. Results of factor analysis with orthogonal target rotation of the current sample as compared to Rentfrow et al. (2012)

			Factors									
			M		U		S		I		C	
Artist	**Piece**	**Genre**	SIN	RF	SIN	RF	SIN	RF	SIN	RF	SIN	RF
The O'Neill Brothers	Through the Years	Smooth Jazz	***.61***	***.74***	.28	.17	.40	.22	-.09	-.09	.05	.13
*Frank Josephs	Mountain Trek	R&B / Soul	***.28***	***.72***	*.60*	.18	.14	.17	-.24	-.05	-.03	.14
Human Signals	Birth	Soft rock	***.48***	***.67***	.34	.08	.27	.27	-.24	-.03	.07	.16
Bruce Smith	Children of Spring	Adult contemporary	***.74***	***.65***	-.06	.14	.27	.38	-.01	-.05	-.09	.01
Walter Rodriguez	Safety	Electronica	***.57***	***.59***	.15	.01	.08	.13	.09	.02	.29	.45
Language Room	She Walks	Soft rock	***.42***	***.54***	.17	.24	.00	.06	.23	.19	.09	.08
Ali Handal	Sweet Scene	Soft rock	***.46***	***.52***	.18	.38	.26	.31	-.14	.03	-.23	.01
Taryn Murphy	Love Along The Way	Soft rock	***.57***	***.50***	.25	.43	.03	-.02	.14	.11	.22	.16
*Kush	Sweet 5	Electronica	***.23***	***.50***	.08	.02	*.70*	.49	.08	-.06	.17	.29
James E. Burns	I'm Already Over You	New Country	.29	.30	***.73***	***.79***	.15	.08	-.04	-.04	.03	.03
Babe Gurr	Newsreel Paranoia	Bluegrass	.24	.13	***.51***	***.76***	.24	.18	-.06	-.04	.20	.03
Bob Delevante	Penny Black	New Country	.08	.25	***.63***	***.75***	.04	.13	.06	.01	-.08	.05
*Five Foot Nine	Lana Marie	Country-rock	*.55*	.34	***.38***	***.71***	.00	.10	.03	-.05	.10	.03
Curtis	Carrots & Grapes	Rock-n-roll	.09	-.05	***.25***	***.69***	.19	.29	.08	.14	.21	.02
Anglea Motter	Mama I'm Afraid to go There	Bluegrass	-.12	-.11	***.37***	***.65***	.27	.35	.06	.14	.30	.06

Carey Sims	Praying for Time	Mainstream country	.30	.47	***.62***	***.65***	-.03	-.02	-.14	.05	-.04	.08
Hillbilly Hellcats	That's Not Rockabilly	Rock-n-roll	-.13	-.11	***.46***	***.64***	.10	.27	.16	.03	.30	-.02
Laura Hawthorne	Famous Right Where I Am	Mainstream country	.28	.46	***.57***	***.62***	.14	-.12	.13	.01	-.14	.10
Ljova	Seltzer, do I Drink Too Much?	Avant-garde classical	-.02	.04	.02	.11	***.62***	***.82***	-.10	.06	.08	.07
Laurent Martin	Scriabin Etude Opus 65 No 3	Avant-garde classical	.16	.03	.04	.00	***.70***	***.76***	.01	.05	-.07	-.02
Antonio Vivaldi	Concerto in C	Classical	.06	.21	.26	.05	***.43***	***.75***	.12	.00	.08	-.06
Various Artists	La Trapera	Latin	.14	-.01	.12	.19	***.69***	***.75***	.14	-.04	.21	.15
Bruce Smith	Sonata A Major	Classical	.48	.31	-.04	.08	***.58***	***.70***	-.06	-.04	-.28	-.03
DNA	La Wally	Classical	.15	.18	-.01	.06	***.59***	***.69***	.22	.07	.09	.00
Paul Serrato & Co.	Who are You?	Traditional jazz	-.03	.07	.16	.10	***.65***	***.68***	-.09	.02	.37	.25
Moh Alileche	North Africa's Destiny	World beat	-.17	.06	.07	.20	***.48***	***.67***	.25	.01	.17	.07
Daniel Nahmod	I Was Wrong	Traditional jazz	.44	.34	-.02	.13	***.52***	***.64***	.05	-.04	.09	.23
Lisa McCormick	Fernando Esta Feliz	Latin	.00	.06	.26	.27	***.64***	***.63***	.05	-.03	.25	.25
AB+	Recess	Electronica	.14	.40	-.02	-.01	***.64***	***.54***	.13	-.04	.38	.34
Lisa McCormick	Let's Love	Adult contemporary	.35	.30	.03	.29	***.41***	***.51***	.05	.02	.20	.16
Bankrupt	Face the Failure	Punk	-.04	-.05	-.05	.00	.11	-.01	***.81***	***.85***	.17	-.02
Squint	Michigan	Punk	-.03	-.04	.08	.03	-.11	-.02	***.85***	***.83***	.13	-.06
Straight Outta Junior High	Over now	Punk	.05	-.06	-.11	.11	-.05	.05	***.83***	***.82***	.20	.00
Exit 303	Falling Down 2	Classic rock	.24	.15	-.02	.05	.12	-.01	***.70***	***.82***	.14	-.01
Five Finger Death Punch	Death Before Dishonor	Metal	.01	.08	.00	-.12	.12	-.02	***.74***	***.80***	.26	-.01
The Tomatoes	Johnny Fly	Classic rock	-.09	-.06	.06	.14	-.06	.04	***.73***	***.79***	.19	.01
Cougars	Dick Dater	Classic rock	-.14	-.15	.15	.21	-.15	.08	***.76***	***.76***	.23	.08
The Stand In	Frequency of a Heartbeat	Punk	-.05	.10	.17	.09	.20	.01	***.69***	***.75***	-.10	.07
Five Finger Death Punch	White Knuckles	Heavy metal	-.14	-.11	.05	-.12	.38	-.05	***.59***	***.74***	.01	.01
Dawn Over Zero	Out of Lies	Heavy metal	.09	.14	-.18	-.10	.11	-.03	***.79***	***.72***	.13	.10
Sammy Smash	Get the Party Started	Rap	.12	-.02	-.02	.13	-.06	-.09	.46	.06	***.60***	***.76***
Mykill Miers	Immaculate	Rap	.09	.08	.15	.03	.14	.10	.25	.08	***.63***	***.75***
Ciph	Brooklyn Swagger	R&B / soul	.14	-.10	-.18	.15	.01	-.05	.37	.07	***.63***	***.75***
Preston Middleton	Latin 4	R&B / soul	-.23	.16	.15	.04	.29	.23	.07	-.01	***.56***	***.73***
Robert LaRow	Sexy	Europop	-.34	.05	.10	.13	.23	.13	.27	-.06	***.51***	***.72***
Leo The Lionheart	Electro	Electronica	.17	.33	.05	-.10	.25	.20	-.11	-.04	***.57***	***.69***
DJ Come Of Age	Thankful	R&B/Soul	.30	.23	-.02	.22	.17	.03	.03	-.11	***.48***	***.62***
Magic Dingus Box	The Way It Goes	Electronica	.19	.35	.08	-.13	.32	.18	.14	.04	***.53***	***.52***
The Cruxshadows	Go Away	Europop	.20	.30	.12	-.21	.40	.25	.13	.08	***.38***	***.50***
Benjamin Chan	MATRIX	Electronica	-.18	.04	.22	-.23	.10	.15	.30	.34	***.49***	***.46***

Notes. **SIN**: factor loadings of current sample. Only factor loadings from orthogonal target rotation are included for SIN given that orthogonal and oblique rotations produced comparable results. **RF**: factor loadings of Rentfrow et al. (2012) derived by Principal Components Analysis with varimax rotation. *musical excerpts that showed a higher factor loading on other factors instead of the targeted factor

Table 2. Coefficients of congruence between current sample and Rentfrow et al. (2012)

	Rentfrow et al. (2012)				
Current sample	**Mellow**	**Unpretentious**	**Sophisticated**	**Intense**	**Contemporary**
Mellow	**.87**	.45	.39	-.02	.26
Unpretentious	.57	**.82**	.33	.06	.23
Sophisticated	.53	.29	**.93**	.08	.43
Intense	.00	.09	.09	**.93**	.25
Contemporary	.23	.20	.35	.27	**.89**

Notes. Range of .85-.94 corresponds to a fair similarity, a value higher than .95 indicates that the two factors or components compared can be considered equal (Lorenzo-Seva & Berge, 2006). The above matrix is not symmetrical due to the nature of the formulation of the coefficient.

Results of the factor comparisons showed that the five-factor structure in the current sample was congruent but not perfect with the factor structure in Rentfrow et al. (2012) (Table 1). Some musical excerpts showed higher factor loadings on another factor instead of the targeted factor. For instance, "Lana Marie" by Five Foot Nine (Country-rock) had a higher factor loading on *Mellow* factor instead of the targeted factor *Unpretentious*. Factor congruence indices of the current sample compared to Rentfrow et al. (2012) ranged from .82 to .93, with the lowest index being the *Unpretentious* factor (factor congruence = .82), and the highest indices being the *Sophisticated* and *Intense* factors (factor congruence = .93) (Table 2).

Discussion

The current study yielded a five-factor structure underlying music preferences in a Singaporean context. Factor analysis with orthogonal target rotation showed that the factors *Mellow*, *Sophisticated*, *Intense* and *Contemporary* have fair similarity with that of the original MUSIC model as reported in Rentfrow et al. (2012). Specifically, the factors *Sophisticated* and *Intense* yielded the highest factor congruence, while the *Unpretentious* factor showed the lowest factor congruence.

As Singapore is a multicultural society with substantial exposure to Western media, art and music, it is likely that similar factor structure underlying the *Sophisticated* and *Intense* factors emerged as Singaporean participants in the current study have comparable exposure to music genres comprising these two factors, such as classical music and rock genres. This is consistent with previous findings that listeners tend to prefer musical styles that they most closely identified with their own cultural background (Teo et al., 2008; Hui, 2009).

In contrast, listeners in Singapore do not frequently listen to country and folk music (National Arts Council, Singapore, 2017), which are music genres comprising the *Unpretentious* factor. It is therefore plausible that less familiarity with these music genres (Peterson, 2013), coupled with cultural differences in mood perception (Hu & Lee, 2012) contribute to lower factor congruence between the current sample and Rentfrow et al. (2012).

Contrary to prior studies that did not find a factor resembling the *Mellow* factor, we found that the current sample has a factor structure that has fair similarity as that of the *Mellow* factor. However certain musical excerpts showed a higher factor loading on other factors instead of the targeted factor *Mellow*. For instance, "Mountain Trek" by Frank Josephs (R&B/soul) had a higher factor loading on *Unpretentious*; "Sweet 5" by Kush (Electronica) had a higher factor loading on *Sophisticated* factor instead. As preference for internal and external properties of music was postulated to predict the location of a specific musical excerpt on the MUSIC model (Rentfrow et al., 2012), it is likely that the current sample showed different degree of liking for configurations of auditory and psychological attributes in these musical excerpts as compared to Western listeners.

While the current study provides preliminary support to extend the validity of the original MUSIC model to a non-Western context, it is not without limitations. The musical excerpts employed in this study only included genres from Western origins. A more comprehensive approach to study music preferences in Singapore would be to include a wider range of music genres that Singaporeans are exposed to, such as Chinese, Korean, Japanese popular music, as well as Ethnic music, including Malay and Tamil songs (National Arts Council, Singapore, 2017).

In addition, more detailed information about the characteristics of listeners, such as the frequency of listening to particular of music genres, cognitive styles when listening to music, and personality would be useful to elucidate the relationship between music preferences and listeners' characteristics.

Acknowledgements. This research is supported by the Nanyang Technological University Start-Up Grant: Learning with Emotions.

References

Browne, M.W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111-150. doi: 10.1207/S15327906MBR3601_05

Colley, A. (2008). Young people's musical taste: Relationship with gender and gender-related traits. *Journal of Applied Social Psychology, 38*, 2039–2055. doi: 10.1111/j.1559-1816.2008.00379.x

Cross, I. (2001). Music, cognition, culture, and evolution. *Annals of the New York Academy of Sciences*, *930*, 28–42. doi: 10.1111/j.1749-6632.2001.tb05723.x

Delsing, M. J. M. H, ter Bogt, T. F. M., Engels, R. C. M. E., & Meeus, W. H. J. (2008). Adolescents' music preferences and personality characteristics. *European Journal of Personality, 22*, doi:109–130.10.1002/per.665

Hu, X., & Lee, J. H. (2012). A cross-cultural study of music mood perception between American and Chinese listeners. *In Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, 535–540.

Hu, X., & Lee, J. H. (2014). Cross-cultural similarities and differences in music mood perception. *iConference 2014 Proceedings*, 259–269. doi: 10.9776/14081

Hui, W. V. (2009). Music listening preferences of Macau students. *Music Education Research, 11*(4), 485-500. doi: 10.1080/14613800903391749

Lorenzo-Seva, U., & Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2*(2), 57-64. doi: 10.1027/1614-2241.2.2.57

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, *535*(7613), 547–550. doi: 10.1038/nature18635

Muthén, L. K. & Muthén, B. O. (1998-2017). Mplus User's Guide. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

National Arts Council, Singapore. (2017). National Music Consumption Survey.

Peterson, R. A. (2013). Creating Country Music: Fabricating Authenticity. University of Chicago Press.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236–1256. doi: 10.1037/0022-3514.84.6.1236

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology*, *100*(6), 1139–1157. doi: 10.1037/a0022406

Rentfrow, P. J., Goldberg, L. R., Stillwell, D. J., Kosinski, M., Gosling, S. D., & Levitin, D. J. (2012). The song remains the same: A replication and extension of the MUSIC model. *Music Perception*, *30*(2), 161–186. doi: 10.1525/mp.2012.30.2.161

Revelle, W. (2018). psych: Procedures for Personality and Psychological Research [R package]. Retrieved from https://CRAN.R-project.org/package=psych Version = 1.8.4

Schäfer, T., & Sedlmeier, P. (2009). From the functions of music to music preference. *Psychology of Music*, *37*(3), 279–300. doi: 10.1177/0305735608097247

Schäfer, T., Tipandjan, A., & Sedlmeier, P. (2012). The functions of music and their relationship to music preference in India and Germany. *International Journal of Psychology*, *47*(5), 370–380. doi: 10.1080/00207594.2012.688133

Suinn, R. M., Rickard-Figueroa, K., Lew, S., & Vigil, P. (1987). The Suinn-Lew Asian Self-Identity Acculturation Scale: An Initial Report. *Educational and Psychological Measurement*, *47*(2), 401–407. doi: 10.1177/0013164487472012

Teo, T., Hargreaves, D. J., & Lee, J. (2008). Musical preference, identification, and familiarity: A multicultural comparison of Secondary students from Singapore and the United Kingdom. *Journal of Research in Music Education, 56*(1), 18-32. doi: 10.1177/0022429408322953

Tucker, L. R. (1951). A method for synthesis of factor analysis studies (Personnel Research Section Report No. 984). Washington, DC: Department of the Army.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Influx of Different Language Rhythms and Cultures into Musical Rhythms because of the Occupation by Other Countries

Tomomi Hida [1]

[1] *Graduate School of Informatics, Kyoto University, Japan*

[1] hida.tomomi@cog.ist.i.kyoto-u.ac.jp

Abstract

It has been found that musical rhythms are often influenced by language rhythms in the composers' native language. The objective of this study was to test whether Japanese history has influenced the Japanese native speaker musical compositions. As Japan was occupied by the United States after World War II, many languages and cultures were imported. In this study, Japanese musical phrases composed after 1800 were evaluated and the index values (nPVI; normalized Pairwise Variability Index) were calculated. It was found that Japanese musical rhythms changed significantly after it was occupied by the United States; however, it was also observed that Japanese musical rhythms differed from one historical period to another historical period. It was concluded that complex historical and cultural influences from other countries have indeed affected Japanese musical rhythms.

Introduction

There has been an ongoing discussion on the relationship between language and music. In 1871, Darwin proposed the "musical protolanguage" hypothesis (see Fitch, 2010 for an overview). The Generative Theory of Tonal Music (GTTM) (Lerdahl & Jackendoff, 1983) applied linguistic analysis to music, which dramatically changed music analysis. More recently, Evolutionary Developmental Biology (Evo-Devo) scholars have also shown more interested in this field (Fitch, 2010; Fitch & Martines, 2014; Masataka, 2007; Patel, 2010). It has been found in many studies that linguistic skills are strongly related to musical ability and vice versa (Aleander et al., 2005; 2008; 2011; Anvari et al., 2002; Chan et al., 1998; Marie et al., 2011; Milovanov et al., 2004; 2007; 2008; 2009; 2010; Parberry-Clark et al., 2009; Perfors & Ong, 2012; Piro & Oriz, 2009; Sadakata & Sekiyama, 2011; Shabni & Torkrh, 2014; Slater et al., 2014; Slevc & Miyake, 2006; Skoe & Kraus, 2012).

Rhythm is one of the most important elements in both music and language (Patel, 2002; 2003; 2008). Language rhythm can be categorized into stress-timed languages, syllable-timed languages, and mora-timed languages (Abercrombie, 1967; Bloch, 1942; Ladefoged, 1975; Ladefoged & Johnson, 2010; Pike, 1945). Grabe and Low (2002) used a pairwise variability index (PVI) to categorize these three linguistic speech rhythms. PVI gives the average value of the differences between the vocalic intervals or the intervocalic intervals in a sentence and is calculated from the differences in the durations between the adjacent vocalic intervals or the intervocalic intervals. The PVI values in stress-timed languages tend to be greater as the vowels in the stressed syllables are longer and the vowels in the unstressed syllables are shorter. However, the PVI values in syllable-timed languages tend to be smaller as the variations in vowel durations are relatively smaller.

Table 1. nPVI for Each Language (Based on Grabe & Low, 2002)

Language	nPVI	Language rhythm
Dutch	65.5	
German	59.7	Stress-timed
British English	57.2	
Japanese	40.9	Mora-timed
Spanish	29.7	Syllable-timed
Mandarin	27.0	

Patel and Daniele (2003) adapted the nPVI (normalized PVI) language measurement to music and found that the nPVI value for French (a syllable-timed language) was similar to the nPVI value in pieces by French composers and the nPVI in English (a stress-timed language) was also similar to the nPVI in pieces by British composers. These results suggested that native language rhythms influenced native musical rhythms. Jekiel (2014) compared the nPVI values in British English and Polish (a mixed stress-timed/syllable-timed language) pieces composed in the 19th century; however, the results were not significant. It was therefore surmised that Poland's history may have been a factor. First, Poland had been divided into the Russian Empire, Preussen, and Austria in the 18th century, and in the early 19th century, a Polish Duchy, a French-bloc country was founded by Napoléon Bonaparte, which meant that Polish language rhythm had been influenced by Russian (a stress-timed language) and French (a syllable-timed language) through the various occupations. Therefore, it made sense that it was difficult to find significant differences between British English and Polish musical pieces as the influx of other languages had influenced Polish musical rhythms. In the same manner, as Japan was occupied by the United States after World War II (WWII)/the Pacific War (1945), language rhythms and musical rhythms may have been affected by history. A recent study also found that history impacted musical rhythms (Daniele & Patel, 2013, 2015).

In this study, two hypotheses about Japanese language and musical rhythms are proposed: (a) the composers' native language (Japanese) influences musical rhythms, and (b) Japan's history of the US occupation influenced more recent musical rhythms. To assess the validity of these hypotheses, the musical nPVIs of native Japanese compositions were calculated and analyzed.

Method

Musical Materials

One thousand and fifty-nine phrases from 220 pieces composed by 104 native Japanese speakers were collected (the mean number of notes per phrase = 16.19, SD = 6.11, for more details about the composers, see Supplemental Table 1). All composers were born after 1800, and 562 of the 1059 phrases were taken from 136 pieces that also had lyrics (79 composers, mean number of notes per phrase = 14.70, SD = 4.37), and 497 of the 1059 phrases were taken from 85 instrumental pieces (30 composers, mean number of notes per phrase = 17.87, SD = 7.27). Each phrase had at least 10 notes and no internal pauses/rests; grace notes and ornamentations were omitted from the calculation.

In this study, the musical nPVI values for the Japanese pieces with and without lyrics were not found to be significantly different (Fig. 1, Song-Instrumental: t-test, p = 0.85, Cohen's d = 0.005) and the mean for the nPVI values in the Japanese music phrases were similar (mean total: nPVI = 40.99, Song: 41.11, Instrumental: 40.86); therefore, the "Song" and "Instrumental" pieces were not distinguished in the advanced analyses.

The Japanese language nPVI value identified by Grabe and Low (2002) came from "modern Japanese." Therefore, modern musical and modern Japanese language nPVI values were compared. Of the 159 musical phrases from composers born after 1950, the results were found to be similar (mean nPVI: 44.53, SD = 18.15) and the differences between the musical nPVI values with and without lyrics were not significant, with the nPVI values being very similar to the linguistic nPVI reported in Grabe and Low (2002). Therefore, the materials in the present study confirmed that Japanese linguistic speech rhythms affected Japanese composed musical rhythms regardless of whether or not there were lyrics.

Measurement of the Musical nPVI

The rPVI is the row PVI, defined as:

$$\mathrm{rPVI} = \left[\sum_{k=1}^{m-1} |d_k - d_{k+1}|/(m-1)\right],$$

where m is the number of intervals, and d is the duration of the kth item. The nPVI is the normalized rPVI, which is defined as:

$$\mathrm{nPVI} = 100 \times \left[\sum_{k=1}^{m-1} \left|\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2}\right| /(m-1)\right],$$

where m is the number of items in an utterance, and d is the duration of the kth interval.

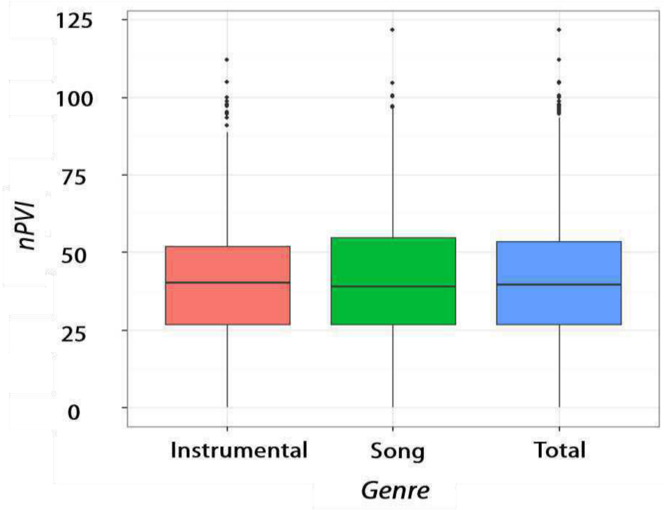


Figure 1. Musical nPVI values for three groups (Instrumental, Song, and Total; 0.05 < p; n.s.): the black dots are the outliers, the bottom of each box is the lower quartile, the top is the upper quartile, and the line in the middle is the median; the upper and lower whiskers are drawn.

All durations between the notes in the phrases were counted: a quarter note was counted as "one" in four-four time phrases, and an eighth note was counted as "one" in six-eight time phrases (Figure 2). To simplify the calculations, Patel and Daniel (2003) set the first note in each phrase as "one" regardless of the meter when they counted the note duration; that is, both calculation results were the same. The nPVI calculator (The Neuroscience Institute, California, http://www.nsi.edu/~ani/nPVI_calculator.html) was used to calculate the nPVI values from the counted note durations using formula (2). R (Ihaka & Gentleman, 1996) and EZR (Kanda, 2013) were used for the statistical analysis.

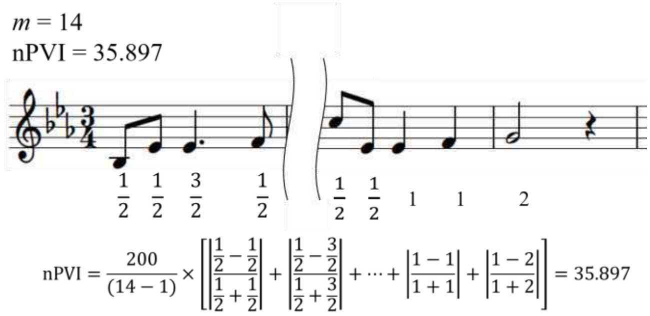


Figure 2. Example of a musical nPVI calculation ("Akatombo (Red Dragonfly)" composed by Kosaku Yamada): where m is the number of notes in the phrase. As this phrase is three-four time, each note was counted as follows: a crotchet was 1, a semiquaver was 1/2, a minim was 2, and a dotted crotchet was 3/2.

Background Analysis

Context of history in Japan. The years between 1800 (the Edo period) and 2015 (the Heisei period) were broken down into eight periods based on *gengo* (i.e., era names; most of them are changed when a new emperor ascends the throne): "Edo (the late Edo period, 1800–1868)," "Meiji A (the early Meiji period, 1868–1890)," "Meiji B (the late Meiji period, 1890–1912)," "Taisho (1912–1926)," "Showa A (from pre-WWII to wartime: 1926–1945)," "Showa B (the Allied Occupation after the WWII/the Pacific War: 1945–1952)," "Showa C (post-WWII: 1953–1989)," and "Heisei (1989–present)."

The late Edo period (Edo, 1800–1868). The "Tokugawa shogunate" was the period between 1603 and 1868 when Japan was under Tokugawa rule. The third shogun, Tokugawa Iemitsu, issued a National Isolation Edict from 1633 to 1639 (known as *sakoku* or the "locked country"), which was maintained until the arrival of the Black Ships of Commodore Matthew Perry in 1853. Therefore, during this period, no foreigners could enter Japan and no Japanese could travel to or return from foreign countries; however, the Netherlands had trade privileges in Dejima, Nagasaki. Japan opened its doors to the world in 1854 when it realized the necessity to establish a modern state government to resist Western influences. The last and fifteenth shogun, Tokugawa Yoshinobu, stepped down, and Emperor Meiji came to the throne in 1867.

The Meiji (Meiji A and B: 1868–1912) and Taisho periods (Taisho, 1912–1926). The Meiji period began in 1868, at which time the samurai warrior class was disbanded, Japan moved toward democratization, and Western culture began to influence Japan (Meiji A). In the latest Meiji period (Meiji B), the Sino-Japanese War and the Russo-Japanese War occurred as Japan began to expand its territories. After the Meiji period, Emperor Taisho took over the throne from 1912 to 1926 (Taisho), and as Japanese culture and Western culture began to merge, a new Japanese culture emerged.

The Showa (Showa A, B and C: 1926–1989) and Heisei periods (Heisei, 1989–present). The Showa period started in 1926. The Japanese empire entered war and launched the Pacific War in 1941. Consequently, Japan was one of the defeated countries and was occupied by the GHQ (General Headquarters) and the offices of the Supreme Commander for the Allied Powers (SCAP) in 1945 (Showa A: pre-WWII-wartime), with the main occupier being the United States. As many Americans were living in Japan, they brought their culture with them. The Treaty of San Francisco was concluded in 1952, which ended the occupation of Japan by the United States, and Japan gained sovereignty (Showa B: the Allied Occupation). Subsequently, the cultures that had been introduced in the period of occupation by the United States became naturalized. The Showa period ended (Showa C: post-WWII) and the Heisei period started in 1989 (for details, see Totman, 2014).

Context of musical history in Japan. The Japanese musical history classification by Yoshikawa (1965) was referred as: "Before the arrival of the Black Ships by Matthew Perry (1800–1852)," "the period of imported Western music (1853–1912)," "the period of digested Western music (1912–1945)," and "the period of the rising of folk music (1946–present)" (author's translation into English).

Western music was imported from 1853–1912. After the arrival of the Commodore Matthew Perry's Black Ships in 1853, Japanese music began to be influenced by Western music and Japanese traditional music began to decline, with many *koto* (Japanese traditional stringed musical instrument) pieces barely surviving. The *Rokumeikan* ("Dear-cry Hall") was built in 1883 to wine and dine state guests and diplomats from foreign countries, and high-ranking Japanese officials began to practice Western manners there for the first time, which became famous for its parties and balls. Therefore, the opportunities to perform Western music increased and Japanese sheet music was also westernized.

The digested Western music (1912–1945) in the Taisho period and the rise in folk music (1946–present) in the early Showa period (1912–1945) saw developments and improvements in Japanese traditional instruments; however Western music provided the initiative. The new words *haikara* and *bankara* were created; *haikara* stemmed from the English phrase "high color" and indicated a Westernized lifestyle, while *bankara* came from the Japanese word *yaban* ("barabarism") and was the antithesis of *haikara*. In music and the arts, *haikara* referred to Western cultural influences such as the violin and moving pictures ("*katsudo shashin*"), whereas *bankara* referred to traditional Japanese music and the arts such as *Kabuki* (classical Japanese drama) and the Shamisen (Japanese traditional instrument). However, at this time *bankara* was in decline. After the Great *Kanto* area earthquake in 1923, the radio was introduced and popular songs spread among the common people with the introduction of records and record players.

In "the period of the rising of folk music," folk (classical) music underwent a renaissance, and sensationalism morphed into intellectualism. In addition, music that dispelled stereotypes and was iconoclastic spread among the Japanese.

Statistical Analysis

The Brown–Forsythe test, a one-way Analysis of Variance (ANOVA), Tukey's HSD test, the Kruskal–Wallis test, and the Steel–Dwass test (Dwass, 1960; Steel, 1960) were performed as statistical analyses using R and MATLAB (Mathworks Inc., Natick MA). The Brown–Forsythe test was used to determine whether the group variances were equal or not, and if the results indicated that the group variance was equal, a one-way ANOVA was used; when the group variance was not equal, the Kruskal–Wallis test was used. The Tukey's HSD test is a subordinate test of the one-way ANOVA, and the Steel–Dwass test is a subordinate test of the Kruskal–Wallis test. Judging the results of the Brown–Forsythe test, the Kruskal–Wallis test was used for the analysis of the context of history in Japan ($F(7, 1051) = 4.10, p < 0.001$), and the one-way ANOVA was used for the analysis of the context of musical history in Japan ($F(3, 1055) = 1.99, p = 0.11$)

Permutation (randomized) test. After each test, the data sets were automatically shuffled into eight and four groups randomly to break the time structure, and a one-way ANOVA or Kruskal–Wallis test was performed 1000 times in MATLAB to estimate the statistical thresholds of each empirical distribution of the test statistics to confirm whether the results were affected by historical effects or not. If the statistical thresholds were lower than the value in the exact grouping, the effects were a result of historical events.

Results

Figure 3 shows the results for the musical nPVI measurements for Japanese music in the different historical periods (Kruskal–Wallis test: $\chi^2 = 16.21$, $p = 0.001$, $\eta^2_G =$

0.02), and the value of the test statistic was higher than its permutation test ($p < 0.0001$). The Steel–Dwass test showed the effect of the influx of other languages and cultures after the "locked country" period and blended with the original Japanese language (Meiji A-Showa A: $p = 0.040$, Cohen's $d = 0.36$), the effects of the influx of other languages and cultures in the United States occupation (Taisho-Showa B: $p = 0.015$, Cohen's $d = 0.25$; Showa A-Showa B: $p = 0.0002$, Cohen's $d = 0.38$), and the effect of the merging of the languages and cultures after the war and the original Japanese (Showa B-Showa C: $p < 0.0001$, Cohen's $d = 0.47$; Showa B-Heisei: $p = 0.0001$, Cohen's $d = 1.05$). Additionally, the musical rhythms were different from they were before (Edo-Heisei: $p = 0.015$, Cohen's $d = 1.11$; Meiji A-Showa C: $p < 0.0001$, Cohen's $d = 0.41$; Meiji A-Heisei: $p = 0.003$, Cohen's $d = 0.84$; Meiji B-Heisei: $p = 0.029$, Cohen's $d = 0.77$). It was concluded that historical events had a significant influence on Japanese musical rhythms at around the US occupation.

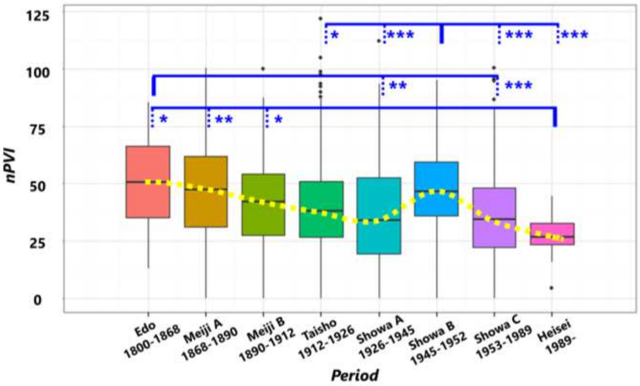


Figure 3. Musical nPVI values for historical periods: the black dots are the outliers, the bottom of each box is the lower quartile, the top is the upper quartile, and the line in the middle is the median; upper and lower whiskers are drawn. Blue lines are the results of the Steel–Dwass test. Solid lines are the starting point compared to each connected dashed line (*$p < 0.001$, ** $0.001 < p < 0.010$, * $0.010 < p < 0.050$)**

Figure 4 shows the results of the musical nPVI measurements for Japanese music from period to period (one-way ANOVA: $F(3,1055) = 5.51$, $p < 0.001$, $\eta^2_G = 0.02$), with the F-value being higher than its permutation test ($p < 0.0001$). Tukey's HSD test showed the changes in the times and the musical rhythms in Japan (the period of imported Western music- the period in which Western music was digested: $p = 0.006$, Cohen's $d = 0.22$; and the period of imported Western music-the period of the rising of folk music: $p = 0.028$, Cohen's $d = 0.18$).

Conclusion

Japanese musical nPVI values changed because of the cultural influences of other countries and particularly because of the US occupation. The results supported the hypotheses that Japan's complex historical background influenced musical rhythms, which was inconsistent with the findings in Jekiel (2014).

In this research, the nPVI value in the Edo period was found to be high even though this period included the period of the "locked country." It is presumed that most of the data in this period was after the opening of the country in 1853,

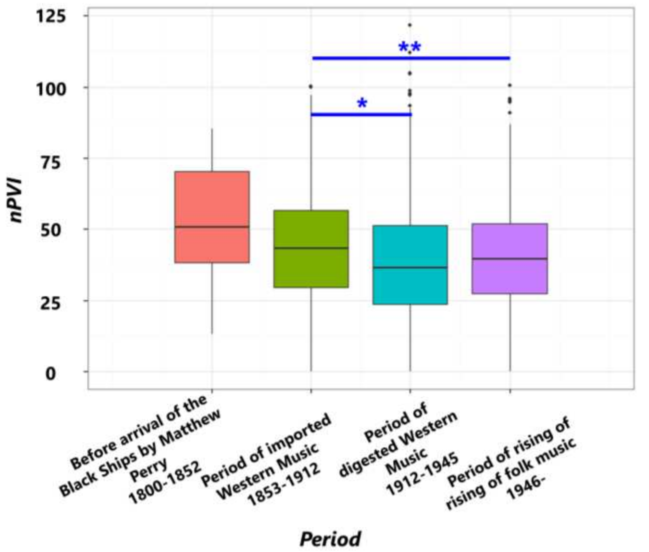


Figure 4. Musical nPVI values for music historical periods: the black dots are the outliers, the bottom of each box is the lower quartile, the top is the upper quartile, and the line in the middle is the median; and the upper and lower whiskers are drawn ($0.001 < p < 0.010$, * $0.010 < p < 0.050$).**

and therefore, it might have been influenced by other languages.

London and Jones (2010) noted that "the relation between musical and linguistic rhythm seems more subtle and complex than that proposed by Patel & Daniele." In particular, the culture has significantly changed many times in the last 200 years in Japan; the "Galapagosization" (the process of the isolation) or "locked country," the importing of new cultures from other countries as Japan opened to the world, the merging of Japanese culture with others, the importing of American culture as a result of the US occupation, and the restoration of sovereignty; have all affected the musical nPVI values.

Therefore, researchers need to consider the historical and cultural background of countries when studying the relationships between language and music.

Acknowledgments. Thanks are due to Dr. Hiroaki Mizuhara, Graduate School of Informatics at Kyoto University, for offering me constructive comments, especially statistical analysis, and for his warm encouragement.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Aldine: Chicago.

Alexander, J. A., Bradlow, A. R., Ashley, R. D., & Wong, P. C. (2011). Music-melody perception in tone-language and non-tone-language speakers. In *Psycholinguistic representation of tone*. Hong Kong.

Alexander, J. A., Bradlow, A. R., Ashley, R. D., & Wong, P. C. M. (2008). Music melody perception in tone-language and nontone-language speakers. *Journal of the Acoustical Society of America*, *124*, 2495.

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. In *Proceedings of the*

9th European Conference on Speech Communication and Technology, 397-400, Lisbon, Portugal.

Anvari S. H., Trainor L. J., Woodside J., & Levy B. Z. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, *83*, 111-130.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language, 26*, 86-125.

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature*, *396*, 128.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.

Dwass, M. (1960). Some k-sample rank-order tests. *Contributions to probability and statistics: essays in honor of harold hotelling* (pp. 198-202), Stanford University Press.

Daniele, J. R., & Patel, A. D. (2013). An empirical study of historical patterns in musical rhythm: Analysis of German and Italian classical music using the nPVI equation. *Music Perception*, *31*, 10-18.

Daniele, J. R., & Patel, A. D. (2015). Stability and change in rhythmic patterning across a composer's lifetime: A study of four famous composers using the nPVI equation. *Music Perception*, *33*, 255-265.

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Fitch, W. T. & Martins, M. D. (2014). Hierarchical processing in music, language, action: Lashley revisited, *Annals of the New York Academy of Sciences*, *1316*, 876-104.

Grabe, E. & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Warner, & C. Gussenhoven (Eds.), *Papers in laboratory phonology*, *7* (pp. 515-546). Berlin: Mouton de Gruyter.

Han, M. S. (1962). The feature of duration in Japanese. *Onsei no kenkyuu, 10*, 65–80.

Ihaka, R. & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*, 299-314. Available via http://www.R-project.org.

Jekiel, M. (2014). Comparing rhythm in speech and music: The case of English and Polish. *Yearbook of the Posnań Linguistic Meeting*, *1*, 55-71.

Kanda, Y. (2013). Investigation of the freely available easy-to-use software 'EZR' for medical statistics. *Bone Marrow Transplantation*, *48*, 452-458.

Ladefoged, P. (1975). *A course in phonetics*. New York: Harcourt Brace Jovanovich.

Ladefoged, P. & Johnson, K. (2010). *A course in phonetics* (International ed. of 6th revised ed.). Wadsworth: Cengage Learning.

London, J. & Jones, K. (2010). Metrical hierarchies and musical *nPVI*: A re-analysis of Patel and Daniele. In *Proceedings of the 11th International Conference on Music Perception and Cognition*, Seattle, 379-340.

Marie, C., Delogu, F., Lapis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–2715.

Masataka, N. (2007). Music, evolution, language. *Developmental Science*, *10*, 35-39.

Milovanov, R., Tervaniemi, M., & Gustafsson, M. (2004). The impact of musical aptitude in foreign language acquisition. S. D. Lipscomb, R. Ashley, R.O. Gjerdingen, P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition (ICMPC8)*, Causal productions, Adelaide, Australia (2004), 717-718.

Milovanov, R., Huotilainen, M., Esquef, P. A. A., Välimäki, V., & Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration determination in school-aged children. *Neuroscience Letters*, *460*, 161-165.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A. A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research*, *1194*, 81-89.

Milovanov, R., Pietilä. P., Tervaniemi, M., & Esquef, P. A. A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences*, *20*, 56-60.

Milovanov, R., Tervaniemi, M., Takio, F., & Hämäläinen, H. (2007). Modification of dichotic listening (DL) performance by musico-linguistic abilities and age. *Brain Research*, *1156*, 168-173.

Parberry-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience*, *29*, 14100-14107.

Patel, A. D. (2003). Rhythm in language and music: Parallels and differences. *Annals of the New York Academy of Sciences*, *999*, 140-143.

Patel. A. D. (2006). Musical rhythm, linguistic rhythm, and human evolution. *Music Perception*, *24*, 99-104.

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press: New York.

Patel, A. D. (2010). Music, biological evolution, the brain. In M. Bailar (Ed.), *Emerging disciplines* (pp. 91–144), Houston, TX: Rice University Press.

Patel, A. D. & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*, B35-B45.

Perfors, A. & Ong, J. H. (2012). Musicians are better at learning non-native sound contrasts even in non-tonal languages. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society, 839-844.

Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Piro, J. M. & Ortiz, C. (2009). The effect of piano lessons on the vocabulary and verbal sequencing skills of primary grade students. *Psychology of Music*, *37*, 325-347.

Sadakata, M. & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica*, *138*, 1-10.

Shabani, M. B. & Torkrh, M. (2014). The relationship between musical intelligence and foreign language learning: The case of Iranian learners of English. *International Journal of Applied Linguistics & English Literature*, *3*, 26-32.

Skoe, E. & Kraus. N. (2014). A little goes a long way: How the adult brain is shaped by musical training in childhood. *The Journal of Neuroscience*, *32*, 11507-11510.

Slater, J., Strait, D. L., Skoe, E., O'Connell, S., Thompson, E., & Kraus, N. (2014). Longitudinal effects of group instruction on literacy skills in low-income children. *PLoS ONE*, *9*, e113383.

Slevc, L. R. & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science 2006*, *17*, 675-681.

Steel, R. G. D. (1960). A rank sum test for comparing all pairs of treatments. *Technometrics*, *2*, 197-207.

Totman, C. (2014). *A history of Japan* (2nd ed.). Wiley-Blackwell.

Yoshikawa, E. (1965). *Nihon ongaku no rekishi* (History of Japanese Music) [in Japanese]. Sogensha: Osaka

Appendix

Supplemental Table 1. Composer information of musical materials (for details of pieces, see http://www.hida-net.jp/downloads/icmpc15_suppleInfo.pdf)

Name of Composers	*Dates Lived*	*Name of Composers*	*Dates Lived*	*Name of Composers*	*Dates Lived*	*Name of Composers*	*Dates Lived*
Yoshizawa Kengyo	1800-1872	Simizu Osamu	1911-1986	Uchida Masato	1940-1997	Okamura Takako	1962-
Hayashi Hiromori	1831-1896	Ifukube Akira	1914-2006	Murai Kunihiko	1945-	Fujii Fumiya	1962-
Koyama Sakunosuke	1864-1927	Otaka Hisatada	1914-2006	Oda Kazumasa	1947-	Hotei Tomoyasu	1962-
Okano Teiichi	1878-1941	Koyama Kiyoshige	1914-2009	Kato Kazuhiko	1947-2009	ak.homma	1964-
Taki Rentaro	1879-1903	Ishiketa Mareo	1915-1996	Inoue Yosui	1948-	Tortoise Matsumoto	1966-
Atarashi Seijiro	1880-1970	Isobe Toshi	1917-1998	Kako Takashi	1948-	Inoue Yoshimasa	1966-
Komatsu Kosuke	1884-1966	Nakada Yoshinao	1923-2000	Serizawa Hiroaki	1948-	Hirose Komi	1966-
Motoori Nagayo	1885-1945	Ohnaka Megumi	1924-	Hisaishi Jyo	1950-	Tachikawa Toshiyuki	1966-
Yanada Tadashi	1885-1959	Dan Ikuma	1924-2001	Hoshi Katsu	1948-	Okui Kaori	1967-
Nakada Akira	1886-1931	Akutagawa Yasushi	1925-1991	Ise Syozo	1951-	Yamazaki You	1967-
Yamada Kosaku	1886-1965	Eguchi Hiroshi	1927-2010	Chiba Kazuomi	1951-	Hakase Taro	1968-
Nobutoki Kiyoshi	1887-1965	Okumura Hajime	1929-	Sada Masashi	1952-	Igarashi Michiru	1969-
Funabashi Eikichi	1889-1932	Mamiya Michio	1929-	Hamada Shogo	1952-	Fukuyama Masaharu	1969-
Sugiyama Haseo	1889-1952	Izumi Taku	1930-1992	Ogikubo Kazuaki	1953-	Fukushima Hirokazu	1971-
Matsushima Tsune	1890-1985	Kobayashi Hideo	1931-	Yoshimatsu Takashi	1953-	Suemitsu Atsushi	1971-
Kashima Meisyu	1891-1954	Takemitsu Toru	1930-1996	Sakamoto Ryuichi	1953-	Fukuda Yosuke	1975-
Hirota Ryutaro	1892-1952	Nakamura Hachidai	1931-1992	Matsutoya Yumi	1954-	Kobuchi Kentarou	1977-
Narita Tamezo	1893-1945	Hayashi Hikaru	1931-2012	Kuwata Keisuke	1956-	Kitagawa Yujin	1977-
Ohno Tadasuke	1895-1929	Hagihara Hidehiko	1933-2001	Mitake Akira	1956-	Mashiko Tatsuro	1978-
Onaka Toraji	1896-1982	Hara Hiroshi	1933-2002	Nagura Akira	1957-1980	Maeyamada Ken'ichi	1980-
Ohkawa Hachiro	1901-	Yuyama Akira	1933-2002	Oda Tetsuro	1958-	Nakata Yasutaka	1980-
Hashimoto Kunihiko	1904-1949	Miyoshi Akira	1933-2013	Nakamura Masato	1958-	Ai	1982-
Takagi Touroku	1904-2006	Kubota Satoshi	1935-	Funky Sueyoshi	1959-	Mizuno Yoshiki	1982-
Taku Koji	1904-2006	Kaneda Bin	1935-2002	Ito Shintaro	1960-	Fukase	1985-
Hattori Tadashi	1908-2010	Sato Toshinao	1936-2002	Matsumoto Takahiro	1961-	miwa	1990-
Hirai Kouzaburo	1910-2002	Ikebe Shin'ichiro	1943-	Ohshima Michiru	1961-	Sumida Shinya	1994-

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Music in the body: Does music listening influence the reality of pain? A scoping review.

Claire Howlin,[1] Darragh Lynch,[2] Suzanne Guerin[3] and Brendan Rooney[4]

[1,] *School of Psychology, University College Dublin, Ireland*

[1]claire.howlin@ucdconnect.ie, [2]darraghkevinlynch@gmail.com, [3]suzanne.guerin@ucd.ie, [4]brendan.rooney@ucd.ie

Abstract

Music listening interventions (MLIs) significantly reduce pain experience and opioid intake and are widely implemented in pain medicine, obstetrics, neurological dysfunction and palliative care. However, in the absence of a music therapist or music creation, it is not yet clear how MLIs can reduce people's experience of pain. The aim of this study is to explore the proposed cognitive principles underpinning MLIs in the context of analgesic effects, by using thematic synthesis to describe the theoretical underpinnings of MLIs. A systematic literature search was conducted using the three search terms; music, listening and pain, in four electronic databases EBSCO Music Index and RILM, and EBSCOhost Psychology and Behavioural Sciences Collection, CINAHL Plus, Pubmed. In total 474 documents were retrieved across the four databases, which were carefully assessed by two reviewers for inclusion according to a strict inclusion, exclusion and quality appraisal criteria, which left 75 articles to be included in the qualitative thematic synthesis. Together the four identified analytical themes; Intramusical Features Directly Command Attention, Cognitive Agency of the Listener, Multidimensional Integration, and Cognitive Strengthening; provide a more comprehensive understanding of the cognitive principles underpinning MLIs in pain contexts compared to any one singular approach. Future MLI development should support the listener as an active agent within the MLI and reduce competing material such as noises or pain relevant stimuli in order to facilitate optimal engagement.

Introduction

Music Listening Interventions (MLIs) are now being implemented in a wide range of analgesic contexts, including chronic pain, surgical recovery, routine procedures, palliative medicine and even during labour (Mainka, Spingte, & Thaut, 2016).

However, although some studies have attempted to standardise MLIs (Bradt, 2012), with 15-minute sessions being advocated (Chen, Wang, Shih, & Wu, 2013), there is a huge level of variability in how MLIs are delivered (Bradt, Dileo, Grocke, & Magill, 2011). For example, Clark (2006) asked patients to listen to a personalised tape whenever they felt like it, with optional muscle relaxation and visual imagery techniques; which ultimately led to some patients not listening to the music, making evaluation impossible. Furthermore, this variation in how MLIs are delivered is reflected in low to moderate effect sizes in meta-analysis, which highlights that there are times when MLIs work and there are times when MLIs don't work.

One of the reasons that there is so much variability in how MLIs are delivered is because the underlying mechanisms of MLIs in the absence of a music therapist are poorly defined and need to be understood more clearly (Fancourt Ockelford, & Belai, 2014; Keenan & Keithley, 2015; Koelsch, 2015; Krishnaswamy & Nair, 2016; Lee, 2016).

So far, several lines of enquiry have helped to shed light on how MLIs work. From a neurochemical perspective, the endogenous opioid circuit is activated by music, leading to the theory that MLIs reduce pain because they encourage the release of endogenous opioids (Jeffries, Fritz, & Braun, 2003). Recently this has been supported through meta-analysis which demonstrates a reduced intake of opioid-based medication as a result of music listening. Initial attempts to maximize the benefits of MLIs emphasized that music with a slow tempo and an absence of strong rhythms should be used, with the rationale that people would entrain to this tempo which would directly lower their physiological arousal. Subsequently it has been recognized that physiological responses to the same piece of music tends to be idiosyncratic between individuals due to a number of extramusical factors, which undermines the idea that specific types of music will lead to uniform physiological responses across individuals. Accordingly, the benefits of MLIs seem to be much stronger when the music is chosen by the individual, compared to music chosen by the experimenters (Lee, 2016), which further implicates the role of individual musical engagement to achieve any physiological effects. In line with this, several studies have emphasised the role of distraction in MLIs, advocating MLIs as an attentional based pain management strategy rather than a direct physiological (Kwekkeboom, 2003; Kemper & Danhauer, 2005; Koelsch, 2012). This suggests that attentional differences to preferred music may underpin successful MLIs and is reflected by patient self-reports, with 65% of patients claiming that music distracts them from their pain (Huang et.al., 2010).

In order to clarify the cognitive principles that underpin MLIs, the current study will examine what is known from the existing literature about the characteristics of MLIs used in analgesic settings. This will be done using a systematic scoping literature review, which provides more breadth and depth than is achievable with meta-analysis. The primary objective of this scoping review is to explore the proposed cognitive principles underpinning MLIs in the context of analgesic effects, by conducting thematic synthesis on the provided rationales in MLI studies and describing the theoretical underpinnings of MLIs. This will help identify what is known from the existing literature about the characteristics of effective music listening interventions used in analgesic settings and will contribute to future guidance on optimal music listening intervention design.

Method

This study comprises a systematic scoping literature review using a descriptive-analytical method in line with previous guidelines (Arksey & O'Malley, 2005; Levac, Colquhoun, & O'Brien, 2010) and the protocol is outlined in full in a previous publication (Howlin, Lynch, Guerin, & Rooney 2018). This review involves five stages, (i) development of the research question, (ii) identifying relevant studies, (iii) study selection (iv) Data extraction and analysis, and finally, (iv) summarizing and reporting the results. The research team consisted of four psychologists with expertise in music psychology, health psychology, cognitive psychology and research methodology.

Search Strategy

A university librarian with expertise in systematic literature reviews was consulted to assist with developing the search strategy and designing a sensitive search string. Three search terms; specifically, music, listening and pain (see table 1), were used as the core terms to develop a three-pronged search string, along with their variations and MESH terms. First direct synonyms of each term known to the research team were added to the string, which was then refined by the research librarian. Second each term was used to find additional keywords or MESH terms specific to each database and added to the string using the database search builders. Thirdly, each string was then searched for separately, before combining the searches. This procedure was followed, using identical limiters, expanders and date range for each database to ensure consistency in the resultant search string produced. Each search was limited to include English articles as this was the only language of the research team, peer reviewed journals, and studies where the abstract was available for further examination. Additionally, the search was limited to include articles with available abstracts, published between January 1st 2006 and June 15th 2017. This was considered necessary to ensure that the review could be completed within the planned time frame. A systematic search was conducted using a subgroup of the electronic databases included in the study by Lee (2016); EBSCO Music Index and RILM, and EBSCOhost Psychology and Behavioural Sciences Collection, CINAHL Plus (EBSCO), Pubmed (Medline). These databases were chosen as they include music listening studies in a range of healthcare contexts from a multitude of health professionals, including music therapists. Unlike Lee (2016) additional hand searching of music therapy journals was not performed, given that these journals are included in the Music Index RILM database and searched systematically. It was considered unnecessary to locate every available study, because the results of a conceptual synthesis will not change if ten rather than five studies contain the same concept but will depend on the range of concepts found in the studies, their context and whether they are in agreement or not (Thomas & Harden, 2008). Accordingly, this review focused on 'conceptual saturation' with maximum variability and the resulting set of studies to be heterogenous rather than the homogeneity seen in meta-analyses.

Table 1. Inclusion and Exclusion criteria

Inclusion Criteria

1. Evaluate pain experience in wakeful patients in the presence of music listening
2. Are applied in a healthcare or laboratory setting
3. Use receptive music interventions where the person actively or passively observes music, either live or recorded for at least one part of the research study. This can be in conjunction with other activities, i.e. visualisation but this is the minimum requirement
4. Measure pain by either self-report measures or reduced pharmacological analgesic requirements.
5. Are empirical studies with primary data collection, including randomised, quasi-randomised, one armed trials, and qualitative accounts of MLIs

Exclusion criteria

6. Only focus on the process of making music
7. Focus on pain or injuries caused by performing music
8. Primarily examine issues of hearing loss, hearing disorders or other issues of aural health
9. Focus on the use of music as part of a patient information material

Study Selection

Document retrieval. In total 474 documents were retrieved across the four databases (see figure 1 for PRISMA flow chart). Two independent reviewers agreed to exclude 175 articles based on the abstract and title content and include 89 articles for further review. Where there was no agreement between the two reviewers, articles were included for full text review.

Full text review and Quality Assessment. Each article was reviewed in full by at least one author to deduce suitability for inclusion, according to the inclusion / exclusion criteria in Table 2. At this stage, it was considered necessary to conduct a quality appraisal on articles to deduce their suitability for inclusion in a systematic way. A minimal quality appraisal was conducted to ensure that all studies included met the appraisal criteria of the five 'fatal flaws' (see table 2; Dixon-Woods et. al. 2006). Out of the 104 articles that were reviewed in full, 75 articles were included in the qualitative synthesis (see figure 1).

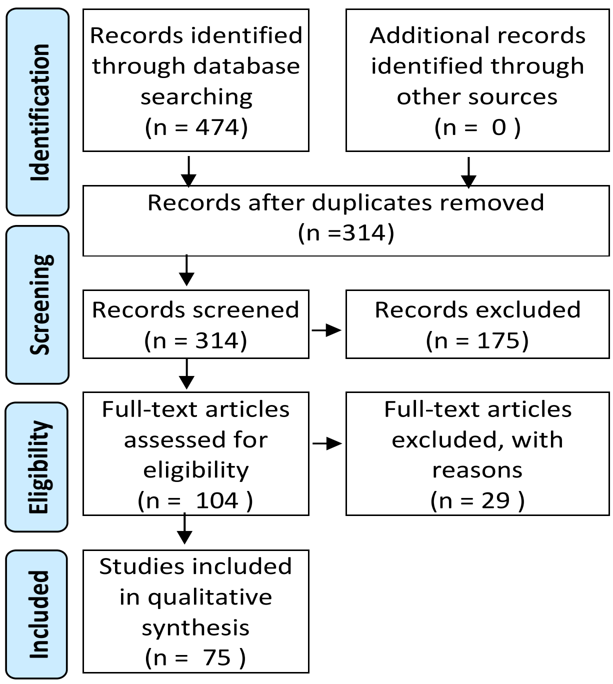


Figure 1.Prisma Flow Diagram of Article Selection

Data Extraction and Analysis

In order to extract the data systematically and consistently a data extraction form was developed in Microsoft excel by the research team in line with previous recommendations (Colquhoun, Levac, O'Brien, Straus, Tricco, Perrier et. al., 2014). The semi-structured form consisted of a series of data points deemed necessary to answer the research question posed in the systematic review protocol, as well as sections of text that could be developed into themes using thematic synthesis. The extraction form was piloted on a sample of 10 papers and amended before being applied to every article. Partial double data extraction was completed on a 10% random sample, and single data extraction was completed on the remaining papers.

In order to describe the theoretical underpinnings of MLIs in pain contexts, thematic synthesis was conducted on the provided rationales of MLI studies in the introduction section and by examining how researchers interpreted their results in the discussion section. A thematic synthesis strategy was developed in line with guidance outlined by Thomas and Harden (2008), involving three stages; (1) the reviewer coded the text, line by line according to its meaning or content, using EPPI reviewer 4 software. There was no limit imposed on the number of codes that could arise in the text. (2) Descriptive themes were developed by grouping and combining codes based on similarity to develop analytic themes. Sometimes this meant incorporating two similar codes with a slightly different meaning to produce a descriptive theme, and the overall code would be relabelled to reflect this. The original codes would then be added to the description of the descriptive theme so that it was clear which codes were reflected in each descriptive theme. Codes that only applied to a very small number of studies (i.e. three or less) were re-coded into larger analytical themes where possible using the process above. (3) Descriptive Themes were then grouped together to form analytical themes. This was done by two reviewers who created a narrative description of each theme based on collections of descriptive themes.

Results

Thematic synthesis of the researchers provided rationales of the underlying cognitive principles of MLIs generated four analytical themes. The themes were Intramusical features directly command attention; cognitive agency of the listener; Multidimensional Integration; and cognitive strengthening, which are briefly outlined below in more detail, with some sample quotations. It should be noted that these themes are not mutually exclusive, with several papers identifying aspects of both approaches in their delivery of MLIs.

Intramusical Features Directly Command Attention

This theme is derived from descriptions of the role of intramusical features of MLIs, such as tempo, or rhythm and their direct impact on pain responses and attentional resources. These descriptions focus on music as a stimulus that induces a response, almost in a mechanistic way, with an expectance of specific physiological changes in response to the music. Additionally many descriptions acknowledge neural activations in response to music, but often characterise neural processing as a one way process without reflecting on the top-down cognitive processes involved in musical experiences .

“Since the tempo of the music so influenced the subjective and physiological arousal of the participants, we conclude that the effect of tempo on pain is related to arousal”
Kennter-Mabiala (2007)

The cognitive mechanisms emphasised by this theme is direct attentional processes to mediate the effects of MLIs and is compatible with limited capacity models of attention, which many studies cited as the theoretical underpinning of their studies. However some studies also acknowledge the limited value of this model or question the usefulness of exclusively using this approach.

"Traditional models of attentional capacity can provide a plausible explanation for music- induced analgesia"
Finlay 2014

“Although evidence for the limited capacity model has been provided by a number of studies, unequivocal support has not been found.”

This theme also incorporates experimenter chosen music, selected for specific intramusical features, emphasised as having universal relaxing properties. Within this framework the same piece of music would be selected for every patient, on the assumption that a single piece of music would be perceived as universally relaxing or arousing to all patients.

Cognitive agency of listener

This theme recognises that the patient can become an active agent within the music listening intervention with the ability to make choices in terms of the music chosen and how their attention is directed. This theme acknowledges that participants active participation plays a role in MLIs, and that a sense of personal meaning of the music may play a role in maintaining this participation. Cognitive agency of the listener is partially derived from descriptions in the literature that emphasize that cognitive effort and active listening may be required for a successful intervention, or may be insufficient where the intervention has been unsuccessful.

“It has been suggested that music is more effective if patients are able to concentrate on the music.”
Allred 2010

However it is also noted by many researchers that while active participation may be required, that this participation may not feel effortful, compared to other cognitive tasks, because music may be easier to concentrate on. Additionally levels of active participation may vary considerably depending on the specific circumstances of the music listening intervention, the individual listening and the constituent components of the music being listened to, which will offer different opportunities for engagement.

“Future research could aim to investigate the extent to which participants are immersing themselves in potentially active components of music listening, such as analysis of the

instrumentation, recognition of compositional patterns, and communication through lyrics."
Finlay 2015

Cognitive agency is also exemplified by an emphasis on self-selected music, often chosen from an unlimited choice, and influenced by the participants individual background and previous experience with music.

The choice of music is important, and the response to music is influenced by earlier experience of music as well as by gender, age, culture, mood, and attitude."
Björkman 2013

This emphasises that people have specific reasons for selecting music depending on their mood, and making it more likely that the music has an increased sense of meaning or resonance with the person than generic music. By bringing a sense of personal meaning to the MLIs, people have the opportunity to reflect on whatever they wish during the musical experience. This helps to increase the listeners' agency in the intervention and enhance their locus of control within the healthcare environment.

"The findings of this study provide further evidence for the efficacy of music listening for pain relief, with use of preferred choice of music leading to significantly longer tolerance, less anxiety, and a greater perceived control over the experience than both a silence control and a visual distraction condition."
Mitchell 2008

Thus, by selecting their own music, the person moves from being a passive listener to becoming an active agent participating in the intervention, through enhanced cognitive engagement with the music. This is facilitated by a personal sense of meaning from the music or previous experience with the musical piece.

Multidimensional Integration

This theme addresses the multidimensional nature of pain and indeed the multidimensional experience offered by MLIs, both of which involve integrated emotional components with attentional and cognitive processes. This gives a more complex overview of the holistic interaction between music and pain and describes how cognitive and emotional processes are intrinsically combined and shaped by musical engagement in an integrated manner leading to analgesic benefits alongside emotional outcomes.

"Music, like pain, is a powerful multidimensional experience with sensory, emotional, cognitive, behavioural and social elements. Like pain, the power of music may be related to its operating simultaneously on multiple levels."
Gold 2013

Multi-dimensional integration is derived from descriptions that emphasise the role of emotional engagement alongside cognitive engagement to mediate analgesic benefits of MLIs. This is in line with the neuromatrix theory of pain, which has evolved from the more simplistic gate control theory of pain and is cited in in many studies as the underlying theoretical rationale for MLIs. The neuromatrix theory of pain emphasises how different dimensions of pain and the persons expereince interact to impact to shape the experience of pain. It also acknowledges that a multitude of additional factors such as arousal or stress, can impact how these factors are integrated.

"The current article argues that the mechanism is indeed multifaceted, involving optimal physiological arousal, positive affect and directed attention towards a self-chosen (i.e. preferred) musical stimulus in order to achieve attenuated pain perception"
Garcia 2016

The theme of multidimensional integration is reflected in several neuro-imaging studies which highlight that neural hubs for multimodal integration are involved during MLIs, as opposed to describing neural activity simply in terms of activations or chain reactions.

"Furthermore, it seems that the AnG is an important hub for multimodal integration, as it structurally connects with parietal, temporal, and frontal areas."
Garza-Villarreal 2015

Additionally multi-dimensional integration is also reflected in studies offering opportunities for multi-modal engagement within the music intervention through the introduction of visual, olfactory or virtual reality components alongside the music listening intervention.

"*The stimulation consisted of ocean shore sounds, combined with a virtual animation of an ocean shore in the audio and visual stimulation group."*
JJ, 2010

As well as acknowledging the contributory role of each dimension of the MLI, each dimension must be integrated in an optimal way in order to optimise engagement. Evidence of successful optimised engagement occurs when there is sufficient immersion or absorption within the MLI to compete with pain signals. Optimised engagement is characterised by immersion and an altered sense of time perception, whereby people estimate that time is passing faster than it really is in reality.

'Similarly, all music speeded the passing of time via changes in perceived pain tolerance, expressed through retrospective judgements of time passing: perceived pain tolerance was consistently lower than actual pain tolerance, suggesting participants felt time had passed more quickly than in reality"
Finlay 2016

This impact of MLIs on the persons impression of reality is elaborated in constructivism based accounts, which hold that music does not simply distract people from their pain, or provide an escape from pain, but instead prevent pain schemata from being constructed in the first place.

"We propose that highly engaging activities may prevent pain by creating competing constructions of reality that draw on the same processing resources. "
Bradshaw 2013

Cognitive Strengthening

This theme identifies different mechanisms through which patients may be enabled, through enhanced cognitive capacity, a strengthened sense of self, increased self-efficacy, enhanced motivation or generally energized. As patients increase in cognitive strength they are in a better position to cognitively dissociate from pain or embrace and accept the pain in chronic pain contexts. This theme is derived from several descriptions that emphasise feelings of energy and motivation as being central to the analgesic effects of MLIs. This was often reflected in qualitative studies, where patients were directly asked how they benefited from music.

Subjects were able to consciously acknowledge bolstered feelings of control and positive valence: ''The specific music I chose makes me feel like I'm going somewhere, and can't be stopped. It makes me feel less vulnerable...''; ''Music inherently makes me happy''; ''music made me feel...energized and in control of myself''.
Hseih 2014

Additionally some studies emphasised the benefits of music to the person on a deeper level, in terms of providing an opportunity to reflect on existential issues that may arise from changes that arise through painful illness. These issues may be important from the patients perspective to build a coherent narrative of their experience and help them to move forward in their illness journey.

"First, music helped to bridge pre-illness identity to present identity and facilitated reflection on existential issues"
Bradt 2015

As the patients self-efficacy increases they can are more able to choose whether to dissociate from the pain or accept the pain. This means that the patient is cognitively strengthened to use a top-down strategy to cognitively control their pain, which was demonstrated to some degree in neuroimaging studies. Cognitive dissociation from pain, requires constant attentional effort and patients may use music as an aid to structure their attention away from painful sensations. In contrast to this, patients may decide to accept their pain experience, in line with third wave therapies, which may be more appropriate in a chronic pain setting.

"The current study has shown that all music is advantaged over no music in its ability to extend the amount of time people are able to tolerate pain, suggesting that it has value in supporting patients in living with daily pain in the context of third wave behavioural therapies"
Finlay 2016

Discussion

The aim of this study was to examine the current literature on MLIs in pain contexts to identify the theoretical rationales underpinning MLI development. Using thematic synthesis this study revealed four themes that align with different theoretical paradigms that explain the cognitive processes during MLI in pain contexts. The four themes outlined are (1) Intramusical Features Directly Command Attention, (2) Cognitive Agency of the Listener, (3) Multidimensional Integration, (4) Cognitive Strengthening.

Considerable focus has been placed on intramusical features in MLIs, based on an assumption that a single piece of music will elicit a similar physiological response across individuals. However this fails to recognize that the physiological response to music is characteristically idiosyncratic from person to person, with fast tempo sometimes leading to an increase in heart rate, and other times to a reduction in heart rate (Thaut, 2016). This precludes the notion that a one to one relationship exists between musical features and physiological responses, (Brattico, 2015), and suggests that the cognitive agency of the person in directing and maintaining the musical experience must be considered. This places the listener as an active agent participating with the music and removes the idea that music is something that is done to the listener.

The third theme that emerged from the literature was multi-dimensional integration, which addresses the multidimensional nature of pain and indeed the multidimensional experience offered by MLIs. This expands the music listening intervention beyond the interaction between the listener and the music, and incorporates the wider environment and additional indirect influences on pain such as emotional interactions. In this approach emotional components are combined with attentional processes; meaning that emotionally relevant material becomes more engaging, and directly implicates any additional emotional or distracting material that may enhance or disrupt the listeners musical engagement. Multi-dimensional integration is directly in line with the neuromatrix theory of pain which acknowledges that the pain experience is derived from a combination of factors such as noiceptive sensory inputs, visual and other sensory inputs, and fluctuations in cognitive and emotional states including the stress-regulation system (Melzack, 1999; Melzack & Katz, 2013). Each of these factors influence the cognitive interpretation of the pain experience, meaning that music could be beneficial at a range of different levels including emotion regulation, stress reduction, or by making the experience seem less unpleasant overall. However, many studies included in this review cited the gate control theory of pain, such that music triggers a gating mechanism that modulates pain inputs (Melzack, 1996). This is a much more limited explanation of the music-pain interaction, however it explains why the majority of studies rely on intramusical features, to mediate the benefits of MLIs. Instead MLIs need to reflect the complexity of both pain experiences and music listening in order to adequately facilitate optimal musical engagement. Evidence of successful optimised engagement occurs when there is sufficient immersion or absorption within the MLI to compete with pain signals. Optimised engagement is characterised by immersion in the MLI and an altered sense of time perception, whereby people estimate that time is passing faster than it really is in reality.

The final theme that was emphasised in the literature was cognitive strengthening, which suggests that through the interaction with music patients feel energized or motivated leading to a strengthened sense of self-efficacy. This provides listeners with the cognitive resources to use a top-down strategy to either cognitively dissociate or willingly accept the nature of their pain experience, in line with third wave approaches. Cognitive dissociation from pain, requires constant attentional effort and patients may engage with music to re-structure their experience away from painful stimuli.

Conclusion

Together the four themes; Intramusical Features Directly Command Attention, Cognitive Agency of the Listener, Multidimensional Integration, and Cognitive Strengthening. provide a more comprehensive understanding of the cognitive principles underpinning MLIs in pain contexts compared to any one singular approach. Future MLI development should support the listener as an active agent within the MLI and reduce competing material such as noises or pain relevant stimuli in order to facilitate optimal engagement.

Acknowledgements.
This project was funded by a government of Ireland postgraduate scholarship from the Irish research council grant number R17777

References

Arksey, H., & Malley, L. O. (2005). Scoping Studies: Towards a Methodological Framework. The International Journal of Social Research Methodology, 8(1), 19–32 https://doi.org/10.1080/1364557032000119616

Bradt, J. (2012). Randomized controlled trials in music therapy: guidelines for design and implementation. Journal of Music Therapy, 49(2), 120-149

Bradt, J., Dileo, C., Grocke, D., Magill, L. (2011). Music interventions for improving psychological and physical outcomes in cancer patients (Review). The Cochrane Collection: Wiley

Brattico, E. (2015). From pleasure to liking and back: Bottom-up and top-down neural routes to the aesthetic enjoyment of music. In Art, Aesthetics and the Brain (pp. 303-318). Oxford University Press.

Chen, L., Wang, T., Shih, Y., Wu, L. (2013). Fifteen-minute music intervention reduces pre-radiotherapy anxiety in oncology patients. European Journal of Oncology Nursing, 17, 436-441

Clark, M., Isaacks-Downton, G., Wells, N., Redlin-Frazier, S., Eck, C., Hepworth, J.T., Chakravarthy, B. (2006). Use of preferred music to reduce emotional distress and symptom activity during radiation therapy. Journal of music therapy, 43(3), 247-265.

Colquhoun, H. L., Levac, D., O'Brien, K. K., Straus, S., Tricco, A. C., Perrier, L., ... & Moher, D. (2014). Scoping reviews: time for clarity in definition, methods, and reporting. *Journal of clinical epidemiology*, *67*(12), 1291-1294.

Fancourt, D., Ockelford, A., & Belai, A. (2014). The psychoneuroimmunological effects of music: A systematic review and a new model. Brain, behavior, and immunity, 36, 15-26.

Howlin, C., Guerin, S., & Rooney, B. (2018). Investigating the cognitive mechanisms of music listening interventions in pain management settings: A Scoping Review Protocol. European Journal for Person Centered Healthcare, 6(1), 25-29.

Huang, S. T., Good, M., & Zauszniewski, J. A. (2010). The effectiveness of music in relieving pain in cancer patients: a randomized controlled trial. International journal of nursing studies, 47(11), 1354-1362.

Keenan, A., & Keithley, J. K. (2015, November). Integrative Review: Effects of Music on Cancer Pain in Adults. In Oncology nursing forum (Vol. 42).

Koelsch, S (2012). Brain and Music. Wiley-Blackwell: Chichester.

Krishnaswamy, P., & Nair, S. (2016). Effect of music therapy on pain and anxiety levels of cancer patients: A pilot study. Indian journal of palliative care, 22(3), 307.

Kwekkeboom, K. L. (2003). Music versus distraction for procedural pain and anxiety in patients with cancer. Oncology Nursing Forum, 30(3), 433-440.

Lee, J. H. (2016). The effects of music on pain. Journal of Music Therapy Advance Access, 1-48. doi:10.1093/jmt/thw012

Levac, D., Colquhoun, H., O'Brien, K. K. (2010). Scoping studies: advancing the methodology. Implementation Science, 5(69), 1-9. doi: http://www.implementationscience.com/content/5/1/69

Mainka, S., Spingte, R. & Thaut, M. (2016). Music Therapy in Medical and neurological rehabilitation settings. In: Oxford Handbook of Music Psychology, S. Hallam, I. Cross, & M. Thaut. (Eds.), 2nd edn. pp. 857- 873. Oxford: Oxford University Press.

Melzack, R. (1996). Gate control theory. Pain Forum, 5(2), 128–138. doi:10.1016/ S1082-3174(96)80050-X.

Melzack, R. (1999). From the gate to the neuromatrix, Pain supplement, 7, S121-S126.

Melzack, R. & Katz, J. (2013). Pain. Wiley Interdisciplinary Reviews: Cognitive Science, 4, 1-15. doi: 10.1002/wcs.1201

Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

Thaut, MH. (2016). History and research. In S. Hallam, I. Cross, M. Thaut (Eds.) The Oxford Handbook of Music Psychology (pp. 893 – 904) 2nd Edition. Oxford University Press: Oxford.

Thomas, J. & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8(45), 1-10. doi:10.1186/1471-2288-8-45

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Analysis of Chunk Forming Factors in a Piano Performance Learning System Using Grounded Theory Approach

Manami Ishigaki,[1] Yoshinari Takegawa,[2] Keiji Hirata,[3] and Atsuko Tominaga[4]

[1,2,3,4] *Future University Hakodate, Japan*

[1]g2117005@fun.ac.jp, [2]yoshi@fun.ac.jp, [3]hirata@fun.ac.jp, [4]tominaga@fun.ac.jp

Abstract

In music, chunks work as one type of short-term memory and can form a hierarchical structure. A listener can more easily understand music with well-formed hierarchical chunks as a piece of long-term memory. We are interested in chunk formation in piano performance, which relates to four cognitive factors: visual and auditory information, body motion, and musical knowledge. We suppose chunk formation is influenced by cognitive factors. However, the relationships between chunk formation and the cognitive factors are not clarified, because of the difficulty of observing chunk formation in the mind from the outside, and that of its verbalization. Therefore, the goal of the research is to explore the formation mechanism and types of chunks occurring in piano practice and to identify the features for distinguishing chunk type from the others. We conducted an experiment to investigate various conditions; for example, in one experimental group only visual and auditory information was available, that is, subjects did not perform keying. The obtained experimental data (chunk information) was analyzed by the method using Grounded Theory Approach. As a result, we have found that three types of chunks can be significantly discriminated by the conditions of degree of proficiency, temporal pattern of chunk formation, and restriction on keying.

Introduction

In the playing and listening to of music, chunks are formed as one unit of short-term memory (Snyder, 2001). Firstly, at the level that people most directly recognize music, there is grouping of melody and meter, corresponding to short-term memory. As chunks included in short-term memory are appropriately made into hierarchical groups, areas of longer time spans are made into groups and long-term memories are formed. Accordingly, the more appropriate the hierarchical grouping structure of a piece of music is, the easier formation of long-term memory images, structure recognition, and memory storage and recollection become.

When playing, the majority of performers, such as pianists, are aware of musical groups such as melodies and meters and the relationship between those groups. At first, beginners are unable to read a score and play picking up notes one by one. As they repeatedly practice in this way, they become able to recognize phrases and chords as single groups, in other words as chunks, as they play, and the coordination of the right and left hands also becomes one chunk. As chunks are formed hierarchically, the time-span of chunks also increases and ultimately the player becomes able to play the entire piece smoothly.

In this way, chunks are formed in musical instrument practice. Aspects considered to be factors of the hierarchy formation of these chunks include the hearing of played notes (auditory information), notations and musical symbols written on the score (visual information), and movement of the body. If we are to assume that different types of chunk are born from different factors, it is expected that we can clarify part of the cognition involved in the practice process of instrument playing, by clarifying what kind of variety of chunk there are and what properties they have.

The goal of the research is to investigate the varieties and properties of chunks that emerge in piano performance and decide the criteria for discriminating them. First of all, in our research, we conduct an evaluation experiment to investigate chunk formation factors. We had subjects record the chunks they recognized during practice directly onto the musical score, then obtained comments about the recorded chunks through interviews. Using a method conforming to Grounded Theory Approach (GTA), we classified the data recorded on the score and the speech data into three types of chunk. We then verified the validity of these three types of chunk, based on quantitative data obtained in the evaluation experiment (e.g. number of chunks recorded on the score, time-span of chunks, number of days taken to master the assigned piece of music, etc.).

Related Work

Before now several studies have been carried out in relation to the support of piano learning. There are systems which identify players' weak points in performance from accumulated performance data and conduct practice focused on these areas (Akinaga, Miura, Emura, & Yanagida, 2006; Kitamura & Miura, 2006; Morita, Miura, Akinaga, & Yanagida, 2012; Mukai, Emur, Miura, & Yanagida, 2007; Yukusel, Oleson, Harrison, Peck, Afergan, Chang, & Jacob, 2016). These systems evaluate the user mainly from keying information, such as keying errors and strength of keying, and provide feedback. Piano Tutor has functions that include automatic page turning by performance following recognition, presentation of example performance as video or audio, and analysis of performance data and presentation of points for improvement, in the form of text etc (Dannenberg, Sanchez, Joseph, Capell, Joseph, & Saul, 1990). P.I.A.N.O. has a function to provide visual support by presenting keying position and articulation, such as legato and staccato, as piano roll inscription (Rogers, Röhlig, Weing, Gugenheimer, Könings, Klepsch, Schaub, Rukzio, Seufert, & Weber, 2014). Andante is a system that makes the body learn rhythm visually, by showing animation that changes depending on keying (Xiao, Tome, & Ishii, 2014). Smoliar, Waterworth, & Kellock (1995) is a proposed system that presents information such as tempo and dynamics and the state of articulation such as staccato or legato. Additionally, there are systems that support the movement of fingers and arms using sensors (Hadjakos, Aitenbichler, & Mühlhäuser, 2008), (Hsiao, Li, Yan, & Do, 2015; Tsutsumi, Nishino, & Kagawa, 2017). The above works of research realize systems that support surface activity, e.g. keying information or finger movement, but do not extend to learning support that considers higher cognitive activity, such as increase in score reading ability or performance by memorization of the score.

Proficient piano performance involves not only surface action, such as playing following the score and playing with consistent strength, but also deep cognitive activity. For example, finding regularity from previously learned knowledge, e.g. harmony or counterpoint, and forming groups (Furneaux & Land, 1999; Sloboda, 1974), being aware of the notes one bar ahead of the note one is focusing on (Rayner & Pollatsek, 1997; Truitt, Clifton, Pollatsek, & Rayner, 1997), and rapidly deciding on appropriate fingering considering the preceding and following notes (Sloboda, Clarke, Parncutt, & Reakallio, 1998). It is said by certain researchers (Pike & Cartner, 2010; Schulze, Mueller, & Koelsch, 2010) that this kind of higher cognitive activity has a deep connection to, for example, score memorization ability or learning speed. Furthermore, chunks are also given as a clue to understanding higher cognitive activity (Snyder, 2001).

Takegawa, Hirata, Tayanagi, & Tsubakimoto (2015) analysed chunk formation factors in piano performance by obtaining verbal statements. From the verbal statements and properties of the chunks, they classified chunks as three types, which are formed according to movement, sense of sound, etc. The chunk classification method used in our research is detailed in the following section, Results of Experiments and GTA Classification. Weaver (1943), as the result of an investigation into chunking ability in advanced pianist's score-reading, clarified that advanced pianists utilize chunk formation ability, such as reading and memorizing chords on the score as groups. Sakai, Kitaguchi, & Hikosaka (2003) are investigating chunking ability in visual movement sequences. They demonstrated that as mastery progresses chunks are formed, speed and accuracy of recognition increase, and the player becomes able to perform fingering efficiently, without mistakes.

Plan for Experiment to Clarify Chunk Construction Factors

To analyse chunk construction factors in piano playing, we conducted an experiment using a piano learning support system (Takegawa, Hirata, Tayanagi, & Tsubakimoto, 2015). The experiment was categorized into two conditions: 'with keying activity', in which subjects perform a set piece of music, and 'without keying activity', in which subjects listen to the set piece of music.

Subjects and Set Piece

The subjects were 25 university students. Subjects assigned to 'without keying activity' were six beginners (A – F) and six experiencers (G-L), and subjects assigned to 'with keying' were six beginners (M-R) and seven experiencers (S-Y). Piano beginners were those who had heard piano music in lessons at elementary, middle and high school and who could hum a tune, while experiencers were those who had been learning the piano for over three years. In this experiment, our aim was to investigate the chunk formation process, but we also included experiencers in order to investigate consistency of chunk formation results once proficiency is achieved.

The set piece of music was 8 bars at the beginning of the 3rd movement of Piano Sonata No. 11 "Turkish March" by W. A. Mozart. The reason we chose the Turkish March as a set piece was to evaluate basic piano technique.

Experimental Conditions

In the 'with keying activity' group, we had the subjects practice the set piece for 20 minutes a day using the piano learning support system (Takegawa et al., 2015). After practice, we conducted a proficiency test, in which we presented only the score of the set piece and had subjects play it straight through from beginning to end without using system support.

In the 'without keying activity' group, we had subjects listen to the set piece from beginning to end repeatedly for five minutes, while looking at the score using the piano learning support system (Takegawa et al., 2015). The set piece was played on a loop by the experimenters.

In the 'without keying activity' group, we had subjects listen to the set piece from beginning to end repeatedly for five minutes, while looking at the score using the piano learning support system (Takegawa et al., 2015). The set piece was played on a loop by the experimenters.

In the case of both with and without keying activity, after the achievement test and listening, we had subjects record chunks according to their personal opinions and then interviewed them, based on the Think Aloud method, about the reasons for their chunk formation. The processes described thus far constituted a single trial run. The length of the experiment varied depending on the subject, because we had the 'with keying activity' group continue the experiment until there were no keying errors and the 'without keying activity' group continue until chunks from the most recent trial run ceased to differ from those of the previous run. What are herein referred to as keying errors are: (a) mistakenly keying = Mis-keying, (b) not keying = Non-keying, and (c) run unnecessarily = Over-keying. During the experiment, we recorded keying data produced by the MIDI keyboard in the system, recorded the conditions of subjects playing or listening, using a video camera, and measured keying errors based on keying data.

Results of Experiments and GTA Classification

Experimental Result

The total number of days of experiment was 79 days and the total time was 85 hours. The average duration of the experiment, in days, was 2.5 days/person for beginners and 2.7 days/person for experiencers in the 'without keying activity' group, and 6.2 days/person for beginners and 1.6 days/person for experiencers in the 'with keying activity' group. The total number of chunks was 2852.

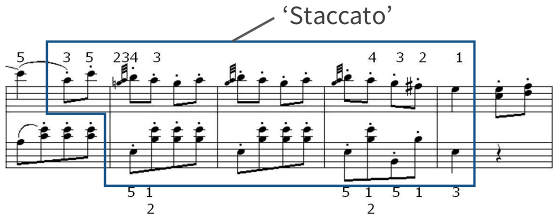


Figure 1. Chunks and comments formed by Subject T

Table 1. Example of GTA analysis on articulation for Subject T

No.	Data	Property/ Dimension	Label Name	Category
T1-5	Staccato	Group Criterion/ Staccato	Staccato	Articulation
		Size of Chunk / 8 beats		
		Number of Chunks / 1		
		Hand(s) / Both Hands		

Application of GTA to Examples

In accordance with GTA method, we analysed the comments relating to chunk formation that we obtained from the 25 subjects. GTA is a well-known method in qualitative research that involves producing theories based on data and is primarily used in the analysis of interview results (Strauss, 1987).

We will now give an example of chunks classified based on the data obtained in the experiment. Figure 1 is the chunks actually produced by subject T, and the corresponding comments, while Table 1 is the result of analysing subject T's comments by GTA. Subject T gave 'staccato' as the reason for the constructed chunk. Accordingly, we assigned staccato to the data heading in Table 1. Additionally, the group criterion in 'property' was staccato, the size of the chunk was 8 beats, as shown in Figure 1, the number of chunks was 1 and 'hands' covered both hands so was recorded as 'both hands'. The label name is a summary of the data and is thus staccato, and the category name, because slur was given as a formation reason in addition to staccato, is generalized as articulation.

Selection of Chunk Classification by a Method Conforming to GTA

In this research, the categories obtained by applying the method conforming to GTA are called small categories. Table 2 presents the large categories and the small categories comprised therein. Small categories were classified into 33 categories. To gain an overview of chunk formation factors, we considered whether it was possible to classify chunks even more largely and finally classified chunks into the following three categories:

- Score chunk: a chunk formed from information on the score, such as musical symbols including rests, phrases, slurs, staccato etc.

Table 2. List of large and small categories

Large category	Small category
Score chunk	timing, rest, note value, similar sound, rhythm, note by note, height of pitch, ornamentation, articulation, descending scales, form, note simplification, note density, bar, convergence, upbeat, accented beat, beam
Performance chunk	practice, consciousness of hands, difficulty level, fingering, position of hands, strength of sound
Knowledge application chunk	phrase, motif, chord, tune, base note, melody, harmony, predicative form, sound simplification

- Performance chunk: a chunk formed from anticipated aspects of performance, such as fingering and practice, or actual movement of the hands during performance.
- Knowledge application chunk: a chunk formed from the application of musical knowledge such as melody and base notes.

Figure 2 shows the result of two-way ANOVA for the factors of beginners/experiencers and three chunk types in terms of the number of chunks recorded a day in average for each chunk type. In the figure, places where there is significant difference or marginally significant one are indicated by * and + respectively and are allocated the codes (†1) - (†10). The meanings of (†1) - (†10) are summarized in Tables 3 and 4.

Discussion Relating to Chunk Classification

Regarding Discrimination of Chunks

Based on the results given in Table 3, Table 5 expresses the transition of each of the with and without keying activity groups from the first day of the experiment to the final day, for beginners and experiencers. *s* represents that there was significance for beginners and experiencers, *m* represents a marginally significant one, *n* represents that there was no significance and the whole table represents the transition from the first day to the last day. For example, where *m* => *s* is written in the knowledge application chunk column of 'with keying activity', this means that, in 'with keying activity', from (†3), on the first day experiencers had a tendency to be higher in significance than beginners and from (†9), on the last day experiencers were higher in significance than beginners.

Table 5 implies that the three types of chunk can be discriminated with the relevant transition patterns for with and without keying activity from the first day to the last day for beginners and experiencers. These patterns are the regulated conditions for discrimination. However, regarding performance chunks and knowledge application chunks, we only find marginally significant one.

Next, we consider the reason why each chunk can be discriminated, using Figure 2 and Table 3. Focusing on score chunks, from (†1), on the first day only, experiencers were higher in significance than beginners (*s* => *n*). From this we

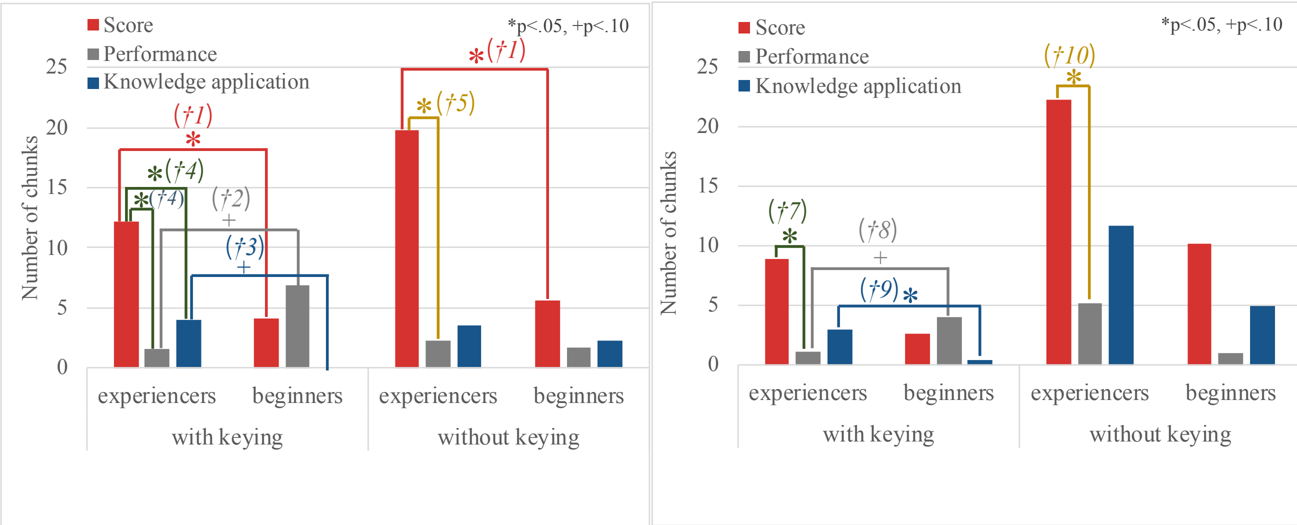


Figure 2. Average number per day of each chunk that subjects annotated and results of ANOVA

can deduce that experiencers are used to looking at a score and can recognize a larger amount of information. In contrast, beginners gradually become able to recognize most of the information on the score, through repeated practice or listening. Accordingly, a larger number of score chunks are formed compared to other chunks.

Looking at performance chunks, from (†2) and (†8), there was a significant difference only in 'with keying activity', and, regardless of whether it was the first or last day, beginners marginally higher significance than experiencers (*m* => *m*). In the case of beginners with keying activity, performance chunks are most widely formed, because the subjects can easily grasp the movement of the hands intuitively by actually playing. In contrast, experiencers, being used to the act of playing from their previous experience, hardly formed any performance chunks.

Looking at knowledge application chunks, from (†3) and (†9) in 'with keying activity' experiencers had marginally higher significance than beginners on the first day and had higher significance on the last day (*m* => *s*). Opposingly, there was no significant difference in 'without keying activity'. From this we can infer that experiencers can apply their knowledge better by performing keying activity, whereas beginners consume up entire cognitive resource only for playing the piano, thus there is no cognitive resource left for applying one's knowledge.

Correspondence to the Gagne Cognizance Model

The classification of chunk formation factors in this research can also be explained from Gagné's five classifications of learning outcomes (Gagné, Wager, Golas, & Keller, 2004). To describe simply Gagné's five classifications of learning outcomes, they are the result of systematically summarizing learning objectives and comprise five viewpoints: verbal information, intellectual ability, cognitive plan, movement ability, and attitude. In our research, it is suggested that classifications correspond in the following manner: score chunks correspond to verbal information because they are basic knowledge that anybody can understand; performance chunks correspond to movement ability because they involve

Table 3. Results of two factor ANOVA between experiencers and beginners and each chunk

Score chunk	First day	According to (†1), regardless of presence or lack of keying activity, experiencers were higher in significance than beginners (with keying activity ($F(1, 22) = 3.95, p < .10$), without keying activity ($F(1, 20) = 8.83, p < .05$)).
	Last day	Regardless of presence or lack of keying activity, there was no significant difference between experiencers and beginners.
Performance chunk	First day	According to (†2), with keying activity beginners had marginally higher significance than experiencers ($F(1, 22) = 3.61, p < .10$). But without keying activity there was no significant difference.
	Last day	According to (†8), with keying activity beginners had marginally higher significance than experiencers ($F(1, 22) = 3.42, p < .10$). But without keying activity there was no significant difference.
Knowledge application chunk	First day	According to (†3), with keying activity experiencers had marginally higher significance than beginners ($F(1, 22) = 4.59, p < .10$). But without keying activity there was no significant difference.
	Last day	According to (†9), with keying activity experiencers were higher in significance than beginners ($F(1, 22) = 5.62, p < .05$). But without keying activity there was no significant difference.

Table 4. Results by one factor ANOVA in each chunk

Score chunk	First day	According to (†4), with keying activity score chunk was higher in significance than performance chunk and knowledge application chunk regarding experiencers (MSe = 46.0635, p<.05).
	Last day	According to (†7), with keying activity score chunk was higher in significance than performance chunk regarding experiencers (MSe = 37.5079, p<.05). But there was no significant difference between score chunk and knowledge application chunk.
Performance chunk	First day	According to (†5), without keying activity score chunk was higher in significance than performance chunk regarding experiencers (MSe = 31.2389, p<.05).
	Last day	According to (†10), score chunk was higher in significance than performance chunk regarding experiencers (MSe = 141.1444, p<.05).

movement methods such as controlling fingering; knowledge application chunks correspond to cognitive ability because high-level musical knowledge can be applied to them. From Figure 2, Table 3, and Table 4, if we focus on score

	Score chunk	**Performance chunk**	**Knowledge application chunk**
With keying	*s => n*	*m => m*	*m => s*
Without keying	*s => n*	*n => n*	*n => n*

Table 5. Comparison of the transition of dates in experiencers and beginners

chunks, from (†4) and (†7), on the first day score chunks were higher in significance than performance chunks and knowledge application chunks. On the last day, score chunks were higher in significance than performance chunks. The reason that there was a clear difference on the first day in particular, is that on the first day easy to understand information, such as rests and slurs, tended to be easily noticeable. Accordingly, it is considered that this corresponds to verbal information, which can be expressed with words.

Paying attention to knowledge application chunks, from (†3) and (†9) it was shown that on the first day experiencers had marginally higher significance than beginners. On the last day experiencers had become higher in significance than beginners. Over the time from the first day to the last day, beginners somewhat got into the habit of forming chunks. This demonstrates that experienced players, having previous experience, are more able to apply knowledge. Thus, we consider that this corresponds to intellectual ability. If we compare (†4) and (†7), on the first day knowledge application chunks had higher significance than score chunks. In contrast, on the last knowledge application chunks and score chunks were not significant. This shows that experienced players form

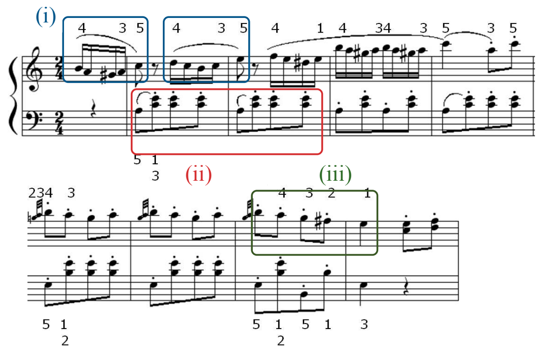


Figure 3. Example of primitive groupings

chunks while thinking, each time they practice. Accordingly, we consider that this corresponds to cognitive planning and is influenced by intellectual ability.

Focusing on performance chunks, from (†2) and (†8) in 'with keying activity' experiencers had marginally higher significance than beginners. Opposingly, in 'without keying activity' there was no significance. The reason that only 'with keying activity' was significant is thought to be that beginners become totally focused on playing whereas for experienced players playing is natural and taxes them very little. From the fact that 'with keying activity' had a significant difference while 'without keying activity' had no difference, a trend can be observed by which, through performing keying activity, subjects most widely form chunks related to movement. Therefore, it is considered that this corresponds to movement ability.

Correspondence to Snyder's 3 Types of Primitive Grouping Rule

If applied to Snyder, the classifications of chunk formation factors in this research are suggested to be connected to primitive grouping and learning grouping (Snyder, 2001). Primitive grouping is largely decided by intuitive sense. Furthermore, it is classified into three rules: adjacency, similarity and continuity. Figure 3 gives an example of primitive grouping. For example, regarding (i), notes that are in places that span rests, which are boundaries, are recognized as far while notes that are surrounded by boundaries are recognized as near, and the latter form a grouping (adjacency). Regarding (ii), a group of similar notes, patterns, fingering etc. is formed (similarity). Regarding (iii), places where notes continuously descend (or ascend) form a group (continuity). In accordance, it can be understood that these 3 rules correspond to score chunks and performance chunks. In grouping by learning, classification is largely decided based on previous experience. Thus, it can be understood that this corresponds to knowledge application chunks.

Conclusion and Future Work

The goal of this research was to identify the discrimination conditions for chunks in piano performance. To investigate the discrimination conditions, we conducted an evaluation experiment relating to chunk formation in piano performance, in accordance with GTA. From the results of the experiment, we were able to classify chunks into three types called score chunks, performance chunks and knowledge application chunks. By regulating three conditions, which were whether or not there was keying activity, whether it was the first or last day of the experiment, and whether subjects were piano beginners or experiencers, it was suggested that we were able significantly to discriminate the three types of chunk to a certain extent. Furthermore, it was understood that the three types of chunk could be explained with reference to Gagné's five classifications of learning outcome and Snyder's primitive grouping and grouping by learning.

As future work, we will compare elimination of chunks, new chunks, length of chunk etc., with previous chunks and analyse whether these aspects influenced the three types of chunk. Additionally, we will investigate whether we it is possible to discriminate between chunks that are established and chunks that don't become established.

Acknowledgements. In the progression of this research, I was greatly aided by the advice of associate professor Terai Asuka. Tamura Hayato and Osumi Yuka, graduates of Hirata / Takegawa laboratory, also provided valuable information concerning related research and the evaluation experiment. Additionally, this research was partially funded by JSPS scientific research funds JP16K12560 and JP15K00279.

References

Akinaga, S., Miura, M., Emura, N., & Yanagida, M. (2006). Toward Realizing Automatic Evaluation of Playing Scales on the Piano. *Proceedings of International Conference on Music Perception and Cognition* (pp. 1843-1847). Australia: International Conference on Music Perception and Perception.

Dannenberg, R. B., Sanchez, M., Joseph, A., Capell, P., Joseph, R., & Saul, R. (1990). A Computer-Based Multi Media Tutor for Beginning Piano Students. *Journal of New Music Research, 19 (2-3),* 155-173.

Furneaux, S. & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society B (*pp. 201-232). The United Kingdom: The Royal Society.

Gagné, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2004). *Principles of Instructional Design*. California, CA: Wadsworth Publishing Company.

Hadjakos, A., Aitenbichler, E., & Mühlhäuser, M. (2008). Potential Use of Inertial Measurement Sensors for Piano Teaching Systems. *Proceedings of the 4th i-Maestro Workshop on Technology-Enhanced Music Education*, 61-68.

Hsiao, C., Li, R., Yan, X., & Do, E. Y. (2015). Tactile Teacher: Sensing Finger Tapping in Piano Playing. *Paper Demonstration in TEI 2015*, 257-260. USA: Association for Computing Machinery.

Kitamura, T. & Miura, M. (2006). Constructing a Support System for Self-learning Playing the Piano at the Beginning Stage. *Proceedings of International Conference on Music Perception & Cognition,* 258-262.

Morita, S., Miura, M., Akinaga, S., & Yanagida, M. (2008). Feature exaggeration in scale performance on the piano. The Journal of the *Acoustical Society of America, 123(5)*, 8357-8362.

Mukai, M., Emura, N., Miura, M., & Yanagida, M. (2007). Generation of Suitable Phrases for Basic Training to Overcome Weak Points in Playing the Piano. *Proceedings of International Congress on Acoustics, MUS-07-018*. Spain: International Congress on Acoustics.

Pike, P. D. & Cartner, R. (2010). Employing cognitive chunking techniques to enhance sight-reading performance of undergraduate group-piano students. *International Journal of Music Education, 28 (3),* 231-246.

Rayner, K. & Pollatsek, A. (1997). Eye movements, the eye-hand span, and the perceptual span during sight-reading of music. *Journal of Current Directions in Psychological Science, 6 (2),* 49-53.

Rogers, K., Röhlig, A., Weing, M., Gugenheimer, J., Könings, B., Klepsch, M., Schaub, F., Rukzio, E., Seufert, T., & Weber, M. (2014). P.I.A.N.O.: Faster Piano Learning with Interactive Projection. *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (pp. 149-158). USA: Association for Computing Machinery.

Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking During Human Visuomotor Sequence Learning. *Journal of Experimental Brain Research, 152,* 229-242.

Schulze, K., Mueller, K., & Koelsch, S. (2010). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience, 33 (1)*, 96--186.

Sloboda, J. A. (1974). The Eye-Hand Span-An Approach to the Study of Sight Reading. *Journal of Psychology of Music, 2 (2)*, 4-10.

Sloboda, J. A., Clarke, EF., Parncutt, R., & Reakallio, M. (1998). Determinants of finger choice in piano sight-reading. *Journal of Experimental Psychology: Human Perception and Performance, 24 (1),* 185-203.

Smoliar, S., Waterworth, J., & Kellock, P. (1995). pianoFORTE: A System for Piano Education Beyond Notation Literacy. *Proceedings of the Third ACM International Conference on Multimedia (pp.* 457-465). USA: Association for Computing Machinery.

Snyder, B. (2001). *Music & Memory: An Introduction* (pp. 53-56). Cambridge, MA: The MIT Press.

Strauss, A. L. (1987). *Qualitative Analysis for Social Scientists*. Cambridge, MA: Cambridge University Press.

Takegawa, Y., Hirata, K., Tayanagi, E., & Tsubakimoto, M. (2015). Analysis of a Chunk-Based Learning Process in a Piano Learning Support System. *Proceedings of World Conference on Educational Media and Technology (EdMedia2015)* (pp. 1577-1585). USA: Association for the Advancement of Computing in Education.

Truitt, FE., Clifton, C., Pollatsek, A., & Rayner, K. (1997). The perceptual Span and the Eye-Hand Span in Sight Reading Music. *Journal of Visual Cognition, 4 (2)*, 143-161.

Tsutsumi, H., Nishino, H., & Kagawa, T. (2017). A Piano Performance Trainer with Tactile Guidance. *IEEE International Conference on Consumer Electronics - Taiwan*, 19-20. USA: Association for Computing Machinery.

Weaver, H. E. (1943). Studies of Ocular Behavior in Music Reading. *Journal of Psychological Mono-graphs, 55 (1),* pp. 1-29.

Xiao, X., Tome, B., & Ishii, H. (2014). Andante: Walking Figures on the Piano Keyboard to Visualize Musical Motion. *Proceedings in the International Conference on New Interfaces for Musical Expression* (pp. 629-632). USA: Association for Computing Machinery.

Yukusel, B. F., Oleson, K. B., Harrison, L., Peck,E.M., Afergan, D., Chang, R., & Jacob, R. J. (2016). Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty based on Brain State. *Proceedings of the 2016 CHI Conference on Human Factors in Computing System* (pp. 5372-5384).USA: Association for Computing Machinery.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Effects of Music on Episodic Memory

Rokaia Jedir[1], Flaithri Neff[2]

[1,2] *Interactive Systems Research Group, Limerick Institute of Technology, Ireland*

[1]rokaia.jedir@lit.ie, [2]flaithri.neff@lit.ie

Abstract

The universal use and often important status given to music in many human societies is due, in part, to the emotional responses it evokes and its ability to influence mood. Furthermore, the association between music and events where emotion plays a role has been shown to make certain musical content more memorable (Mikutta et al., 2015). It is suggested that this phenomenon occurs due to a significant interaction between emotion and memory mechanisms in limbic structures, alongside the significant influence of auditory stimuli on memory (Cahill & McGaugh, 1990; Heuer & Reisberg, 1990). Time-specific experiences are held in episodic memory, and can be interpreted as events in a serial form. Autobiographical events can be reconstructed from episodic memory, and time-sensitive details can be regurgitated. In this paper, the authors discuss how specific musical dimensions (such as rhythm, meter, timbral content and structure) engage with episodic memory in order to gain a deeper understanding of key influential components. The authors propose that an understanding of the interactions between specific auditory components and episodic memory is necessary in the development of perceptual and computational models that attempt to simulate discrete facets of human auditory perception. The outcome of this review has already informed the design of user-studies currently underway, which are testing the effects of certain auditory stream constituents on memory - specifically the level of disruption they cause during certain cognitive tasks. In studying the effects of specific musical dimensions on memory processes, the authors assert that this research could be effectively adapted to multimodal interface design processes.

Introduction

Several neural substrates of musical perception and musical memory processes are still not fully understood. The interaction of auditory stimuli and memory is described in Alan Baddeley's latest memory model (2015). This model supports the view that episodic memory permits the association of emotions and other contextual knowledge with auditory details, while semantic memory stores a more structured record of adopted facts and meanings. Distinct neural networks serve the retrieval of information from semantic and episodic memory (Wiggs et al., 1998; Platel et al., 2003). Functional imaging studies have exhibited similarities in the effects of verbal and visuospatial material and musical material. Separate groups of neural connections enable the interpretation of discrete musical components (such as loudness, pitch, timbre, rhythm), and networks combine to facilitate our perception of music as a whole.

Emotional events hold a privileged status in human memory (Labar & Cabeza, 2006; Suddendorf & Corballis, 2007). In addition, the perception of musical contenthas been shown to employ multiple brain regions concurrently, some of which are also responsible for emotional control and reactions (Levitin, 2006). The authors propose that mechanisms in control of emotional reactions may influence the perceptual effects of certain sonic attributes, and are therefore useful considerations in sonification design for various applications.

By drawing from particular constructs embodied in existing models and by extrapolating data from their series of user-studies, the authors aim to formulate sonification guidelines, which in turn will focus on utilising auditory streams to convey information in scenarios that also require users to engage in memory tasks in the visual domain. The framework will be implemented using a modelling language so that developers may readily adapt its design principles to algorithmic constructs.

Episodic Memory

The concept of episodic memory encompasses the ability to store and re ct autobiographical events, and associated contextual information such as time, location, emotional responses and relevant human interactions. This recollection process elicits the retrieval of contextual information pertaining to a specific event or experience that has occurred. The term "episodic memory" was advanced by Tulving (1972) with the aim to emphasise the distinction between the acts of "knowing" and "remembering", which was a prevalent topic in the field of cognitive psychology during that time. Research suggested that the act of "knowing" referred to the storage of factual information, which was based on the assumption that semantic links within long-term memory (LTM) could have been formed at any time. In order to "remember" a certain event, conversely, an individual's experience consists of emotional processes, as episodic memory provides access to other relevant information from a specific event. Today, cognitive psychologists consider human memory to comprise two main divisions known as declarative memory and implicit memory.

An integral process that occurs within declarative memory is the link between episodic memory and semantic links in LTM. Comprehension that is achieved by the combined episodic and LTM processes form the basis of declarative (or explicit) memory (Tulving, 1984; Tulving and Markowitsch, 1993; Buckner et al., 1995). The process of episodic memory recollection consists of three main properties: a subjective sense of time; a sentimental or emotional connection; and autonoetic consciousness. Autonoetic consciousness, in this context, is said to enable self-awareness within human memory, and associations of this awareness to specific timestamps stored in episodic memory (Baddeley, Eysenck & Anderson, 2009). Further research suggests that many other perceptual elements are largely influential in episodic memory recollection processes, such as visual imagery, narrative structure, and feelings of familiarity during the retrieval of specific semantic information (Hassabis & Maguire, 2007).

A crucial feature of the concept of episodic memory is the conscious role of an individual. Research by Baddeley (2001) supports the view that episodic memory is partly responsible for the ability to theorise and formulate structured plans. It is assumed that each individual creates a personal "self-concept” based on an accumulation of episodic experiences, and that this speculation contributes to the ability to relate to many different social situations, and enables the understanding of relevant content to enable contextualisation in these situations. Conway (2001), also outlined a review of this topic, which demonstrated a distinction between the term episodic memory, and the long-term accumulation of knowledge. Conway concluded that episodic memory stores were limited to relatively recent recollective experiences, while the long-term information accumulation can be seen as general autobiographical memory.

The experimental study of human memory has progressed to focusing on the differentiation and interactions between multiple subsystems in memory. This differs largely from the earlier theory of the unitary faculty, which previously resounded amongst researchers in psychology (Nairne, 1990; Baddeley, 2001). Critical differences between semantic and episodic memory were concluded by Tulving and Donaldson (1972), whereby relations between semantic memory and accumulated knowledge were reinforced. They proposed that it was proposed that LTM holds a collection of generic information that has been acquired through semantic processes, across multiple different contexts, and that this information can later be applied to many different interactions and situations. Episodic memory, in contrast, was considered to refer to an individual’s capacity to recollect specific events. In this respect, researchers today consider the essence of episodic memory to be its specificity and its associations to contextual information (Allen & Fortin, 2013).

Neurophysiological experiments have demonstrated a clear separation between explicit (declarative) and implicit memory systems (Parkin, 1992; Baddeley, 2001; Vargha-Khadem, Gadian & Mishkin, 2001; Wilson & Baddeley, 1988). These studies examined densely-amnesic patients abilities to perform various learning tasks. Findings consistently showed that patients were capable of performing cognitive tasks and apply learning techniques, including priming; the acquisition of motor skills; classical conditioning; and habit learning. This suggests a clear separation between the brain systems responsible for explicit and implicit memory processes. Furthermore, Squire published a review in 2013 that demonstrated enhanced learning processes in amnesic patients. Patients had no prior training or exposure to the specific tasks presented. These findings therefore support the view that multiple learning mechanisms exist, and are dependent on separate neurological structures. Moreover, this suggests that these mechanisms are unrelated to the recollective or episodic system, as it is severely impaired in amnesic patients. In particular, priming, which is the ability to enhance understanding in more difficult conditions through prior presentation of the stimuli in question (Squire et al., 1993), appeared to be enhanced in amnesic patients. Squire (2013) also suggested that the multiple learning mechanisms in question have a tendency to operate concurrently and interact with each other, independently of episodic memory. In 2001, Kopelman and Kapur provided further evidence in relation to episodic memory and retrograde amnesia (the loss of memory for events preceding a memory deficit). This research outlined the range of factors influencing retrograde amnesia, both in terms of specific episode recollection and knowledge retrieval, which is now referred to as semantic autobiographical memory (Kopelman et al 1989).

Studies have shown that information recorded in episodic memory stores can trigger episodic learning processes in various situations (Baars & Gage, 2010; Terry, 2015). The retrieval of information from episodic memory is dependent on two distinct processes. Initially, an individual accesses memory stores in search for relevant information. Additional event details can then be retrieved in a second more elaborate process (Hennessey, Morris and Kensinger, 2014). This staggered retrieval process enables more a energy and time-efficient retrieval process, as the individual gathers smaller amounts of information, depending on the progression of the cognitive task at hand. This is possible as the conscious working memory system is constantly updated with information that is relevant to the task at hand. Furthermore, emotions also reserve an important role in the establishment of episodic long-term memories in the mammalian brain. In 2010, researchers in cognitive psychology (Dere, Pause & Pietrowsky, 2010) proposed that learning processes are largely affected by emotional connections through episodic memory in both humans and non-human mammals. Evidence for this effect was observed in findings from clinical, neurophysiological, and neuroimaging studies on both animals and humans. A systematic review of multiple studies outlined the neuroanatomical and neurochemical foundations of the interaction between the different brain systems that contribute to emotional responses. From this, researchers could specify which brain areas contribute to the integration and comprehension of multi-dimensional stimuli within episodic memory. Functional neuroimaging studies have been conducted to examine the episodic memory retrieval (Spaniol et al., 2009; Cabeza & Nyberg, 2000). Generally, these studies have focused on identifying the brain regions involved in the initial search process outlined by Hennessy et al., (2014), and have identified that a bilateral memory network exists. Activity in this network spans across the prefrontal cortex (PFC), the medial-temporal lobe (MTL), as well as parietal, occipital and cerebellar regions. Full reviews of this brain activity during episodic memory retrieval has been published by Spaniol et al. (2009), and Cabeza & Nyberg (2000).

In 2010, a clinical review by Dere et al. suggested that changes in memory performance were dependent on the positive or negative valence of the stimulus material presented to participants. Data gathered from patients with emotional affective disorders were examined in particular, and in many cases, their performance exhibited a reduced ability to recall specific episodic memories in detail. This suggests a relationship between memory deficits and neurodegeneration in the brain systems responsible for mediating emotional responses to stimuli (Dere et al., 2010).

Auditory Perception and Episodic Memory

A significant number of reports indicate that music listening has positive effects on emotion and motor behaviour (Karageorghis & Terry, 1997; Menon & Levitin, 2005). However, the reports are more mixed when describing the effects of music listening on cognitive processing. Irrelevant sound present in an environment is generally disruptive to tasks such as reading, and other cognitive processes relying on a system referred to as “working memory" (Baddeley, 2015; Jones & Macken, 1993; 1995; LeCompte, 1996). This negative effect has been demonstrated in studies whereby the presentation of auditory stimuli remarkably impairs participants' ability to perform tasks involving serial recall. Amongst researchers in both psychology and psychoacoustics, this is referred to as the “Irrelevant Sound Effect", and is widely supported by numerous studies that consistently demonstrate negative effects on a listener’s cognitive performance caused by irrelevant sound. However, it must be noted that the auditory content is particularly influential when that individual is focused on a specific complex task (LeCompte, 1996, Beaman & Jones 1997; Macken et al. 1999; Campbell et al., 2002; Parmentier & Beaman, 2015, Colle & Welsh 1976). The ISE occurs due to a significant interaction between short-term memory systems and auditory perceptual processing.

Working Memory (WM) is an established model for verbal short-term memory (Baddeley, 1992; 2015), which is in constant use as it provides the means for comprehension and contextualisation in various situations. WM processes are relevant to the effects of musical content on mental performance, and provide a critical pathway for information to reach episodic memory stores. Within the WM system, a mechanism referred to as the "phonological loop” provides a basis for the interaction between auditory perception and working memory. Alan Baddeley (2012) describes this mechanism as the “model for verbal short-term memory”, which indicates its importance in auditory perceptual processing, both during focused auditory tasks and when the sound present in the environment is irrelevant to the cognitive task at hand.

Episodic memory retrieval has been studied in relation to an individual's strength of handedness, an individual difference variable, and associations with WM (Sahu, 2013). Findings have demonstrated that handedness and WM performance are dissociated from one another, however, both comprise mechanisms that significantly affect episodic memory performance. The authors of this paper hypothesise that episodic memory, being responsible for the encoding of specific experiences and events, is somewhat reliant on WM capacity, due to the notion that sensory input is required to pass through WM before it can be stored in episodic memory (Baddeley, 2015). To this end, the view that auditory perception holds a critical connection with the episodic memory system is widely accepted amongst researchers (Tulving et al., 1994).

The effectiveness of the encoding and storage processes within episodic memory are greatly influenced by the emotional qualities perceived. Emotionally striking percepts are therefore more memorable to the perceiver (Labar & Cabeza, 2006; Suddendorf & Corballis, 2007). The retrieval process in episodic memory implies a mental reconstruction of an earlier experience, and can often be recalled in great detail. This detail is dependent on a number of factors and is greatly influenced by emotion.

This enhanced ability to encode emotionally striking content is due to the specific physiological and neural mechanisms responsible for the processes. Cognitive neuroscientists (Cahill et al., 1996; Kilpatrick & Cahill, 2003; Hamann, 2001) have examined this phenomenon in studies which allowed them to identify the brain regions responsible or emotional episodic memory processes. To this end, the amygdala, which is the integrative centre for emotional behaviour, has been shown to facilitate processes in brain regions responsible for memory consolidation, including the hippocampus and the prefrontal cortex (Labar & Cabeza, 2006). Interactions between emotion and memory processes occur at multiple stages of information processing, which indicates a strong relationship between brain operations that reactivate emotional associations and the recollection of information from episodic memory.

The authors propose that interactions between the auditory perceptual system and episodic memory processes may be linked through the brain operations responsible for integrating emotion and emotional responses to various stimuli. The human auditory system is in a constant state of preattentive “ background” processing, even when a listener is not focused on a specific auditory stimulus. Background listening has been shown to affect other perceptual and cognitive processes, depending on the particular stimulus presented, as well as the cognitive capacities of the subject (Jones & Macken 1995; Macken et al., 1999). Particularly in the context of episodic memory, Proverbio et al. (2015) demonstrated this phenomenon in a study whereby the effects of music on episodic memory were examined, as well as autonomic responses. Findings exhibited a significant interaction between musical content and episodic memory capacity. Furthermore, these findings by Proverbio et al. indicated a significant distinction between the effects of emotionally striking musical content and non-musical auditory content, such as the sound of naturally occurring rainstorms, or white noise.

References

Allen, T. A., & Fortin, N. J. (2013). The evolution of episodic memory. *Proceedings of the National Academy of Sciences* , *110* (Supplement 2), 10379-10386.

Baars, B. J., & Gage, N. M. (2010). *Cognition, brain, and consciousness: Introduction to cognitive neuroscience* . Academic Press.

Bacon, F. (2000). *Francis Bacon: the new organon* . Cambridge University Press.

Baddeley, A. D. (1992). “Working Memory: The Interface Between Memory and Cognition.” in the *Journal of Cognitive Neuroscience*, *Vol. 4(3).* pp. 281-288.

Baddeley, A. D., Eysenck, M.W. & Anderson, M.C. (2009). Memory. East Sussex.

Baddeley, A. D. (2015). “Working Memory” in *A. Baddeley, Michael W. Eysenck & Mickael C. Anderson (Eds.), Memory* , Ch. 3. Psychology Press, NY.

Beaman, C.P. & Jones, D.M., 1997. Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), p.459.

Cabeza, R., Daselaar, S. M., Dolcos, F., Prince, S. E., Budde, M., & Nyberg, L. (2004). Task-independent and task-specific age effects on brain activity during working memory, visual attention and episodic retrieval. *Cerebral cortex* , *14* (4), 364-375.

Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of cognitive neuroscience* , *12* (1), 1-47.

Cahill, L., Haier, R.J., Fallon, J., Alkire, M.T., Tang, C., Keator, D., Wu, J. and Mcgaugh, J.L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information.

Proceedings of the National Academy of Sciences , *93* (15), 8016-8021.

Cahill, L. & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends Neuroscience, 21* , 294–299.

Chanda, M. L., & Levitin, D. J. (2013). The neurochemistry of music. *Trends in cognitive sciences* , *17* (4), 179-193.

Clark, C. N., Downey, L. E., & Warren, J. D. (2014). Music biology: all this useful beauty. *Current Biology* , *24* (6), R234-R237.

Clayton, N. S., Salwiczek, L. H., & Dickinson, A. (2007). Episodic memory. *Current Biology* , *17* (6), R189-R191.

Colle, H. A., & Welsh, A., (1976). "Acoustic Masking in Primary Memory" *Journal of Verbal Learning and Verbal Behavior., Vol. 15*. pp. 17-31.

Dhanjal, N. S., Warren, J. E., Patel, M. C., & Wise, R. J. (2013). Auditory cortical function during verbal episodic memory encoding in Alzheimer's disease. *Annals of neurology* , *73* (2), 294-302.

Ford, J. H., Morris, J. A., & Kensinger, E. A. (2014). Effects of emotion and emotional valence on the neural correlates of episodic memory search and elaboration. *Journal of cognitive neuroscience* , *26* (4), 825-839.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in cognitive sciences* , *5* (9), 394-400.

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences* , *11* (7), 299-306.

Jones, D. M. & Macken, W. J. (1993) "Irrelevant Tones Produce an Irrelevant Speech Effect: Implications for Phonological Coding in Working Memory" in the *Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol. 19*. No. 2. pp. 369-381.

Jones, D. M. & Macken, W. J. (1995). "Phonological Similarity in the Irrelevant Speech Effect or Between-Stream Similarity" in the *Journal of Experimental Psychology: Learning, Memory and Cognition, Vol. 21*. No. 1.

Jones, D. M., & Macken, W. J. (1995). "Organizational Factors in the Effect of Irrelevant Speech: The Role of Spatial Location and Timing." in *Memory and Cognition, Vol. 23*. pp. 192–200.

Karageorghis, C. I., & Terry, P. C. (1997). The psychophysical effects of music in sport and exercise: A review. *Journal of Sport Behavior* , *20* (1), 54.

Kilpatrick, L., & Cahill, L. (2003). Amygdala modulation of parahippocampal and frontal regions during emotionally influenced memory storage. *Neuroimage* , *20* (4), 2091-2099.

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience* , *7* (1), 54.

LeCompte, D. C. (1996). "Irrelevant Speech, Serial Rehearsal, and Temporal Distinctiveness: A New Approach to the Irrelevant Speech Effect" *Journal of Experimental Psychology: Learning, Memory and Cognition,* Vol. 22, pp. 1154–1165.

Macken, W. J., Tremblay, S., Alford, D. & Jones, D. (1999). "Attentional Selectivity in Short Term Memory: Similarity of Process, Not Similarity of Content, Determines Disruption" *International Journal of Psychology*, 34, pp. 322–327.

Menon, V., & Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* , *28* (1), 175-184.

Mikutta, C.; Strik, W. K.; Knight, R. & Altorfer, A. (2015). Memory in Music and Emotions. *Synaptic Fundamentals of Memory Performance.* Nova Science, Bern.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition* , *18* (3), 251-269.

Parkin, A. J. (1992). Functional significance of etiological factors in human amnesia.

Parmentier, F. B. R & Beaman, C. P. (2015). "Contrasting Effects of Changing Rhythm and Content on Auditory Distraction in Immediate Memory" in the *Canadian Journal of Experimental Psychology,* Vol. 69, No. 1, pp. 28–38.

Platel, H., Baron, J.C., Desgranges, B., Bernard, F. & Eustache, F., (2003). Semantic and episodic memory of music are subserved by distinct neural networks. *Neuroimage, 20(1),* 244-256.

Proverbio, A. M., Nasi, V. L., Arcari, L. A., De Benedetto, F., Guardamagna, M., Gazzola, M., & Zani, A. (2015). The effect of background music on episodic memory and autonomic responses: listening to emotionally touching music enhances facial memory capacity. *Scientific reports* , *5* , 15219.

Rajah, M. N., & D'esposito, M. (2005). Region-specific changes in prefrontal function with age: a review of PET and fMRI studies on working and episodic memory. *Brain* , *128* (9), 1964-1983.

Sahu, A., Christman, S. D., & Propper, R. E. (2016). The contributions of handedness and working memory to episodic memory. *Memory & cognition* , *44* (8), 1149-1156.

Sahu, A. A. (2013). *The roles of individual differences and working memory in episodic memory* . The University of Toledo.

Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: how musical sounds become rewarding. *Trends in Cognitive Sciences* , *19* (2), 86-91.

Schacter, D. L., Gilbert, D. T., & Wegner, D. M. (2011). Semantic and episodic memory. *Psychology,* , 240-241.

Spaniol, J., Davidson, P. S., Kim, A. S., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: meta-analyses using activation likelihood estimation. *Neuropsychologia* , *47* (8), 1765-1779.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans?. *Behavioral and brain sciences* , *30* (3), 299-313.

Terry, W. S. (2015). *Learning and memory: Basic principles, processes, and procedures* . Psychology Press.

Tulving, E. (1972). "Episodic and Semantic Memory" in *E. Tulving & W. Donalson (Eds.), Organisation of Memory,* New York: Academic Press.

Tulving, E. (1984). Precis of elements of episodic memory. *Behavioral and Brain Sciences* , *7* (2), 223-238.

Tulving, E., Kapur, S., Markowitsch, H. J., Craik, F. I., Habib, R., & Houle, S. (1994). Neuroanatomical correlates of retrieval in episodic memory: auditory sentence recognition. *Proceedings of the National Academy of Sciences* , *91* (6), 2012-2015.

Vargha-Khadem, F., Gadian, D. G., & Mishkin, M. (2001). Dissociations in cognitive memory: the syndrome of developmental amnesia. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* , *356* (1413), 1435-1440.

Wiggs, C. L.; Weisberg, J. & Martin, A. (1998). "Neural correlates of semantic and episodic memory retrieval" in *Neuropsychologia, 26*

Wilson, B., & Baddeley, A. (1988). Semantic, episodic, and autobiographical memory in a postmeningitic amnesic patient. *Brain and Cognition* , *8* (1), 31-46.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Human Perception of Melodic Similarity in Theme and Variation Pieces

Alexandrea Jonker,[1] and Leigh VanHandel[2]

[1] *College of Music, Michigan State University, USA*
[2] *College of Music, Michigan State University, USA*

[1]jonkeral@msu.edu, [2]lvh@msu.edu

Abstract

The generation of musical forms relies heavily on a listener's ability to hear and understand repetition of and similarity between musical ideas. Due to its importance in comprehending music, this research project intends to investigate how listeners perceive melodic similarity in varied repetitions of musical ideas. Providing insight to the similarity processes in the brain could be applied to our understanding of theme and variation compositions as well as the human perception of similarity outside the musical realm. The goal of this research is to examine theme and variation pieces to determine which musical factors affect a listener's ability to perceive melodic similarity. The first study examined three factors: harmony, timbre, and ornamentation. Participants were presented with pairs of themes and permutations of variations based on Beethoven's WoO 64 and WoO 77 and asked to rank their similarity. The second study attempted to unpack this issue further by asking participants to rate how similar variations from different pieces are to a given theme. Themes and variations from Mozart K. 354 and Beethoven Op. 35 were compared. From the first study, results showed that a change in harmony or in timbre were not considered to be variations on the theme. However, varying degrees of ornamentation did affect the participants' ability to recognize the variation as being similar to the theme. In the second study, no difference was observed by participants when listening to a variation that belonged to the theme and ones that did not. Based on these results, it is likely that listeners are focusing on surface elements rather than structural elements.

Introduction

The generation of musical forms relies heavily on a listener's ability to hear and understand repetition of and similarity between musical ideas. Strophic form, ritornello, concerto, rounded binary, ternary, rondo, sonata form, theme and variations, and small tight-knit themes such as sentence and period structures were all founded on the principle of repetition. Even more modern verse-chorus pop and folk songs and 12-bar blues capture the importance of reiteration. The topic of repetition was the basis of Lisa Margulis' book *On Repeat: How Music Plays the Mind* (2014), in which she explained, "repetition makes it possible for us to experience a sense of expanded present characterized not by the explicit knowledge that *x* will occur at time point *y*, but rather by a heightened sense of orientation and involvement.". Since music is a temporal art in which listeners cannot go back to rehear a section, at least not in a live performance setting, repetition has been used by composers for centuries to give listeners a sense of order and understanding. Denis Arnold also supports this notion in his definition of form on *Oxford Music Online*: "the effect of such repetition will be registered, at least by the experienced listener, and give him or her a sense of order. However, repetition alone would be monotonous. Musical form also consists of the relationship between different patterns of sound" (Arnold, n.d.). Due to its foundational importance in comprehending music, this research project intends to investigate how listeners perceive melodic similarity in varied repetitions of musical ideas.

Review of Literature

With the rapid growth of interest in MIR in recent years, much of the research done on melodic similarity focuses on the creation of computational algorithms. Byrd and Crawford (2012) outline much of the history and research done in the area, as well as the problems that were being discovered through these studies. The largest challenge facing researchers in this area is the ability to address polyphonic music. Since much of the information produced by audio data relies heavily on time-stamped events, Byrd and Crawford explained that information on individual voices is usually not available. A fundamental problem with MIR is the primary focus on pitches and contour, and without the ability to differentiate between voices in polyphonic music, this information is inadequate. This paper largely serves as a literature review and assists in painting a picture of where research on this topic is headed and what still needs to be addressed by future studies.

Several studies have examined melodic similarity through a variety of approaches. Welker (1982) asked participants to draw an abstract contour line representing the theme after hearing several transformations of a melody. He found that both novice and expert musicians were able to complete this task. Deliège (2001) also investigated how listeners categorize themes. In her study, participants heard the opening to a J.S. Bach violin sonata, followed by either parts of what they had already heard, parts from an unheard section of the piece, or parts that had been altered in pitch and rhythm. They were then asked to determine whether or not they had heard the excerpt before. Deliège found that unheard items were commonly judged as "heard items" due to their similarity to the opening of the piece, and that alterations which were unstylistic were easily detectable as not belonging. Finally, studies by Lamont and Dibben (2001) and Ziv and Eitan (2007) required participants to sort musical excerpts into two categories based on which theme they belonged to. These studies demonstrate a common philosophy for how humans perceive and understand similarity: both studies concluded that similarity perception relies on "surface elements," including texture, orchestration, register, and pace.

The present study builds on this prior research but takes the task of melodic similarity in a slightly different direction. First, this paper is less focused on understanding human perception of similarity in order to test or build upon computational algorithms. Instead, the goal of this research is to understand the listening experience as it relates to theme and variation pieces with hopes of understanding which musical factors affect a listener's ability to perceive similarity between repeated musical ideas. In light of this goal, another key difference between the present study and other research is presented: with an emphasis on the listening experience, this study sets out to test musical factors beyond pitch, rhythm, and contour in polyphonic music. Although these factors are the primary focus of many of the above studies, there are more aspects to the perception of similarity in classical, orchestral music than just pitch and rhythm. Timbre, harmony, and ornamentation are examined in this paper to see what role these play in the perception of similarity. Finally, this study does not examine similarity across a genre as a whole, but rather similarity as a linking device within theme and variation pieces. This research hopes to contribute to the knowledge of how humans perceive similarity and, ultimately, comprehend music.

Experiment 1

Participants

Although untrained musicians would be able to perform the task of rating similarity between two melodic excerpts, it is probable that formal musical training would influence their decisions. In order to narrow the scope for this pilot study, only trained musicians were recruited. The participant pool contained twelve trained musicians (7 female, M=24.1 years) from Southwestern Ontario and Michigan State University. There were seven females and five males. All participants had a minimum of four years of music studies and two years of collegiate music theory instruction to classify them as trained musicians. The average years of music study was 9.46 and the average years of collegiate music theory instruction was 4.77. No participants self-identified as having absolute pitch. All participants were right-handed, and all are first-language English speakers, although these factors are unlikely to play a significant role in the results.

Stimulus Materials

The stimuli consisted of two theme melodies from lesser-known Beethoven Theme and Variation pieces, WoO 64, "Six Easy Variations on a Swiss Song" (Figure 1a) and WoO 77, "Six Easy Variations on an Original Theme" (Figure 1b). These pieces were chosen with hopes that most participants would not have heard them before in an attempt to control similarity ratings based on familiarity with the variations. Two different themes were employed in the study in order to assess similarity judgement consistency. Both themes were rendered in MIDI piano timbre, and were then altered by three factors: harmony, timbre, and ornamentation. Apart from timbre, the basis of each of permutation was drawn from the variations composed by Beethoven, although not always an exact rendition. Each melody was paired with two different harmonizations and the texture was controlled so that listeners were not presented with block chords in one version and single note harmonizations in the other. Each melody was also heard in three additional MIDI timbres: flute, trumpet, and violin. Finally, each melody was ornamented to three varying degrees based on Beethoven's composed variations (variations one, five, and six from both WoO 64 and WoO 77) with slight modifications.

a)

b)

Figure 1. The two themes used and altered for the stimuli. Melody A is the theme from Beethoven's WoO 64 and Melody B is the theme from Beethoven's WoO 77.

Procedure

Participants were sent a link to Qualtrics, an online survey and research data collection platform hosting the study. After answering a short demographics survey, they were presented with 46 pairs of permutations to be compared. Each pair consisted of one excerpt followed by five seconds of silence and then the second excerpt. Participants were asked to rate how similar the pair were on a scale from one to seven, with one being "not very similar" and seven being "very similar."

Results

In order to determine which of the factors (harmony, timbre, or ornamentation) played a role in perception of melodic similarity, an Analysis of Variance was completed on the data. This analysis includes results from both melodies and reveals a statistically significant difference in the overall average ratings of harmony (H), ornamentation (O), and timbre (T) (see Figure 2).

As the graph shows, timbre received the highest average similarity ratings, averaging around 6.5 on the seven-point scale; harmony was next, receiving an average of about 5.5 on the scale; finally, ornamentation was overall rated as having the lowest similarity of the three factors, averaging around 3.5 on the similarity scale. This means that if timbre was the factor being varied, participants were not likely to hear it as a variation but rather considered it to be very similar to the original melody. If ornamentation was the varied factor, on the other hand, participants were more likely to hear it as a variation and dissimilar to the original melody.

A pairwise probability was also examined for these individual variables. There was not a significant difference between ratings for Ha and Hb (p=0.97), however there were significant differences between Ha and each of the other variables, as well as Hb and the other variables. Interestingly, Oa did not have a significant difference between Ob (p=0.2938) and C (p=0.2715), however Ob and C were significantly different from one other (p=0.0018). All ornamentations were significantly different from the three timbre variations, and the timbres were not significantly different from one another. Figure 3 shows a visual representation of the pairwise probabilities for the individual variables.

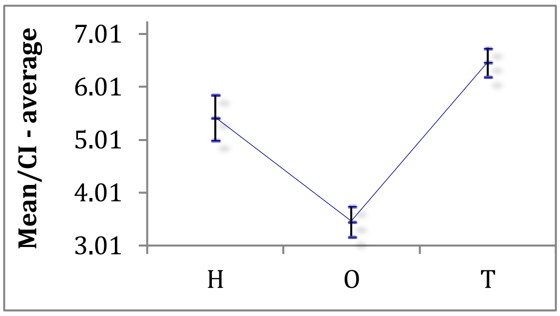


Figure 2. An Analysis of Variance revealed statistically significant differences between the overall average ratings of harmony (H), ornamentation (O), and timbre (T) across both melodies.

Finally, Melody A and Melody B were compared to note any consistencies or discrepancies between the similarity ratings for each variable across the two melodies. The interactions can be seen in Figure 4. This graph shows a remarkable consistency in similarity ratings for all three timbres and both harmonies across Melody A and Melody B. This leads to the conclusion that regardless of which timbre or harmony was varied, or which melody was heard, participants were rating timbral differences as very similar and harmonic differences as slightly less similar but still consistent. Ob also has consistent ratings in Melody A and Melody B, however Oa and Oc are drastically different. This leads to the question, what is different about Melody A Oa, and Melody B Oa? What about Melody A Oc, and Melody B Oc? In order to determine what specific factors in these ornamentations creates a different perception of similarity, a critical look at these ornamented variations is required.

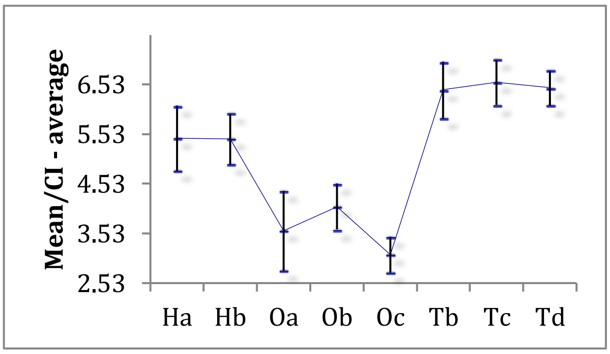


Figure 3. Pairwise probabilities for individual factors reveals statistically significant differences between similarity ratings of different variable kinds (harmony, ornamentation, timbre) but not between the two harmonies or three timbres. Additionally, Ornamentation B and C had a statistically significant difference.

Discussion

As the results show, most listeners had very similar perceptions of timbral and harmonic differences across Melody A and Melody B. On average, listeners perceived timbral changes to be insignificant—these alterations to the melody were rated very highly on the similarity scale.

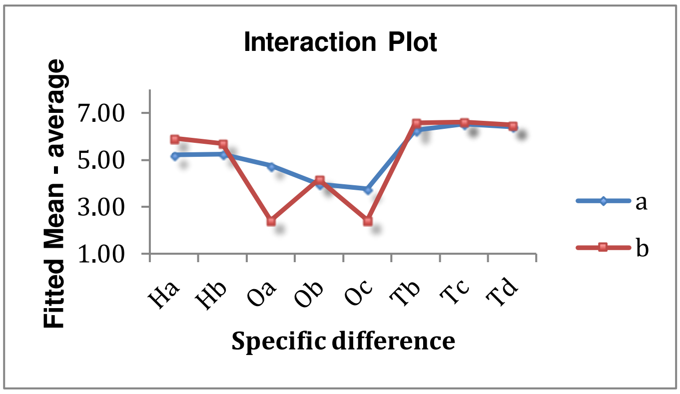


Figure 4. An interaction graph demonstrates the similarity rating consistencies for each variable across the two melodies, except for Ornamentation A and Ornamentation C.

Harmonic alterations were perceived as less similar than timbral changes, however the average ratings for harmony still remained high on the similarity scale. It can be concluded that these factors play little role in the perception of melodic similarity: changing the timbre or the harmony without any alterations to the melody allowed listeners to perceive similarity with very little hindrance.

Ornamentation, however, seemed to play a significant role in the perception of melodic similarity. Since this study uses the term "ornamentation" very broadly to describe any alterations to the melody, including rhythmic differences, non-chord tones, and chromaticism, a further look into what individual factors could be at play in each ornamentation would provide more information on which specific factors should be examined in future studies.

Oa is where consistency in ratings between the two melodies started to diverge. In Melody A Oa was rated the highest on the similarity scale of the three ornamentations. After examining the variation in detail, a possible reason for why this might be becomes evident: although a triplet rhythm is introduced as well as some chromatic passing tones, the upper-most melodic line remains the same as the theme. In this case, it seems as though the other notes are creating more of a harmonic setting than interfering with the perception of the melody, so listeners are still able to latch onto the melodic line and recognize its similarity to the original theme.

Oa of Melody B is when similarity ratings declined substantially. When comparing the theme to this variation, a few key factors stick out as possibly having an influence on the perception of similarity. First, a new rhythm is consistent throughout this variation (all 16th notes) and many strong beats are dropped, creating a sense of syncopation. The lack of strong beats could be disorienting for the listener, especially in comparison to the original theme, which was grounded in recurring eighth notes and a steady, predictable beat. Although the overall harmonic motion of the theme is retained in this variation (i.e. a half cadence is reached in measure 4, and a perfect authentic cadence is reached at the down beat of measure 8), as well as some of the same notes in relatively the same place, some of the notes and harmonies in corresponding spots between theme and variation are different. These factors could be obscuring the original melody so much that listeners are beginning to hear this variation as quite dissimilar to the theme.

The final ornamentation of both melodies, Oc, was rated the lowest on the similarity scale for both Melody A and Melody B, although significantly lower ratings were given to Melody B Oc than Melody A. Taking a closer look at Melody A, some prominent features of this variation that could be impacting the perception of similarity are the chromaticism and rhythms that are not present in the original melody. Although the melodic notes from the theme can be found in much of this variation, the melody seems to be obfuscated by the other notes. This leads to the conjecture that in order for the melody to be heard and noticed as similar to the original, it needs to be in the highest sounding voice.

Finally, Oc of Melody B was rated the lowest on the similarity scale of all the variations in the study. Just from looking at the score, the differences between the two excerpts are obvious: the ornamented variation consists of four times as many notes per measure, which makes the overall pace seem quicker than the original theme. Like in Melody A Oc, this variation incorporates a lot of chromaticism and the melodic notes of the theme are complete hidden by the sheer number of new notes present. Another factor at play is the widening of the overall range—this variation expands to covering just short of three octaves, which starkly contrasts the less-than-an-octave range of the theme. The combination of these factors makes it unsurprising that this variation was perceived, on average, as very dissimilar to the original melody.

Experiment 2

One factor that was not examined in the first study was the comparison between two excerpts that belonged to different themes. This type of study was completed by Ziv and Eitan (2007) as well as Alexandra Lamont and Nicola Dibben (2001), and is required to determine whether listeners are relying on surface elements of the music (such as surface rhythm, pitch, and contour) or structural elements of the music (such as underlying harmonic progressions) when determining similarity. Our second study attempted to complete this task, however in a different way than previous studies—rather than asking listeners to categorize variations by which theme they belong to, we simply asked participants to rate how similar each variation was to two different themes. By asking participants to rate similarity between excerpts that do and do not belong to the same piece as a given theme, we hoped to explore the full range of the similarity possibilities (i.e. those that are extremely similar and those that are completely different) as well as gain more insight on what kind of listening strategies listeners utilize when determining similarity.

Participants

The participant pool contained twelve trained musicians ranging in age from 23 to 40 (8 female, M=28.17) from Southwestern Ontario and Michigan State University. All participants had a minimum of 4 years of music studies and 3 years of collegiate music theory instruction to classify them as trained musicians. The average years of music study was 13.83 and the average years of collegiate music theory instruction was 6.95. No participants self-identified as having absolute pitch. All but one participants were right-handed, and all are first-language English speakers, although these factors are unlikely to play a significant role in the results. 4 participants took part in both this study as well as the first, and only one participant was familiar with the musical excerpts, although this did not seem to have any effect on the results of the study.

Stimulus Materials

The stimuli consisted of two theme melodies from theme and variation pieces, one composed by Mozart (K.354) and the other by Beethoven (Op. 35). Additionally, 4 variations were used from each piece focusing on four variation techniques: 16th notes in the right hand, octaves in the left hand, continuous triplet rhythms, and minor mode setting. Both themes were in E-flat major, in duple meter, eight measures long, and reached a tonicized half cadence in measure eight. All excerpts were rendered in MIDI piano timbre (see Figure 5a and 5b).

a)

b)

Figure 5. The two themes used for the second study. Melody A is Mozart's K. 354 and Melody B is Beethoven's Op. 35.

Procedure

Participants were presented with one theme and all eight variations and asked to rate the similarity of each variation to the theme on a scale from one to seven, with one being "completely different" and seven being "exactly the same." This process was then repeated with the second theme and all eight variations. Participants had access to the theme while rating all eight variations and were able to go back and listen to it as often as they wished. Following the study, participants answered a short demographics survey, including whether or not they had heard either of the themes before the study, and whether or not they took part in the first phase of the study.

Results

Figure 6 below shows the average similarity ratings (with seven being "exactly the same" and one being "completely different") for each individual variation. The first eight variations were compared against the Beethoven melody (abbreviated to "B.") and the last eight were compared against the Mozart theme (abbreviated to "M."). This graph reveals that there was no difference observed by the participants when they were listening to a variation that belong to the theme and when they were listening to one that did not.

The same information can be observed in Figure 7. This ANOVA compares variations that were categorized as belonging to the same theme (categorized as "S") and those that belonged to a different theme (categorized as "D"). Once again, it can be observed that there is no significant difference in the ratings when participants were listening to variations that did and did not belong to each theme.

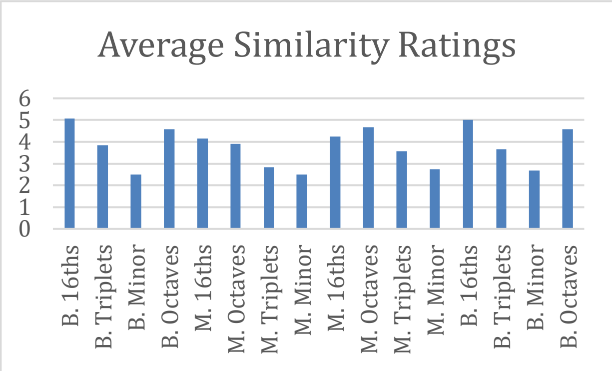


Figure 6. Average similarity ratings for each individual variation. B. represents variations from the Beethoven theme, and M. represents those from the Mozart theme. The first eight variations were compared against the Beethoven theme, and the final eight were compared against the Mozart theme.

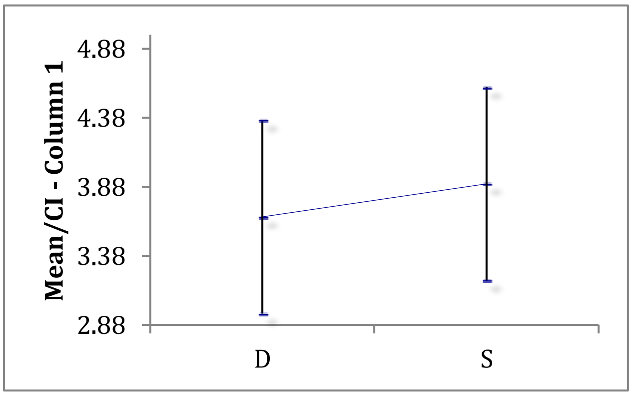


Figure 7. An ANOVA also reveals no difference in similarity ratings between the variations that belong to the heard theme ("S") and those that belong to a different theme ("D").

From Figure 6, additional trends can be observed. As the graph shows, the variations in the minor mode were consistently rated lower than those in the same tonality as the theme, averaging around 2.5 on the similarity scale. The next lowest rating was consistently given to the triplet variations from both themes. Finally, the 16th note variations and those that featured octave leaps in the left hand were consistently rated higher on the similarity scale, with the 16th notes usually being rated as more similar than the octave variations.

A second ANOVA was completed to examine the differences between each of the varied elements (16th notes, octaves, triplets, and minor mode) in order to determine if there were any significant differences between these categories. The results are shown in Figure 8. This analysis revealed significant differences between the minor mode variations compared to all others, the triplet variations compared to all others, and no significant differences between the octave and 16th note variations (p=0.4950).

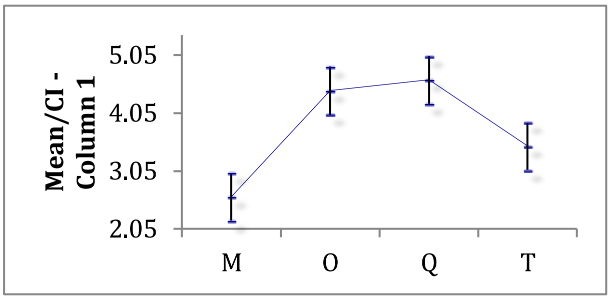


Figure 8. An ANOVA revealed statistically significant differences between the overall average ratings of the minor mode variations (M) and the triplet variations (T), but no significant difference between the octave variations (O) and the 16th note variations (Q).

Experiment 2 Discussion

As the results show, listeners were unlikely to be listening to structural elements of the variations when determining their similarity ratings. Since the excerpts that similar in their variation technique (i.e. those that were varied by use of minor mode) received consistent similarity ratings, regardless of whether they were being compared to the theme they belonged to or not, it appears that these surface elements had a more significant impact on their ability to perceive similarity. If structural elements played a more significant role, than a clear distinction between similarity ratings given to the variations that belonged to the given theme and those that did not would arise.

As Figure 8 exhibits, there seemed to be a clear and significant difference for participants when listening to the minor mode variations and the triplet variations in comparison with the similarity ratings given to the octave variations and the 16th note variations. It can be conjectured that the contrast between mode cued listeners into a difference and was more of a focal point than the structural melody that was still prevalent within these melodies. Additionally, when examining the minor-mode variations, there are other factors at play that could have contributed to the significantly lower similarity ratings: in the Beethoven variation, the main melodic material in the right hand is derived from the left hand of the theme. Similarly, the melodic line of the minor-mode Mozart variation is also changed from the theme. Although the overall harmonic structure of these variations remains, the surface elements of change in mode and change in melodic content could explain the significantly lower ratings for these variations.

The triplet variations in both themes were also received statistically-significant differences in their similarity ratings compared to the others. It can be conjectured that these lower similarity ratings are the result of a faster surface rhythm. Furthermore, the melody of these variations, like those in the minor mode, feature a more obscured melody, which could contribute to the lower similarity ratings, despite the similarities in contour and overall harmonic motion.

Overall, the variations featuring 16th-notes in the right hand and octave leaps in the left hand showed no significant difference in their similarity ratings. On average, these two variations were consistently rated the highest on the similarity scale, with the 16th note variations usually being rated as the most similar to the themes. In each of these cases, the original melody from the theme was most clearly stated once again, which could have helped the listener notice similarity between the variations and the themes.

Interestingly, when comparing the results of this study with the results of the first study, there seems to be a discrepancy in terms of the results for a change in harmony. The first study revealed that changes in harmonization were not perceived as a variation, however, the harmonization of the melody into the minor mode in this study had the most significant impact. This could reveal the difference between an overall change in tonality (as observed in the second study) and a change in single chords with the same function (as observed in the first study). It can be conjectured, based on this small sample size, that altering harmonies while retaining the same key and overall harmonic motion has little effect on the ability to perceive similarity between two excerpts, however reharmonizing the melody into a minor mode can have a significant effect on this perception.

Conclusion

From this study, three main conclusions can be drawn and several areas for further research can be identified. First, it can be concluded that, with this small sample size, ornamentation such as 16th notes in the melody or octave leaps in the accompaniment has little effect on the human perception of melodic similarity, as enough of the original melody is retained. When these factors are altered but the melody remains the same, listeners perceive significantly less of a difference than when the melodic notes are altered. Overall, adding 16th notes is the least detrimental to the perception of similarity, and octave leaps in the accompaniment are not far behind. The second conclusion that can be drawn is that minor mode harmonizations and triplet rhythm ornamentations have a significant impact on a human's ability to perceive similarity between melodic excerpts. In these variations, it is likely that the melodic material from the original theme was masked by the change of mode and of rhythm so that listeners did not hear them as similar. Finally, it can be conjectured that listeners use surface elements, such as mode and rhythm, when determining melodic similarity more than they use structural elements, such as overall harmonic progressions and structural melodies.

This leads to possible areas for further research. First and foremost, running this study again with more participants will determine whether these results are consistent amongst many listeners. It would also be interesting to examine both trained and untrained musicians to determine if listening strategies between these two groups are the same or different. Additionally, it would be ideal in future studies to compose variations so that each was controlled to only one variation technique. Since the variations composed by Beethoven and Mozart contain multiple variation techniques (i.e. the minor mode variation also has a change in melody), it is difficult to determine exactly what listeners are perceiving and what strategies they are using when determining their similarity ratings. Narrowing the focus more than the current study will provide better insight and understanding to how humans judge melodic similarity, and, ultimately, understand musical forms.

Acknowledgements. We'd like to thank the College of Music and the Graduate School at Michigan State University for their support of this research.

References

Arnold, Denis et al. (n.d.). Form. *The Oxford Companion to Music. Oxford Music Online.* Oxford University press. Retrieved from http://www.oxfordmusiconline.com/subscriber/article/opr/t114/e2624.

Byrd, D., & Crawford, T. (2002). Problems of music information retrieval in the real world. *Information processing & management*, *38*(2), 249–272.

Cahill, M., & O'Maidín, D. (2006). Assessing the Performance of Melodic Similarity Algorithms Using Human Judgments of Similarity. In *ISMIR* (pp. 355–356).

Cahill, M., & O'Maidín, D. (2005). Melodic Similarity Algorithms--Using Similarity Ratings for Development and Early Evaluation. In *ISMIR* (pp. 450–453).

Deliège, I. (2007). Similarity relations in listening to music: How do they come into play?. *Musicae Scientiae*, *11*(1_suppl), 9–37.

Devaney, J., Arthur, C., Condit-Schultz, N., & Nisula, K. (2015, October). THEME AND VARIATION ENCODINGS WITH ROMAN NUMERALS (TAVERN): A NEW DATA SET FOR SYMBOLIC MUSIC ANALYSIS. In *Proceedings of the 16th ISMIR Conference* (Vol. 729).

Lamont, A. & Dibben, N. (2001), Motivic structure and the perception of similarity, *Music Perception,* 18, 245–274.

Margulis, Elizabeth Hellmuth. (2014). *On Repeat: How Music Plays the Mind.* Oxford: Oxford University Press.

Marsden, A. (2012). Interrogating melodic similarity: a definitive phenomenon or the product of interpretation?. *Journal of New Music Research*, *41*(4), 323–335.

Melen, M. & J. Wachsmann, J. (2001), Categorization of musical motifs in infancy, *Music Perception,* 18, 325–346.

Müllensiefen, D., & Frieler, K. (2004). Cognitive adequacy in the measurement of melodic similarity: Algorithmic vs. human judgments. *Computing in Musicology*, *13*(2003), 147–176

Müllensiefen, D. & Frieler, K. (2007). Modelling experts' notions of melodic similarity. *Musicae Scientiae*, Discussion Forum 4A, 183–210.

Schmuckler, M. A. (1999). Testing models of melodic contour similarity. *Music Perception: An Interdisciplinary Journal*, *16*(3), 295–326.

Typke, R., de Hoed, M., de Nooijer, J., Wiering, F., & Veltkamp, R. C. (2005). A ground truth for half a million musical incipits. *Journal of Digital Information Management, 3*(1), 34–39.

Volk, A., & Van Kranenburg, P. (2012). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae*, *16*(3), 317–339.

Volk, A., de Haas, W. B., & Van Kranenburg, P. (2012). Towards modelling variation in music as foundation for similarity. *Proc. ICMPC*, 1085–1094.

Welker, R.L. (1982), Abstraction of Themes From Melodic Variations*, Journal of Experimental Psychology*: *Human perception and Performance*, 8, 435–447.

Ziv, N. & Eitan, Z. (2007), Themes as prototypes: Similarity judgments and categorization tasks in musical contexts, *Musicae Scientiae*, Discussion Forum 4A, 99–133.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Are musicians at an advantage when processing speech on speech?

Elif Canseza Kaplan,[*#1] Anita Wagner,[*#2] and Deniz Başkent[*#3]

[*] *Research School of Behavioral and Cognitive Neurosciences, Graduate School of Medical Sciences, University of Groningen, Netherlands*

[#] *Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Netherlands*

[1]*e.c.kaplan@rug.nl,* [2]*a.wagner@umcg.nl,* [3]*d.baskent@umcg.nl*

Abstract

Several studies have shown that musicians may have an advantage in a variety of auditory tasks, including speech in noise perception. The current study explores whether musical training enhances understanding two-talker masked speech. By combining an off-line and an on-line measure of speech perception, we investigated how automatic processes can contribute to the potential perceptual advantage of musicians. Understanding the underlying mechanisms for how musical training may lead to a benefit in speech-in-noise perception could help clinicians in developing ways to use music as a means to improve speech processing of hearing impaired individuals.

Introduction

Earlier studies have shown that musical training can grant normal-hearing listeners an advantage on auditory tasks, in particular for speech comprehension in noise or in the presence of background talkers (Başkent & Gaudrain, 2016; Kraus, Strait, & Parbery-Clark, 2012; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Swaminathan et al., 2015). However, not all studies have found a difference between musicians' and non-musicians' speech in noise performance (Boebinger et al., 2015; Ruggles, Freyman, & Oxenham, 2014), or they have found an advantage in some auditory tasks and not in others (Fuller, Galvin, Maat, Free, & Başkent, 2014). Taken together, various studies addressing a musician advantage provided mixed results (for a review see: Coffey, Mogilever, & Zatorre, 2017), which can be partly explained through the use of different measures (e.g. behavioral versus physiological, online versus offline), the variations in target/noise signal properties, and the linguistic complexity of the target and noise stimuli used in the tasks.

In the current study, we aimed to investigate how perception of speech in two-talker masker noise differs between musicians and non-musicians. We first used a sentence-recall task, in which participants recall and repeat target Dutch sentences presented with two-talker Dutch sentence maskers. This task provides an estimate of sensory sensitivity in different Target-Masker-Ratios (TMR) by estimating the percentage of correctly recalled words. This is an offline-measure of speech processing, similar to the single-talker masker study reported by Başkent & Gaudrain (2016). Following, we used a visual-world paradigm, using eye tracking (Cooper, 1974; for a review see: Salverda & Tanenhaus, 2017). Here, while listening to sentences, participants visually search for the image of the target word mentioned by the target speaker. The target image is displayed among a competitor (phonologically similar to the target, i.e. ham-hamster) and two distractor images. During the task, gaze fixations and pupillary responses are measured. This task gives an online measure of speech processing by comparing the time course of gaze fixations to the target and/or the competitor word, as well as the changes in event related pupil dilations. The online measure indicates how quickly participants integrate the acoustic information in the signal when they are accessing the mental lexicon and the extent of the mental effort involved in processing of linguistic information. We used the utterances of the same target speaker and the masker speaker in both tasks in order to be able to capture how the automatic processes may contribute to speech on speech perception in musicians.

Our main research questions are:

1. Would the difference observed in an offline task, manifest itself in an online task?
   - Musicians have been previously shown to perform better than non-musicians in single talker masker (Başkent & Gaudrain, 2016) or four-talker babble (Slater & Kraus, 2016). Based on literature, we expect musicians to be less affected by masked speech than non-musicians on the offline task.
2. Would the results obtained by an online task be clearer in reflecting the difference in the usage of cognitive resources in musicians and non-musicians?
   - Speech masker should have a smaller effect in terms of timing of lexical mapping and pupil dilations for musicians than non-musicians

Method

Participants. Ten musicians (years of musical training = 10.6, age = 23.8) and twenty non-musicians (years of musical training = 1.63, age = 25.1) from Groningen, the Netherlands were recruited for participating in the study. The musicianship criteria were taken from the literature (Fuller et al., 2014) and included the following: having started music before the age of 7, having at least 10 years of musical training, actively practicing music at least for 3 years prior to the study. The non-musician criteria were not meeting the musician criteria and not having more than 3 years of music training.

Materials. The same target and masker speakers' sentences were utilized in both tasks.

In the sentence recall task, 28 semantically neutral Dutch target sentences (Wagner, Toffanin, & Baskent, 2016) uttered by a female speaker were embedded in two-talker maskers. We chose to use a two-talker masker based on previous literature. According to Rosen, Souza, Ekelund, & Majeed (2013), greatest changes in masking occur when the number of talkers change from 1 to 2 or 4 background talkers, and Calandruccio, Buss, & Bowdrie (2017) showed that a two-talker masker was the most effective masker. The masker sentences consisted of simple, meaningful Dutch sentences (Versfeld, Daalder, Festen, & Houtgast, 2000) uttered by a different female speaker than the target speaker.

The onset of the target sentences was delayed by 500 ms from the maskers' onset, and the participants were instructed to repeat the sentence that started later. The sentences were presented at four possible TMRs in a random order; -3, -5, -7 or -9 dB SPL (Calandruccio et al., 2017), where the presentation level of the masker sentence combination was fixed at 75 dB SPL and the presentation level of the target sentence was adjusted depending on the TMR in each condition.

Based on pilot results of the sentence recall task, the eye-tracking experiment's TMR levels were set to 0 and -5 dB SPL. The stimuli were presented either in a two-talker masker with 0 or -5 TMR and, as a baseline in quiet (without any masking).

Procedure. Participants were first tested for normal hearing, where the hearing levels were < 20 dB HL for pure tone thresholds from 250 to 4000 Hz bilaterally, then first completed the visual world paradigm, and following the sentence recall task.

In the visual world paradigm, the experiment started with a practice phase where participants completed 8 trials of the quiet and masked conditions. They were instructed to pay attention to the voice they heard in the quiet condition throughout the experiment and choose the image of the target word from four images displayed on the screen. The quiet condition was presented first at all times. Upon completing the quiet condition, participants could take a break and then complete the masked condition. The two TMR levels were presented in a random order within one block. Each condition included 12 trials. The order of presentation of the sentences was randomized for each participant.

The sentence recall task began with a practice phase of 4 trials, where participants heard an example of each TMR level (-3, -5, -7, -9 dB). They were instructed to repeat the sentence uttered by the same target speaker as in the first experiment. After the practice, participants completed 28 trials, with TMR conditions presented randomly in order to eliminate learning effects.

Results

The preliminary results will be reported; however, none are conclusive since the musician data is based on 10 participants.

In the sentence recall task, musicians perform slightly better as the level of background masker speech increases (Figure 1).

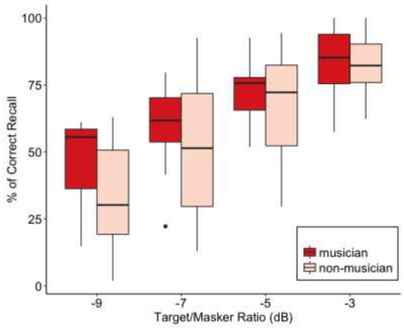


Figure 1. Percentage of correctly recalled keywords as a function of target masker ratios (-9, -7, -5, -3 dB) of the sentence recall task.

The gaze fixations in the quiet condition (Figure 2) indicate that lexical competition occurs in a similar fashion both for musicians and non-musicians. The point of disambiguation is happening earlier in quiet than in masked conditions for both groups. The proportion of fixation to the target is also higher for both groups in quiet versus in masked speech.

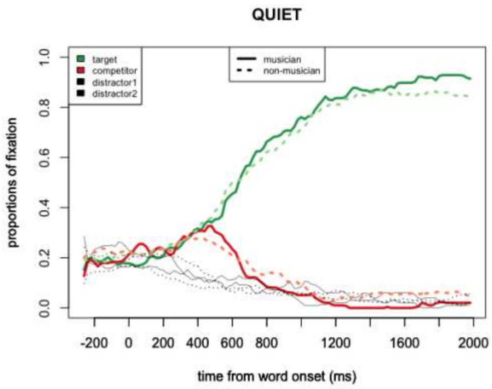


Figure 2. Gaze fixations to the target, competitor and distractor images as a function of time from word onset of the target word in the quiet condition.

At TMR=0 dB, where the intensity of the masker and the target are equal, the point of disambiguation seems to be slightly delayed for non-musicians. The lexical competition is still present for both groups, while the certainty is lowered for both groups as shown by the proportion of fixation to the target (Figure 3).

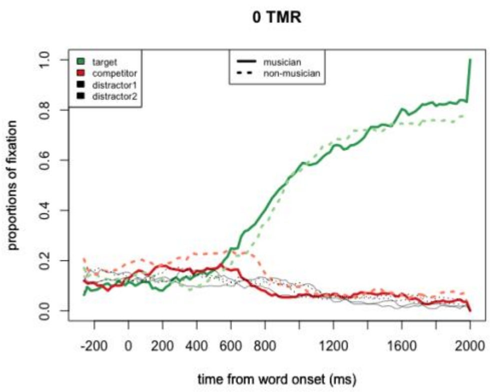


Figure 3. Gaze fixations to the target, competitor and distractor images as a function of time from word onset of the target word in TMR=0 dB condition.

In the most difficult TMR, where the target is presented at 5 dB lower level than the masker the lexical competition is less present for musicians, but not for non-musicians. Overall fixations to the target are reduced compared to the other conditions for both groups, indicating increased uncertainty (Figure 4).

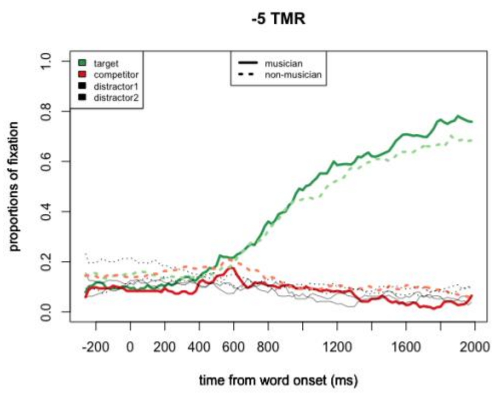


Figure 4. Gaze fixations to the target, competitor and distractor images as a function of time from word onset of the target word in -5 TMR condition.

The event related pupil dilation responses (ERPD) are calculated with the following formula:

$$\% \text{ ERPD} = \frac{\boldsymbol{observation - baseline}}{\boldsymbol{baseline}} * \mathbf{100}$$

The baseline is taken to be the mean of the pupil response occurring 200 ms right before the onset of the target word. The ERPD in response to the processing of the target word reflects that there is less change in pupil size for non-musicians in quiet. Overall, both groups exhibit less change in ERPD in the -5 TMR condition compared to the 0 TMR condition. For the musician group, quiet condition appears to have the biggest change in ERPD across conditions (Figure 5).

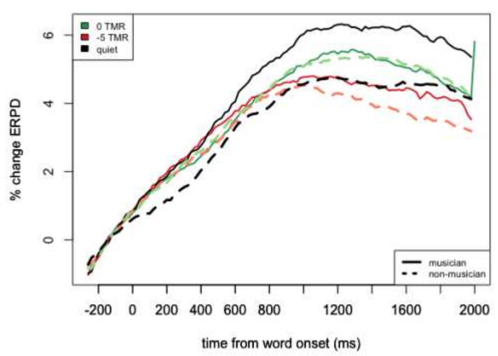


Figure 5. Pupil dilation data time curves for both musicians and non-musicians at all conditions (quiet, 0 TMR, -5 TMR).

Conclusion

The preliminary results so far indicate that musicians may perform slightly better at the sentence recall task as the background speech masker level increases. If this holds as data collection is completed, this finding would be in line with previous findings, despite the difference in masker structure (two talker vs one talker masker) (Başkent & Gaudrain, 2016).

In the visual world paradigm, lexical competition, i.e. listeners' simultaneous consideration of phonologically similar words to be the potential target, which captures automatic process observed in previous research (Salverda, Dahan, & McQueen, 2003; Wagner et al., 2016) is present in the quiet condition, while it becomes less visible as the masking increases, especially for musicians. The overall timing of lexical decision making is delayed for both groups when the target is presented with the masker. The event related pupil dilations, which may be reflecting either increased effort or more attention allocation are inconclusive. We aim to collect more musician data before drawing any further conclusions and conducting statistical analysis.

Acknowledgements. This project is supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie- Skłodowska-Curie grant agreement No 675324, and the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw), VICI Grant (Grant No. 918-17-603). This study has been approved by the Medical Ethics Committee (METc) of University Medical Center Groningen (UMCG).

References

Başkent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America*, *139*(3), EL51-EL56.

Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *The Journal of the Acoustical Society of America*, *137*(1), 378–387.

Calandruccio, L., Buss, E., & Bowdrie, K. (2017). Effectiveness of Two-Talker Maskers That Differ in Talker Congruity and Perceptual Similarity to the Target Speech. *Trends in Hearing*, *21*, 1-14.

Coffey, E. B. J., Mogilever, N., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: a review. *Hearing Research*. *352*, 49-69.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107.

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience*, *8*(179).

Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, *1252*(1), 100–107.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661.

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431–2443.

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of Musical Training on Understanding Voiced and Whispered Speech in Noise, *9*(1).

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89.

Salverda, A. P., & Tanenhaus, M. K. (2017). The Visual World Paradigm. In *Research methods in psycholinguistics: A practical guide* (pp. 89–110).

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. *Cognitive Processing*, *17*(1), 79–87.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports*, *5*(11628), 1–10.

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America*, *107*(3), 1671–84.

Wagner, A. E., Toffanin, P., & Baskent, D. (2016). The timing and effort of lexical access in natural and degraded speech. *Frontiers in Psychology*, *7*(398).

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

In Further Search of Tonal Grounds in Short Term Memory of Melodies

Alperen Karan[1] and Esra Mungan[2]

[1] *Psychological Sciences Graduate Program, Boğaziçi University, Turkey*
[2] *Psychology Department, Boğaziçi University, Turkey*

[1]alperen.karan@boun.edu.tr, [2]esra.mungan@boun.edu.tr

Abstract

Taylor & Pembrook (1983) proposed several factors to affect short-term memory for melodies. We reassessed their findings using a more controlled stimulus set and a 2-alternative forced choice (Experiment 1) or same/different test (Experiment 2) instead of a dictation or singing-back task. Nonmusicians listened to a total of 158 isochronous 5-tone melodies. Each melody was followed by a same-length retention interval filled with silence, nonsense syllables, a nondiatonic melody, or a diatonic melody, and a subsequent test with same-contour lures. In both experiments and across all conditions listeners showed above-chance short-term recognition performance. We replicated Taylor & Pembrook's recency effect for the 5th note of the sequences but also found a full J-shaped serial position curve (recency>primacy>center). Secondly, listeners performed better for tone sequences that were either fully ascending or descending than those with melodic direction changes. Thirdly, listeners were better in noticing a changed note that occurred at a point of melodic direction change (e. g., ∧ or ∨). In Experiment 1 but not 2, we furthermore found that this "corner note effect" was even more pronounced if that note was preceded by a "skip" (3 or more semitones) instead of a "step" (2 or less semitones) pitch interval. The latter finding was somewhat similar to Taylor and Pembrook's finding that listeners were more accurate in their reproduction of skip as opposed to step intervals when they marked a point of change in melodic direction. Fourthly, type of retention interval had a major effect on participants' performances. When tested with a 2AFC setup, the silence group performed best, when tested with a same/different setup, both the silence and nonsense syllable groups performed best. Results are discussed in reference to related STM studies using short tone sequences and Berz' (1995) working memory model for music.

Introduction

Given that most simple songs are built on fairly short melodies, an interesting question arises as to how well listeners retain exact note sequences after first hearing in short-term memory. More specifically, how precise is the immediate configural representation of a short and simple tone sequence, and how much is this precision jeopardized by intervening musical or non-musical events? Are there specific melodic characteristics that lead to more precise and more resistant configural representations? The present study aims to address these questions by expanding upon earlier, related research.

Although encoding of musical information is affected by several organizational factors, which are stored in long-term memory, such as contour, tonality and range (Berz, 1995), several studies showed that around 7-11 notes can be held in short-term memory (Pembrook, 1987). More recent studies with musically untrained participants suggest an average span of 5-7 notes (e. g., Benassi-Werke, Queiroz, Germano, & Oliveira, 2010; Schulze, Dowling, & Tillmann, 2012).

Probably one of the earliest studies to investigate short-term memory for tone sequences was Ortmann's 1933 experiment in an attempt to identify tonal characteristics that boost or reduce short-term memory retention. He used twenty 5-tone sequences as stimuli and asked participants to recall them in a dictation task after each was played to them on a piano. Taylor and Pembrook (1983) did a follow-up on Ortmann's work with the same stimulus set to ensure comparability. But they added a singing-back task as a less demanding memory task and corrected some of the methodological shortcomings of the 1933 study. The simpler singing-back task allowed them to include musically untrained participants to their study. The researchers replicated Ortmann's recency effect finding (i. e., better recall for the last note) for musically trained and untrained participants. Like Ortmann, they also found worse performance when the tone sequences had two or more melodic direction changes compared to sequences with one or none. This was again observed in both musically trained and untrained participants. They also replicated Ortmann's finding of better recall when a melodic direction change occurred at a larger pitch interval ("skip") than a small one ("step", defined as a semitone or whole tone interval). This difference was even greater for musically untrained participants. Taylor and Pembrook explained this interesting "skip-effect" as resulting from attentional processes caused by unexpected skip intervals. Finally, just as in Ortmann's 1933 findings, all participants performed better with ascending than descending sequences.

Even though there were improvements in Taylor and Pembrook's 1983 study, some factors jeopardize the generalizability of their findings. Some sequences, for instance, had full melodic symmetry which possibly served as a strong mnemomic in itself, an issue they acknowledge as a potential pitfall in the Ortmann melody set. Moreover, the fact that all sequences not only started but also ended with a C, makes the recency effect potentially meaningless. Finally, since their corpus had only 20 stimuli, we cannot be sure how generalizable their findings are.

Our goal was to take Taylor and Pembrook's research one step further by using a larger corpus and a short-term memory *recognition* test setup. As such we aimed to bridge Taylor and Pembrook's endeavors with the extensive amount of findings coming from the short-term recognition literature with tonal sequences (e.g., Deutsch, 1970; Dewitt & Crowder, 1986; Dowling, 1973, 1978, 1991; Dowling & Bartlett, 1981; Dowling & Fujitani, 1971; Dowling, Kwak, & Andrews, 1995). A carefully constructed corpus of hundred fifty eight 5-tone sequences was used. All sequences were isochronous and started either with C4 or C5, just like Ortmann's sequences. Further facts such as number of melodic direction changes and note repetitions within each tune were also controlled for. Participants were tested after each 5-tone sequence with either a 2-alternative forced choice (2AFC) test (Experiment 1) or a same/different test (Experiment 2).

In addition, we included a retention interval manipulation. For one group of participants, the retention interval was not filled, for second group it was filled with a sequence of five nonsense syllables, for a third group with a sequence of five nondiatonic tones and for a fourth group with a sequence of five diatonic tones. The purpose of this manipulation was to attempt a comparison between two competing models of short-term memory for music. Whereas Baddeley's model (2000) proposes that both verbal and musical materials are processed in the phonological loop, Berz (1995) proposes a new "slave system" exclusively for musical material. Based on Baddeley's model, we should expect equally reduced short-term recognition performance under all retention interval conditions except the unfilled one. Berz' model, on the other hand, would predict reduced performance for the intervening *musical* sequences but not for the intervening nonsense syllable or silence conditions. We were also curious about how our different retention interval manipulations would affect the expected J-shaped serial position curve. Would they only wipe out mostly the recency portion or create an overall depression of the curve? Earlier, Dowling (1973) had shown that musical material just like verbal material produces J-shaped serial position curves in short-term memory testing. But, to our knowledge, little is known about how the different sections of the curve respond to manipulations such as ours.

Methods

Experiment 1: 2AFC Test

Participants. Seventy-two Boğaziçi University undergraduate students with no to negligible musical background (M= .96; SD=1.33; range 0-4 yrs) participated in return for course credit in introductory psychology courses. None of them had absolute pitch or a hearing impairments.

Stimuli and Apparatus. We created 158 5-tone target-lure melody pairs using MATLAB. All melodies were in the C major scale within the range of C4 and C5. All melodies started with either C4 or C5. Note repetitions occurred in a melody at most twice, and never in succession. Lure items were obtained by changing the second, third, fourth or fifth tone of the target melody by one diatonic note. All lure items preserved the contours of their target melodies. Target-lure identity as well as target-lure position at test were counterbalanced across participants and trials.

We also controlled for the number of melodic direction changes by including a sufficient number of target items for all possible four conditions: zero (in 30 melodies), one (in 48 melodies), two (in 48 melodies), or three (in 32 melodies) direction changes (see Figure 1, for an example with two direction changes). Among sequences without melodic direction changes, half were ascending and half descending. For those with one, two or three direction changes, half started on C4 and half on C5. The base rates of pitch intervals were distributed as 31% "steps" (1 or 2 semitones) to 69% "skips" (> 2 semitones) across all pitch intervals of 158 5-tone sequences. Finally, the likelihood of whether a note was at a point where a melodic direction change occurred or not was 51% to 49%, respectively, across all tunes.

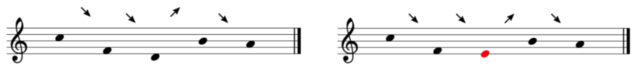


Figure 1: A target and its lure with two direction changes. The first note of the melodies is C5. The lure was created by changing the 3rd note at which a direction change occur in the melody.

For the intervening nonsense syllable condition, 5-syllable sound sequences were created from a set of 23 nonsense syllables via an online text to speech converter (texttospeech.org), and the length of each syllable was adjusted to 0.25s using Adobe Audition 3.0 to match the note durations of the 5-tone sequences. For the intervening melody condition, 158 5-tone diatonic and nondiatonic sequences were created with same pitch averages across five tones. None of these intervening tone sequences contained repetitions, and all were different from the experimental target-lure melodies.

All 5-tone sequences were played with the Steinway grand piano in Logic Pro X software at a speed of 0.25s per note. Over- and on-ear padded headphones were used (Philips SHP 1900, Urbanears Plattan 2, Sennheiser Momentum On-Ear) in cubicles of the Cognitive Processes lab at Boğaziçi University.

Procedure. Participants were tested individually or in groups of up to four. Instructions were followed by four closely monitored practice trials to ensure comprehension of the task. Participants listened to a target melody, which was followed by a retention interval filled with either silence, nonsense syllables, a nondiatonic or diatonic tone sequence. After the retention interval, participants received a 2AFC test where they heard two comparison melodies (Figure 2) from which to pick the target. They were also asked to provide a 3-point confidence rating ("very sure", "sure", "not sure") each time they made their choice. Each experimental session lasted about 35 minutes.

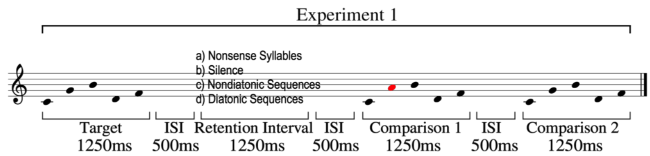


Figure 2: The experimental designs.

Experiment 2: Same/Different Test

Participants. Fifty-eight Boğaziçi University undergraduate students with no to negligible musical background (M= .75; SD=1.25; range 0-4 yrs) participated in return for course credit in introductory psychology courses. None of them had absolute pitch or a hearing impairments.

Stimuli and Apparatus. The apparatus and the stimuli were the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1 except for the recognition test which this time was a same/different test (see Figure 2). Participants were explicitly instructed that the comparison melody which followed the retention interval would either be same or different with an equal likelihood. They were again asked to provide a 3-point confidence rating ("very sure", "sure", "not sure") each time they made their choice. Each experimental session lasted about 25 minutes.

Results

We calculated average performance scores using the proportion of correct responses. The data of seven participants, all from Experiment 2, who showed an average recognition performance of .52 or less but rated 120 or more of their 158 responses as 3 ("very sure") were excluded from the study. We also excluded one participant from Experiment 1 who performed around chance level (.54) but rated all his 158 responses as 3. We took this as an indication that those participants were not using the 3-point confidence rating scale as they should have.

Experiment 1: 2AFC Test

Participants performed above chance level in silence (M=.67, SD=.08, $t(15)$=8.44, p<.001), nonsense syllables (M=.63, SD=.12, $t(19)$=5.00, p<.001), nondiatonic (M=.61, SD=.09, $t(17)$=5.29, p<.001) and diatonic (M=.58, SD=.09, $t(17)$=3.70, p<.01) conditions.

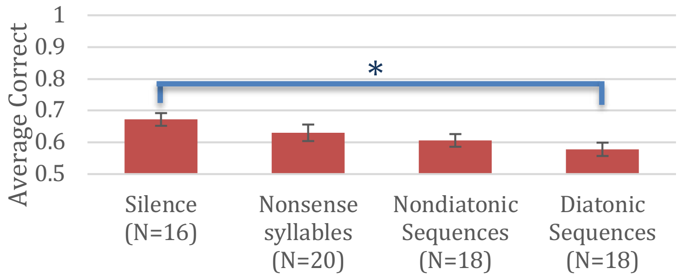


Figure 3. Average correct recognition across groups. Error bars (as in all remaining graphs) indicate standard errors.

A one-way ANOVA revealed a significant effect for type of retention interval, F(3,68)=3.02, p<.05, η_p^2=.12. A Tukey's HSD post-hoc test showed that recognition performance was significantly higher for the silence condition than diatonic sequences condition, p<.05 (Fig. 3).

Although many participants stated written or verbally that the experiment lasted quite long, no fatigue or practice effects were observed. We calculated performances for the first and second half of the trials. A 2 (order) x 4 (group) mixed ANOVA showed that order had neither a main nor an interactive effect on performance, p>.10[1].

Taylor and Pembrook (1983) found that note repetition did not change recall performance. Our findings confirmed their results. A 2 (repetition) x 4 (group) mixed ANOVA revealed neither a main nor an interactive effect for note repetition, all ps>.10.

The lure items contained a tone that was one diatonic tone higher than the target in half of the trials or lower in the other half of the trials. A 2 (repetition) x 4 (group) mixed ANOVA revealed a marginal main effect for type of lure, indicating slightly better performance with lures that had a one diatonic tone up deviation (p=.06).

Participants had a metacognitive understanding of the task. We calculated average performances for low, moderate and high confidence trials (Fig. 4). A 3 (confidence) x 4 (group) mixed ANOVA showed that performances differed between different levels of confidences, $F(1.37,93.21)$=14.60, p<.001, η_p^2=.18. Post-hoc analyses revealed that participants' performances were significantly greater for the high confidence trials than low or moderate confidence trials. No interaction was observed, p>.10.

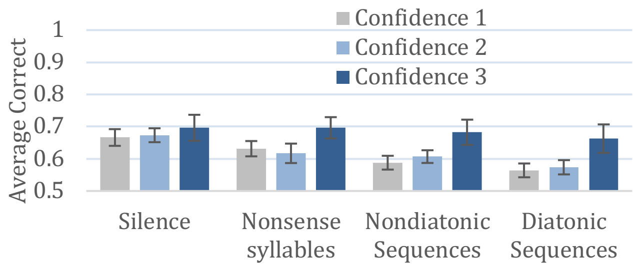


Figure 4. Average correct recognition by confidence across groups.

We also tested the effect of number of melodic direction changes on performance. A 4 (number of direction changes) x 4 (groups) mixed ANOVA showed that number of melodic direction changes had a main effect on performance, F(3,204)=2.94, p<.05, η_p^2=.04. In particular, participants performed higher for melodies without direction changes (which were melodies that were either fully ascending or descending) than melodies with one or three direction changes, p<.05. No interaction was found, p>.10 (Fig. 5). No performance difference was observed between ascending and descending sequences, p>.10.

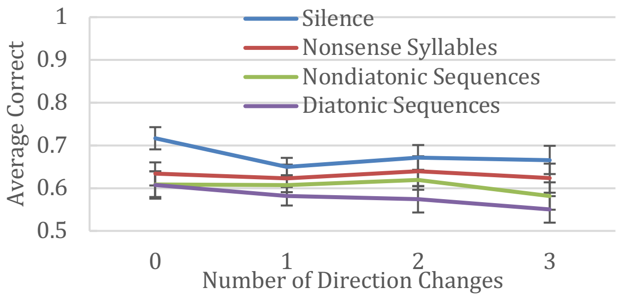


Figure 5. Average correct recognition by number of melodic direction changes across groups.

A 4 (serial position of the changed tone) x 4 (groups) mixed ANOVA revealed a main effect for serial position (F(3, 204)=19.29, p<.001, η_p^2=.22) but no interaction effect (p>.10). Planned comparisons at an overall α-level of .05 confirmed the expected J-shaped curve (recency > primacy > center): changes in the last tone were detected better than changes in the second tone; changes in the second tone were detected better than changes in the middle (third and fourth) tones (Fig. 6).

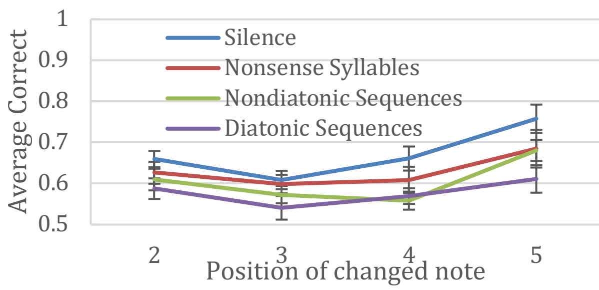


Figure 6. Average correct recognition based on position of changed note within 5-tone sequence across groups.

[1]We also found no main effect for order when splitting trials into three blocks.

In 60 trials, lure melodies differed from target melodies at a note of melodic direction change (see Fig. 1 as opposed to Fig. 2). Note that it is conceptually wrong to talk about the melody's direction at the 1st and the 5th tones. Dyson and Watkins (1984) coined these notes as "corners" and the remainders as "slopes". A 2 (change at "corner" or "slope") x 4 (groups) ANOVA showed that participants were better at detecting note changes at "corners" than "slopes" ($F(1,68)=12.50$, $p<.01$, $\eta_p^2=.16$). No interaction was observed, $p>.10$ (Fig. 7).

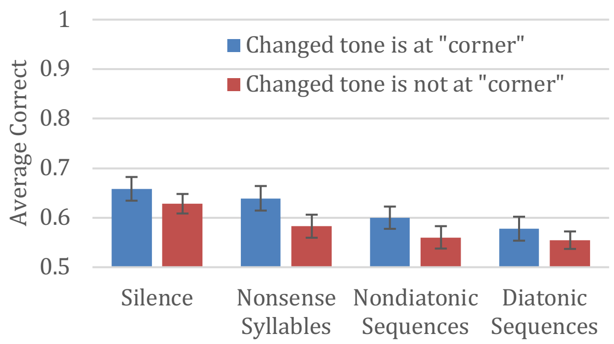


Figure 7. Average correct recognition rates when the change did/did not occur at a "corner" note.

We also replicated Taylor and Pembrook's (1983) "skip effect". For trials where the change occurred at a "corner" note, participants' correct recognition rates were higher if the interval before that note was a "skip" rather than a "step", $F(1,68)=9.08$, $p<.01$, $\eta_p^2=.12$.

Experiment 2: Same/Different Test

We conducted our analyses using area under the receiver operating characteristics curve (AUC). When the research question of interest was about the lure trials, we used correct rejection rates as the dependent variable. For each group, the AUCs were significantly greater than chance level .50, $p<.001$.

A one-way ANOVA showed that discrimination differed between groups, $F(3,54)=7.66$, $p<.001$. Tukey's post-hoc HSD analyses showed that the silence and nonsense syllable groups performed significantly better than the nondiatonic and diatonic sequence groups, $p<.05$ (Fig. 8).

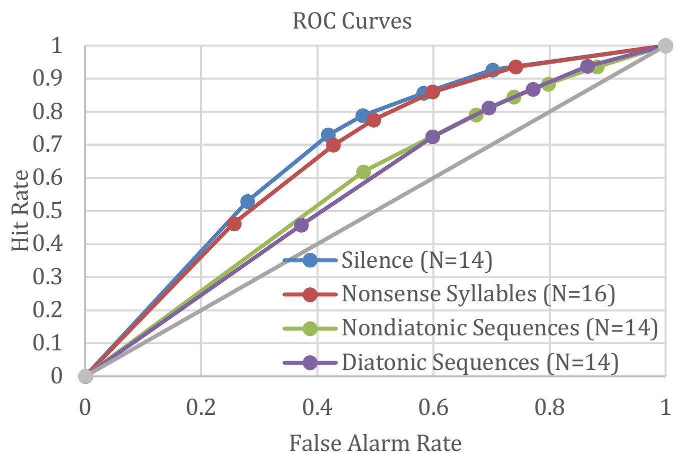


Figure 8. Receiver operating characteristics (ROC) curves across groups.

Participants' performances did not differ in the first and second half of the trials, $p>.10$. We observed a marginal advantage for melodies with a note repetition compared to those without, $p=.07$. Participants performed equally well in trials where the lure item contained a higher or lower tone, $p>.10$. None of these 2 x 4 mixed ANOVAs revealed an interaction effect with type of retention interval.

A 4 (number of melodic direction changes) x 4 (groups) ANOVA showed that number of direction changes had a significant effect, $F(2.55,137.41)=2.89$, $p<.05$, $\eta_p^2=.05$. Participants were better in discriminating melodies that had either no or two direction changes compared to melodies with three direction changes, $p<.05$. No interaction was found, $p>.10$ (Fig. 9). We also did not find any performance difference between ascending and descending sequences, $p>.10$.

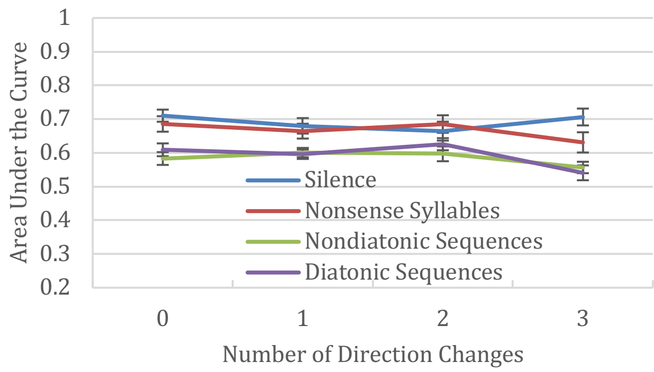


Figure 9. Average AUC values by number of melodic direction changes across groups.

A 4 (serial position of the changed tone) x 4 (groups) mixed ANOVA revealed a main effect for serial position ($F(2.47,133.26)=34.48$, $p<.001$, $\eta_p^2=.39$). Planned comparisons at an overall α-level of .05 confirmed the expected J-shaped curve (recency > primacy > center): changes in the last tone were detected better than changes in the second tone; changes in the second tone were detected better than changes in the middle (third and fourth) tones (Fig. 10).

However, serial position of the changed tone interacted with type of retention interval group, $F(7.40,133.26)=4.83$, $p<.001$, $\eta_p^2=.21$. To better understand the interaction effect, we decided to do a planned comparison in the form of a 2 (silence/nonsense syllable groups vs. non-diatonic/diatonic groups) x 3 (primacy vs. center vs. recency position) ANOVA. We found a main effect for group, such that the combined silence/nonsense syllable group performed significantly better than the combined non-diatonic/diatonic group. This points to an overall depression of performance when the retention interval was filled with *musical* material. Moreover, musical material also curbed the recency effect, as can be seen in Fig. 10.

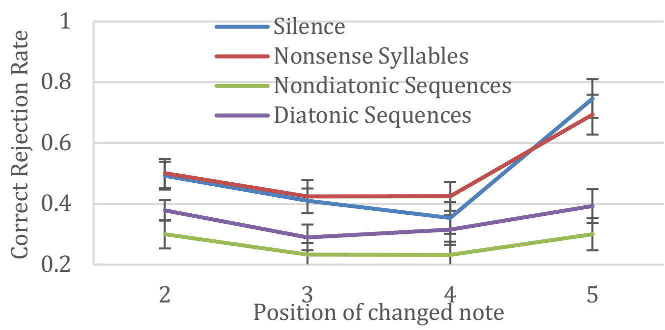


Figure 10. Average correct rejection rates based on position of changed note within 5-tone sequence across groups.

We conducted a 2 (change at "corner" note or "slope") x 4 (groups) ANOVA to see whether the contour-wise position of the changed note had an effect on correct rejection performance. Participants performed better when the change occurred at a "corner" note, $F(1,54)=6.34$, $p<.05$, $\eta_p^2=.11$. No interaction was found, $p>.10$ (Fig. 11).

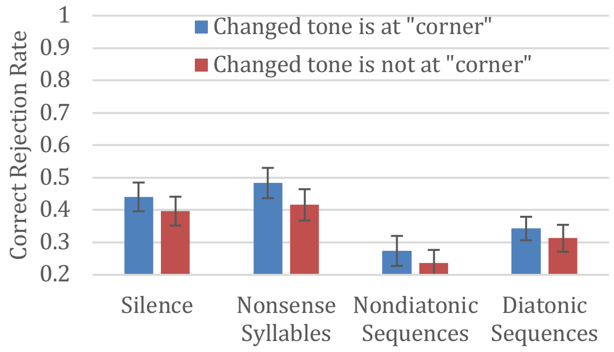


Figure 11. Average correct rejection rates when the change did/did not occur at a "corner" note.

However, we could not replicate the "skip effect" from Experiment 1. When the changed tone was a "corner" note, participants performed equally well regardless of whether the interval before the changed tone was a step or a skip, $p>.10$.

Conclusions

The main motivation of this research was to continue Taylor and Pembrook's (1983) line of research which was a follow-up on Ortmann's 1933 study. The goal of these studies was to detect potential tonal features that improve memorability of short, isochronous 5-tone sequences in a short-term memory (serial) recall setting. Taylor and Pembrook found better recall for (1) melodies that had fewer melodic direction changes (e. g., melodies with up-up-down-down contours were better recalled than melodies with up-down-up-down contours); (2) ascending than descending sequences; (3) notes that were immediately preceded by skips (pitch intervals of 3 or more semitones); and (4) notes in the final position (which was not an meaningful finding since all 5-tone sequences started and ended on a C hence recalling the ending C could have been a simple knowledge-based inference rather than a unique act of remembering).

Our goal, in turn, was threefold. Firstly, we wondered how much Taylor and Pembrook's findings would replicate in a short-term *recognition* test setting. Secondly, we wanted to see how well musically untrained listeners would detect one-diatonic-tone deviations in same-contour lures, which are known to be harder to detect compared to contour-violating lures (e. g., Dowling, 1991). This would also provide us with some clues about the "resolution" of the representation immediately formed after listening to a short isochronous tone sequence. Given that children's songs or lullabies tend to have quite overlapping melodic contours (Trehub, Unyk, & Trainor, 1993), it seems worth pursuing how well a "layperson" attends to, codes, and retains melodic information beyond basic contour. Thirdly, we were curious whether type of intervening stimuli would affect short-term recognition performances as predicted by Baddeley's (2000) working memory model as opposed to Berz' (1995) model. We will start with a summary of our main findings regarding our first goal then move on to the next two points.

Listeners showed above chance short-term recognition after listening to isochronous, diatonic, 5-tone melodies which were tested against same-contour lures with minute one-diatonic-note deviations. This was so regardless of the type of test (2AFC in Experiment 1 and same/different in Experiment 2) and regardless of whether the retention interval between exposure and test was filled with silence, nonsense syllables, nondiatonic or diatonic tone sequences. Second, we not only replicated Taylor and Pembrook's recency effect but also found a full J-shaped serial position curve (recency > primacy > center). Participants were best in detecting a change in the 5th note, second best in detecting a change in the 2nd note, and worst in detecting changes in the 3rd and 4th notes. This was so for all groups and regardless of type of test. Third, in Experiment 2 but not 1, we observed a general depression of the J-shaped curve and a curbing of the recency effect when the retention intervals were filled with musical sequences but not when filled with silence or nonsense syllables. Fourth, we observed that note changes at "corners" (i. e., points of melodic direction change) were better detected than note changes within "slopes" (cf. Dyson & Watkins, 1984). We also replicated Taylor and Pembrook's interesting "skip effect" such that a note change that occurred at the end of a *larger* pitch interval ("skip") was detected (or reproduced, in Taylor & Pembrook's case) better than one that occurred at the end of a *smaller* pitch interval ("step"), provided it marked a point of melodic direction change and tested in a 2AFC test. In other words, when notes formed "corners" in terms of melodic contour, they appeared to be better encoded when they followed a "skip" than a "step" interval. Fifth, we found better short-term memory performance for tone sequences that did not have melodic direction changes, i. e., that were fully ascending or descending. We did not replicate Taylor and Pembrook's (1983) advantage for ascending over descending tone sequences. Finally, whether notes appeared repeatedly (though never in unison) or not did not affect listeners' performance in a 2AFC or same/different test set up (though there was a marginal trend in the latter in favor of note repetition).

The fact that we found above-chance performance in two different short-term recognition setups with only same-contour lures provides evidence for a higher representational "resolution" that incorporates specific pitch interval information. Since we used a recognition instead of recall paradigm our findings also need to be compared to those of Dowling (1978; 1991) and his colleagues (e. g., Dowling, Kwak, & Andrews, 1995), who used comparable isochronous 5- or 7-tone sequences in partly comparable recognition test setups. The closest experiment to ours would be Experiment 1 in Dowling and Fujitani (1971), which showed that when using untransposed same-contour lures in a 2AFC setup with isochronous 5-tone sequences, listeners with mixed musical background were successful in discriminating targets from lures [2] (with a mean area-under-the-memory-operating-

[2] Although Dowling and Fujitani's participants performed close to ceiling in the untransposed condition, this was not the case for our participants. This could be due not only to their musical background-wise mixed participant profile but also their use of only 60 tone sequences (whereas we used 158 tone sequences).

characteristics-curve of M_{AUC}=.92). In all other studies, Dowling and colleagues used transposed targets only, to exclude any "help" from pitch memory. When using transposed targets only, musically trained and untrained participants dropped to chance performance in a same-different task (Dowling, 1978). Above chance performance in discriminating transposed targets against transposed same-contour lures was obtained in later studies with 7-tone sequences when changing *two* instead of a single note in a continuous-running-memory task (e. g., Dowling, 1991; Dowling, Kwak, & Andrews, 1995; also cf. Dewitt & Crowder, 1986). However, because there are many parametric differences between their studies and ours (e. g., their tone sequences had only small pitch intervals, all their ending notes were twice as long, their pauses between two consecutive melodies were 7-sec long, and their shortest unfilled delays between study and test were 7 s, and 12 s and more if filled with intervening melodies), caution is needed when comparing our results. The most major difference is that in all of the Dowling studies, contour change versus pitch-interval-change-only are pitted against each other. When exposed to *both* types of lures, listeners showed an interesting trend of ever improving recognition sensitivity with same-contour lures the longer the delay (filled with other melodies) between exposure and test. In contrast, our main interest was to look specifically at the "resolution" of one's *exact* short-term memory representation of isochronous tone sequences as used by Taylor and Pembrook (1983). This is why we intentionally allowed for the use of pitch memory as it would occur in a natural setting. We made it impossible for participants to use any contour-based proxies to differentiate targets from lures, and instead, forced them to attend to exact pitch interval information, which they were indeed able to do beyond chance, despite being musically untrained. Byron and Stevens (2006) found similar sensitivity to same-contour pitch interval changes with non-isochronous 8-note sequences (derived from unknown folk tunes) and a silent 2-sec delay between exposure and test in a same/different task with musically untrained participants. Their participants could detect differences in pitch interval magnitude with same contours even in transposition hinting at some pitch interval magnitude retention independent of contour information[3]. Cuddy and Cohen (1976) found evidence for similar pitch interval extraction even for 3-tone sequences in a 2AFC task with a 2-sec unfilled retention interval.

The question of what makes a melody more discriminable is beautifully addressed in Harrison, Musil, and Müllensiefen (2016) who modeled responses from same/different tests looking at a large range of melodic (and rhythmic) features. In their model, melody complexity is defined on the basis of number of notes (the more notes, the higher the working memory load) as well as contour and pitch interval related aspects (also cf. Cuddy, Cohen, & Mewhort, 1981). Since our study controlled for a variety of melodic aspects we believe that an assessment of our corpus would render a rather stable complexity level caused by the fact that *all* melodies had to be discriminated against contour-wise *same* and pitch interval-wise *very similar* lures. Given that listeners received little help from lure dissimilarity, their recognition responses had to rely more on the quality of their initial encoding. In that sense, it is worth looking at which features in such an otherwise very homogenous set of targets and lures still help listeners in their discrimination performance. We plan to do an analysis on the best and worst discriminated melodies to see whether we can detect certain shared saliency increasing or decreasing features. If we cannot find such shared characteristics we might think that at least the ones with top discriminability might be tone sequences reminiscent of sections of well-known tunes which can be found out by asking our or a new group of participants to rate those top and worst tunes based on criteria such as "sounds well-known", "reminds me of tune X".

Another interesting and quite "single" study in the short-term recognition literature for melodies is Mikumo's (1992) where she looked at the effects of different types of intervening stimuli in a same-different recognition task. Among others, she found that with tonal (rather than atonal) 6-tone sequences, musically less trained participants performed significantly better when there was a silent 12-sec delay between the standard and comparison stimulus than when the delay was filled with another melody or with shadowing tasks. In her setup, where targets were untransposed and melodies were played at a slow rate (1 s per note), same-contour melodies triggered high false alarms when collapsed over interference conditions. Despite its discrepancies from ours, this study could have been an interesting first study to look at the differential effects of melodic versus nonsense syllabic interference in short-term recognition memory for isochronous melodies. However, since intervening nonsense syllables were shadowed whereas intervening tone sequences were not, we are unable to make clear inferences with respect to the vulnerability of a melody representation to verbal versus musical interference. Any detrimental effect in the nonsense syllable condition could have been due to the motor process of shadowing rather than nonsense syllables per se. We believe that our study provides a unique opportunity to look at the differential effects of intervening verbal (nonsense syllables) versus musical (diatonic or nondiatonic) stimuli on short-term memory performance in melody recognition. This, in turn, may allow us to compare the predictions of Baddeley's (2000) working memory model which proposes that the phonological loop is responsible for the processing of both verbal and musical material and Berz' (1995) proposal of a separate "slave system" for musical material only. Interestingly, when using a 2AFC test, our findings seem more in line with Baddeley's model (though we see a trend for a stepwise worsening when going from nonsense syllables to non-diatonic to diatonic intervening sequences). But when using a same/different test, our findings seem more in line with Berz' model. In a same/different test setup, we were also able to observe a general depression of the J-shaped serial position curve and a curbing of its recency section for the intervening *melodic* sequence groups only.

Last but not least, we found some differences in overall performance as well as in the effects of our variables on performance depending on whether participants were tested with a 2AFC or a same/different test. The same/different test led to a bias to say "same" which was likely driven by the high similarity between target and lures. Given that we used exactly

[3] But unlike our minute one-diatonic pitch interval changes at test, theirs pitted "step" changes against "leap" (>3 semitones) changes.

the same stimulus corpus, this shows that the particular way of testing directly affected the decision making process which, in turn, affected the results (cf. Jang, Wixted, Huber, 2009). Benassi-Werke et al. (2010) showed that when using a forward melodic span test, which is a relatively bias-free test due to being recall instead of recognition, musically untrained listeners had an average span of 5.9 notes for diatonic sequences with max. four semitone leaps, and 4.9 notes for those with unrestricted pitch intervals. This confirms that even in the more challenging same/different setup our listeners were dealing with a task that was fully within their short-term memory capacity for tunes. In other words, it is unlikely to think that their inferior performance in the same/different setup was due to capacity limits.

In conclusion, we believe that further work has to be done on the instantaneous resolution of short-term memory representations for melodies. Given the extensive amount of research on visual-spatial short-term and working memory representations (e. g., Luck & Vogel, 2013; Sligte, Vandenbrouke, Sholte, & Lamme, 2010), we think the musical domain has to follow suit. This will also provide us with a better understanding of how musical material is processed in immediate memory and which models better predict observed findings. Certainly, future research also has to look at short-term memory for melodies with rhythmic elements, such as the ones used by Byron & Stevens (2006; also cf. Byron, 2008).

References

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences, 4(11),* 417-423.

Benassi-Werke, M. E., Queiroz, M., Germano, N. G., & Oliveira, M. G. M. (2012). Melodic memory and its dependence on familiarity and difficulty. In S. Serafin (Ed.) *Proceedings of the 7th Sound and Music Computing Conference* (pp. 1-8). Spain: Universitat Pompeu Fabra.

Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception: An Interdisciplinary Journal, 12(3),* 353-364.

Byron, T. (2008). The processing of pitch and temporal information in relational memory for melodies. (Unpublished doctoral dissertation). University of Western Sidney, Penrith, New South Wales, Australia.

Byron, T., & Stevens, C. (2006). Steps and leaps in human memory for melodies: The effect of pitch interval magnitude in a melodic contour discrimination task. In M. Baroni, A. R. Addessi, R. Caterina, M. Costa (Eds.) *Proceedings of the 9th International Conference on Music Perception and Cognition (ICMPC9)*, Bologna, Italy.

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *The Quarterly Journal of Experimental Psychology, 28(2),* 255-270.

Cuddy, L. L., Cohen, A. J., & Mewhort, D. J. K. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 7(4),* 869.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science, 168(3939),* 1604-1605.

Dewitt, L. A., & Crowder, R. G. (1986). Recognition of novel melodies after brief delays. *Music Perception: An Interdisciplinary Journal, 3(3)*, 259-274.

Dowling, W. J. (1973). The perception of interleaved melodies. *Cognitive psychology, 5(3),* 322-337.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological review, 85(4),* 341.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics, 50(4),* 305-313.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology: A Journal of Research in Music Cognition, 1(1),* 30.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America, 49(2B)*, 524-531.

Dowling, W. J., Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception & Psychophysics, 57(2),* 136-149.

Dyson, M. C., & Watkins, A. J. (1984). A figural approach to the role of melodic contour in melody recognition. *Perception & psychophysics, 35(5)*, 477-488.

Harrison, P. M., Musil, J. J., & Müllensiefen, D. (2016). Modelling melodic discrimination tests: descriptive and explanatory approaches. *Journal of New Music Research, 45(3),* 265-280.

Jang, Y., Wixted, J. T., & Huber, D. E. (2009). Testing signal-detection models of yes/no and two-alternative forced-choice recognition memory. *Journal of Experimental Psychology: General, 138(2),* 291.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in cognitive sciences, 17(8),* 391-400.

Mikumo, M. (1992). Encoding strategies for tonal and atonal melodies. *Music Perception: An Interdisciplinary Journal, 10(1),* 73-81.

Ortmann, O. (1933). Some tonal determinants of melodic memory. *Journal of Educational Psychology, 24(6),* 454.

Pembrook, R. G. (1987). The effect of vocalization on melodic memory conservation. *Journal of Research in Music Education, 35(3),* 155-169.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception: An Interdisciplinary Journal, 29(3),* 255-267.

Sligte, I. G., Vandenbroucke, A. R., Scholte, H. S., & Lamme, V. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in psychology, 1*, 175.

Taylor, J. A., & Pembrook, R. G. (1983). Strategies in memory for short melodies: An extension of Otto Ortmann's 1933 study. Psychomusicology: *A Journal of Research in Music Cognition, 3(1),* 16-35.

Trehub, S. E., Unyk, A. M., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant behavior and development, 16(3),* 285-295.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Variable Critical Bands

Ludwig Kollenz[1]

[1] *Department of Musicology, University of Vienna, Austria*

[1]a0325926@univie.ac.at

Abstract

Critical bands were first discovered in the 1930s by Fletcher and Munson during their research on equal loudness contours and have been a research subject since. This technique of human hearing is connected to the perception of loudness as well as other effects. Zwicker measured the width of critical bands and derived the so-called bark-scale from this. This scale consists of 24 bands at fixed center frequencies, although Zwicker states that a critical band can appear at any center frequency in the human hearing range. Glasberg and Moore researched auditory filters and (assuming critical bands are related to auditory filters) came to the ERB-scale; similar to the bark scale but consisting of narrower bands.

Researchers agree that critical bands (as well as auditory filters) can appear at any center frequency and yet fixed scales (for example bark) are used.

In the presented work dynamic behaviour of critical bands shall be explored. Width and center frequency for different input signals are of interest. From this an enhanced loudness approximation can be derived. Other fields of application could be audio compression or hearing aids.

Introduction

When Fletcher and Munson researched equal loudness contours in 1933, they discovered that tones close to each other are evaluated differently in the ear than tones spread apart. This lead to the investigation of the so called critical bands. Fletcher and Munson gave an initial estimation of these critical bandwidths: 100Hz below 2kHz, 200Hz from 2 – 4kHz, 400Hz from 4 – 8kHz and 800Hz from 8 – 16kHz.

Zwicker, Flottorp and Stevens were able to measure “Frequenzgruppenbreiten” in 1957 and suggested that these were equal to the concept of critical bands. The authors state that Frequenzgruppen are 100Hz wide until a center frequency of 500Hz, above that the Frequenzgruppen are approximatley one third of the center frequency wide. It is noteworthy that the authors present 2 lists of stacked Frequenzgruppen.

The first entries in the lists for the center and cut-off frequencies were:

20	
	65
110	
	155
200	
	250

The left column gives the cut-off frequencies for the right column center frequencies and vice versa the right column gives the cut-off frequencies for the left column center frequencies.

In 1961 Zwicker references an I.S.O. meeting where preferred center frequencies were defined. He stresses that critical bands can occur at any frequency in the hearing range, but a fixed scale makes measurements comparable. The entries for preferred center frequencies started with 50, 150, 250, 350, 450, 570, 700, 840 and 1000Hz.

Center frequencies are again presented by Zwicker and Feldtkeller in 1967. First the authors also state that the center frequencies can occur at any frequency of the human hearing range. Then a division into 24 bands is presented (named the bark scale after Heinrich Barkhausen). This stacking of critical bands is called “willkürlich” (arbitrary) by the authors. They assume based on their measurements, that a broadband noise is divided into these 24 bands, although in one of their measurements the noise was divided into 30 bands. The center frequencies here are the same as in Zwicker 1961 (see above).

Spreng also states in 1975 that the center frequencies of critical bands are not fixed. Their width corresponds to 1.3mm on the basilar membrane and hence has a neural density of 1300 sensory cells. Due to the 1.3mm width a connection to the “Kopplungsbreiten” seems to be given. For loudness perception the complete evoked area in the inner ear is decisive. Two evoked areas can only be discriminated if there is a less evoked area between them, i. e. a local minimum is required. According to Spreng the inner ear is the origin of critical bands.

Moore and Glasberg suggested a new formula for critical bands in 1983. Since the inner ear can be modeled as a filterbank, ciritcal bands could be related to auditory filters. With this relationship to auditory filters the authors showed, that the width of critical bands should shrink below 500Hz instead of being constant at 100Hz. Based on empirical data the authors suggested equivalent rectangular bandwidth (ERB) scale, implicating the critical bands are related to equivalent rectangular bandwidth of an auditory filter.

The ERB-formula was revised in 1990. Moore and Glasberg found in their experiments that the upper slope of auditory filters is constant while the lower slope changes.

Schreiner and Langner researched the frequency organization in the inferior colliculus in 1997, which consists of a layered arrangement. The authors see in this structure a possible substrate for the shape of critical bands with lateral inhibition. This arrangement may be essential for the creation of critical bands as well as the extraction of information.

Aims

Critical bands shall be further explored, specifically their dynamic behaviour. The first phase is to explore their center frequencies if this is continuous or how small possible discrete steps are and if this possible stepwidth changes towards higher frequency regions. For applications it is of interest how dense the sampling of (possibly continuous) critical bands should be.

The second phase will be experiments on the width of critical bands. It is known, that their width changes with temporal effects. This shall be further explored.

Methods

The pre-study was to develop a simulation for a general approach to critical bands. Phase 1 and 2 are psychoacoustical experiments on critical bands, which shall reveal information which is implemented in the simulation, so it will be able to correctly predict dynamic behavior of critical bands.

Simulation

The simulation uses a wavelet-transform to analyze a time-domain signal in the frequency domain. The frequency resolution of this transform reflects the resolution of the human ear better than a short-time Fourier-transform. Patterson et al. (1987) presented the gammatone-filter, which reflects the filtering in the cochlea. Based on this Venkitaraman et al. developed a complex gammatone wavelet transform in 2014 extending the gammatone filter to a complex gammatone wavelet.

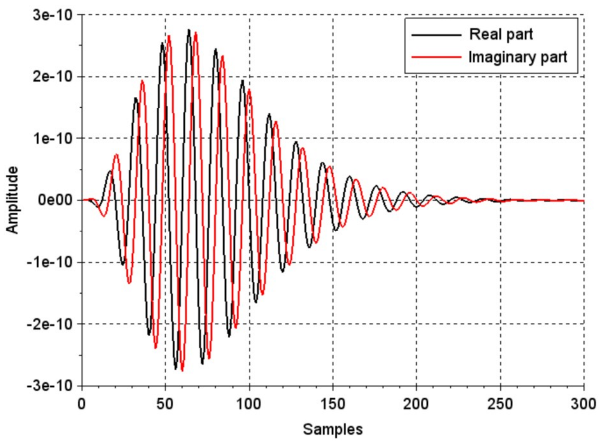


Figure 1. A complex gammatone wavelet.

Since a discrete transform (applied to digital signals) can only evaluate at finite, discrete frequencies the transform has to have a high resolution, in order to approximately allow dynamic critical bands. In the current implementation the wavelet transform quarters the bark-scale.

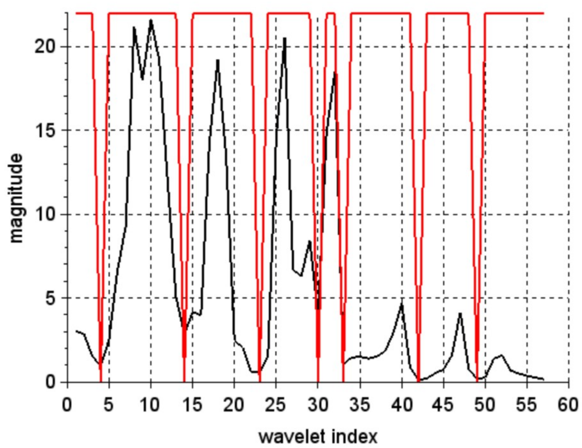


Figure 2. Spectrum (black) and estimated critical bands (red).

The magnitude frequency response is analyzed with a peak-valley-finding algorithm to discriminate different areas. If these areas are considered critical bands, then the intensity per band can be evaluated and hence overall loudness.

Psychoacoustic experiments

Psychoacoustic experiments are conducted as listening tests. Utilizing the notched noise method to prevent off-frequency listening, arbitrary center frequencies for critical bands shall be explored. Three center frequencies have been chosen as samples of the human hearing range: 250, 1000 & 8000Hz. For each frequency the ERB is calculated and dissected into 10 segments. These new frequencies are used as center for critical bands.

For each of these “fine” frequencies the ERB is calculated and the notch tested with a probe to verify that the critical bands indeed build up at these center frequencies.

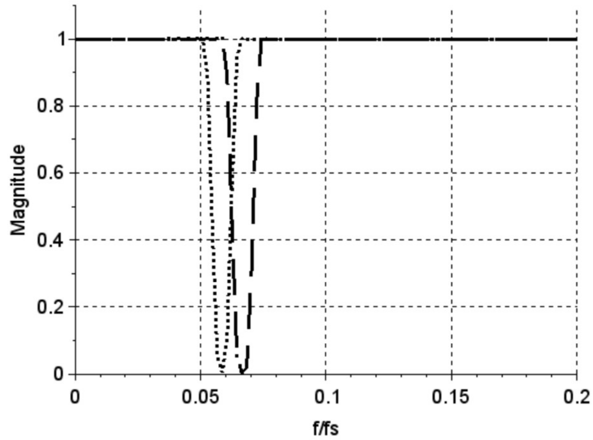


Figure 3. Two notch filters at 933Hz and 1066Hz. These represent the lower and upper limit for the ERB at 1kHz. This range is divided into 10 frequencies which are the center for notch filters each of according ERB width.

Results

The simulation suggests that the main differences for loudness emerge from the intensity per band, which determine masking of the critical band and the specific loudness. Continuous center frequencies with fixed bandwidths do not show significant differences to a filterbank with fixed center frequencies (and bandwidths). For example if a sine is exactly at the crossover region of two filters; both will have a lower power. But if there is a matched (concerning the center frequency) filter, it will have slightly more power. The differences arise rather from flexible bandwidths.

The psychoacoustic experiments are still ongoing. Drawing conclusions would be too early. The results will be published as soon as the experiments have been completed and the data evaluated. It has to be pointed out that the simulation is solely based on theoretical assumptions. Since it defines critical bands from gap to gap it will produce very wide critical bands for certain input signals, which does not seem to be realistic. The empirical data gained by the experiments will be used to trim the simulation, such that it is able to predict expected results.

Conclusion

Critical bands have been a field of research since the 1930s. The width of critical bands and the shape of auditory filters have been explored by others in detail. The dynamic behavior of critical bands has been pointed out by several researches, for example by Reuter (1995). This work focuses on the continuous center frequencies of critical bands as well as their width depending on temporal effects. Currently the work is in phase 1, which includes a simulation and experiments concerning the center frequencies. The second phase will focus on the width of critical bands. The empirical data gained will be reflected by the simulation. This shall provide a basis for possible future applications like loudness measurement, audio coding and enhancement of audio signals for hearing impaired.

References

Fletcher, H. & Munson, W.A. (1933). Loudness, Its Definition, Measurement and Calculation. *Journal of the Acoustical Society of America*, 82 – 108.

Glasberg B.R. & Moore, B.C.J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research 47*, 103-138.

Moore, B.C.J. & Glasberg, B.R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America,* 750-753.

Patterson, R., Nimmo-Smith, I., Holdsworth, J., Rice, P. (1987). An Efficient Auditory Filterbank Based on the Gammatone Function. in *SVOS Final Report, APU Report 2341 annex A*.

Reuter, Ch. (1995). *Der Einschwingvorgang nichtperkussiver Musikinstrumente.* Frankfurt: Lang.

Schreiner, C. & Langner, G. (1997). Laminar fine structure of frequency organization in auditory midbrain. *Nature 388*, 383-386.

Spreng, M. (1975). Mechanisch-elektrische Wandlung: Dynamik der Sinneszellen des Gehörs, Adaption und selektives Verhalten. In Keidel, W.D. (Ed.), *Physiologie des Gehörs* (pp. 102-152). Stuttgart: Thieme.

Venkitaraman, A., Adiga, A. & Seelamantula, C.S. (2014). Auditory-motivated Gammatone wavelet transform. *Signal Processing 94*, 608–619.

Zwicker, E., Flottrop, G. & Stevens, S.S. (1957). Critical Band Width in Loudness Summation. *Journal of the Acoustical Society of America*, 548-557.

Zwicker, E. (1961). Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen). *Journal of the Acoustical Society of America,* 248.

Zwicker, E. & Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger.* Stuttgart: S. Hirzel.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Perception of Stable Tones in Polytonal Structures

Lynnsey Lambrecht,[1] Leigh VanHandel,[2]

[1,2] *College of Music, Michigan State University, United States of America*

[1]lynnsey.lambrecht@gmail.com, [2]lvh@msu.edu

Abstract

Textbooks on twentieth-century techniques place a differing emphasis on the aural distinguishability of two tonal centers in polychords. This raises the question of polytonality being a compositional techniques or something that is perceptible to listeners. The goal of this research is to examine polychords drawn from *Corcovado* from Milhaud's *Saudades do Brasil* Op. 6 to determine if listeners can perceive two simultaneous tonal centers and if the composer's orchestration and registral placement of the individual chords has an effect on listeners' perception. In the first study, participants were presented with three polychords and their transformations which included spacing, voicing, register, and inversion as stimuli. After listening to the chord, participants sang or hummed the most stable tone in the structure and used a piano to determine the name of the pitch. For the second study, two polychords and their transformations were used in conjunction with probe tones. The subjects were asked to evaluate how well the probe tone fit with the polychord on a Likert scale of 1–7. Results from the first study showed that participants selected the tone present in both triads of the polychord as the most stable. The first study demonstrated that extreme registral placement of one triad in the polychord resulted in a preference for the other triad. The second study confirmed the preference of doubled tones, showed a significant difference in the treatment of doubled tones compared to other tones, and revealed a significant difference in the treatment of non-chord tones compared to chord tones. The results suggest that listeners favor the C4-C5 register in polychords to determine the goodness of fit for probe-tones. Overall, participants found a single, prominent tone in each polychord, which was impacted by spacing, register, and doubled tones. This indicates that the compositional choice of utilizing polychords impacts participants' perception of the music.

Introduction

Post-tonal theory courses allocate time to the study and analysis of polytonality and polychords. Leading textbooks differ in the definition of this technique, including the idea that the two sonorities that comprise the polychord must be aurally distinguishable. How do listeners perceive the tonal center of vertical sonorities constructed using polytonality? Are polychords aurally perceptible to listeners or just a technique employed by composers? Does the use of polytonality create organized harmonies or does it obscure the perception of tonality for listeners?

Background

Krumhansel and Schmuckler (1986) found that probe tone ratings from polytonal structures in Stravinsky's Petroushka fit Van den Toorn's (1983) hierarchy of priorities better than the major key profiles. Thompson and Mor (1991) found that listeners were sensitive to two key centers and that when one key predominates in a polytonal context, other keys may not contribute to the overall tonal structure. These studies demonstrate that post-tonal music is perceived to have different tonal hierarchies than music from the common-practice period, and they raise the question of how other polytonal combinations are perceived by listeners.

Experiment 1

To investigate the perception of stable tones in polychords, sonorities were drawn from Milhaud's *Saudades do Brasil Op. 6, VII. Corcovado*. Participants were asked to listen to the vertical sonorities and determine the most stable tone.

Methodology

Using a Qualtrics online survey, participants listened to 6 second audio clips of 46 randomly presented polychords. The participants were then asked to sing or hum the most stable pitch in the structure. Using a keyboard, participants found the note name of the most stable pitch. Participants could repeat audio samples as desired. Each question was separated by 6 seconds of white noise to eliminate any influence of the different questions. Seven control questions contained only C major triads to ensure that participants were able to perform the task of finding stable pitches.

Stimuli

Figure 1 shows the three different polychords drawn from different points in Milhaud's *Corcovado.* Chord 1 originated from m. 1, Chord 2 came from m. 23, and Chord 3 was drawn from m. 53. The audio examples were realized using a MIDI piano timbre.

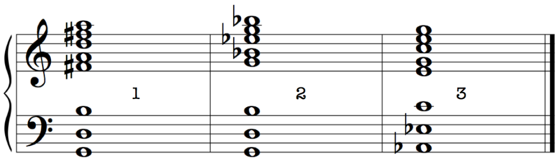


Figure 1. Three polychords drawn from Milhaud's *Corcovado.*

These chords were presented in Milhaud's original voicing (O) plus 12 transformations, including close position (C), hands in root position (R), hands inverted (I), high transposition (H), low transposition (L), as well as combinations of these transformations including close position inverted (CI), root position inverted (RI), close position high (CH), root position high (RH), close position inverted high (CIH), root position inverted high (RIH), close position low (CL), root position low (RL), close position inverted low (CIL), and root position inverted low (RIL). Figure 2 shows all of the transformations used in the study of Chord 1.

The high and low transformations were included to investigate what happens when the chords are voiced with members outside of the preferred listening range. Huron (2001) outlined the optimal region of pitch stability between F_2 and G_5 based on virtual pitch weight.

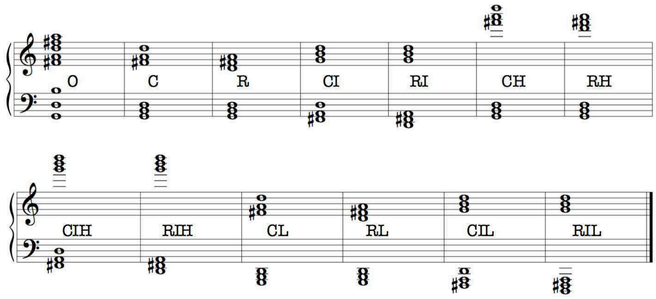


Figure 2. Transpositions of Chord 1 used as stimuli, including original voicing (O), close position (C), root position (R), high transposition (H), low transposition (L), and combinations formed by these transpositions.

Participants

23 participants (n=12 F, *M*=25.5 years old) participated in this study. All had studied undergraduate music theory, and 15 participants had studied graduate music theory. Two subjects self-reported absolute pitch. The participants had an average of 14.5 years of music lessons.

Results

The results for the control questions that only contained C Major triads demonstrated that the participants were capable of performing the task of selecting a stable tone; 94% of responses (132 of 140) selected C as the most stable tone. Results for Chord 1, shown in Figure 3, indicate that most participants selected stable tones that were members of the polychord and gave preference to D, which was the tone that was present in both triads (G Major and D Major). D was selected as the most stable tone for 42% (134/321) of the Chord 1 stimuli, with A selected 24% (78/321) of the time. G was selected as the most stable tone for 15% (50/321) of the Chord 1 stimuli.

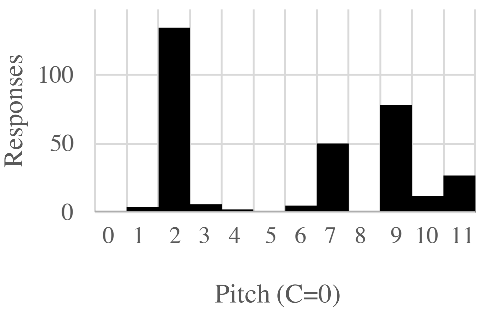


Figure 3. Stable tone pitch responses for all Chord 1 questions.

Figure 4 includes data from Chord 1 that shows when one triad of the polychord is placed outside of the preferred listening range, participants selected chord members of the triad that remained in the preferred listening range. In this case, the G Major triad was in the bass and the D Major triad was transposed up two octaves from its original placement. The participants selected D, G, and B as the most stable tones of the polychord, which are all members of the G Major triad.

The results of Chord 3 were similar to those of Source Chord 1, most likely because of the two chord's similar voicing and ranges. However, Chord 2's overall responses, shown in Figure 5, did not indicate a significant preference for the doubled chord tone. Chord 2 was built from a G Major triad and an E^b Major triad. The doubled tone, G, was selected as the most stable tone for only 23% of the Chord 2 stimuli, while E^b was selected as the most stable tone for 26% of the Chord 2 stimuli.

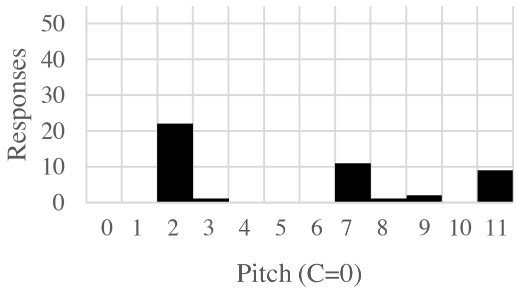


Figure 4. Stable tone pitch responses for Chord 1 in non-inverted and high transposition.

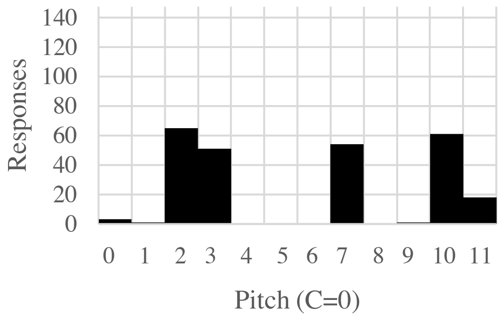


Figure 5. Stable tone pitch responses for all Chord 2 questions.

Discussion

Listeners gave preference in their responses to the Chord 1 questions to the common tone shared between both triads, and selected notes played by the left hand as the most stable tones. When the left-hand chord was moved outside of the optimal hearing range, the task became much more challenging and yielded mixed responses of the most stable tone. The pitch responses from Chord 3 align with those found with Chord 1, but when the left-hand triad was transposed too low for stability, listeners selected tones that belonged to the right-hand triad as the most stable pitches. The pitch responses from Chord 2 did not align with the results from the previous two chords.

The difference in the perception of the chords is significant to the analysis of *Corcovado*. Chord 1 is from m.1 at the beginning of the piece, and Chord 3 from m. 53 leads to the return of the beginning material; therefore, the chords serve a similar function within the piece. However, Chord 2 is drawn from m. 23 during a transition to a new section in the middle of the piece. The unique triad combinations of each source chord help to differentiate the sound worlds of the formal sections of *Corcovado*.

Remaining questions following experiment 1 include: Did humming/singing the pitch give preference to a tone in the subjects' vocal range? Did unaltered loudness of stimuli create bias? Did overtones produced by MIDI stimuli influence a more stable tone?

Experiment 2

To investigate the perception of polytonality further, a second experiment was conducted. In this experiment, participants were asked to evaluate how well specific tones fit with these same structures from Milhaud's *Corcovado*.

Methodology

Experiment 2 consisted of 40 audio examples of polychords delivered by a Qualtrics survey that included Chord 1 and Chord 2 from Experiment 1. Since Chord 1 and Chord 3 had similar results in the prior experiment, it was possible to use just Chord 1 in this experiment. The questions were presented randomly and asked subjects to evaluate how well a probe tone fit with a polychord using a Likert scale of 1 (does not fit) to 7 (fits very well). Participants were able to repeat each audio example as desired.

Stimuli

The audio examples consisted of a 3 second polychord, 1 second of silence, and 3 seconds of a Shepard tone probe-tone. The audio examples were pure sine waves produced by a Max/MSP patch, and all chord members were adjusted to the equal loudness contour. The experiment used five transformations of both polychords, including the original presentation drawn from Milhaud's composition (O), close position without any doubled pitches in each hand (C), root position without any doubled pitches in each hand (R), inverted hands of the close position (CI), and inverted hands of the root position (RI). The transformations of Chord 1 are shown in Figure 6.

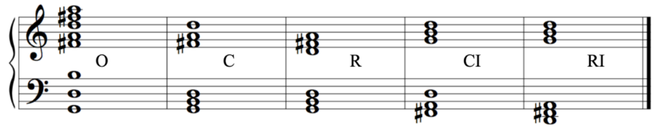


Figure 6. Chord 1 transformations used in experiment 2 including original voicing (O), close position (C), root position (R), and the combination of the transformations.

The probe tones consisted of the tones of each polychord; the tones G-B-D were taken from the G Major triad, D-F#-A from D Major, and Eb-G-Bb from the Eb Major triad from Chord 2. All probe tones were heard for each chord, so some probe tones were non-chord tones. For example, participants had to judge how well Eb fit within the context of the G Major and D Major context of Chord 1, and how well F# fit within the context of the G Major and Eb Major context of Chord 2.

Participants

24 subjects (n=16 female, *M*=31.2 years of age) participated in experiment 2. All but one of the participants (23) had studied undergraduate music theory, and 15 had studied graduate music theory. The participants had an average of 14 years of music lessons. Of these participants, 14 subjects completed both experiment 1 and experiment 2.

Results

In experiment 2, participants rated the doubled tone present in each polychord as a better fit in both Chord 1 and Chord 2. Figure 7 shows the average ratings of each probe-tone separated by the two chords. These averages show that subjects treated chord members and non-chord members similarly across the two polychord sources; the non-chord tones Bb and Eb were rated lower for chord 1 and the non-chord tones of A and F# were rated lower for chord 2.

Figure 8 compares the average ratings for individual probe tones between Chord 1 and Chord 2 in experiment 2. This demonstrates that the subjects created discrete categories with their average probe tone ratings. For both chords, the non-chord tones were rated lower than all chord tone members, and the doubled tone was rated higher than all of the other chord tone members ($F(4,60) = 15.6$, $p < 0.5$).

Participants gave different average ratings to probe tones depending which hand the tone occurred in. Figure 9 shows the average probe tone ratings for both source chords separated by the tones present in each hand. The average rating was highest for the tone present in both hands and both triads of the polychord, followed by the tones present in the right hand only, followed by the tones present in the left hand only, and the lowest rating went to probe tones not present in either hand of the chord. Each pairwise probability is significant at $F(3,66)=28.6$, $p \leq .02$.

Figure 10 includes the average ratings for probe tones separated into the right and left hands for both source chords, showing that the participants gave preference to probe tones that were present in the right hand triad ($F(1,34) = 9.812$, $p = .003$).

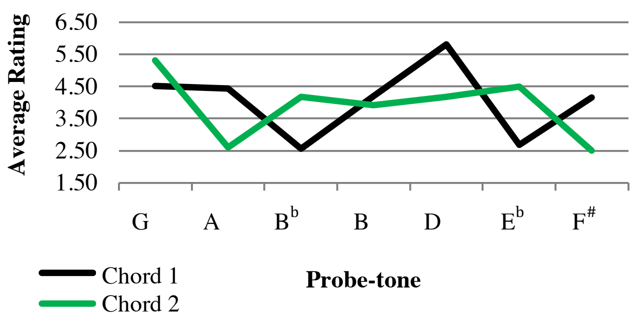


Figure 7. Probe tone rating of probe tones by chord.

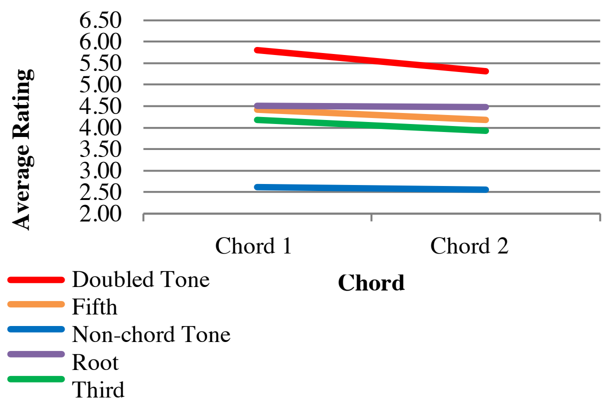


Figure 8. Average probe tone rating of each chord member by source chord.

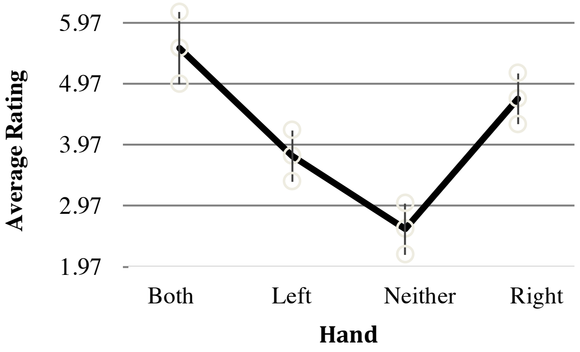


Figure 9. Average probe tone rating of members of each hand of both source chords.

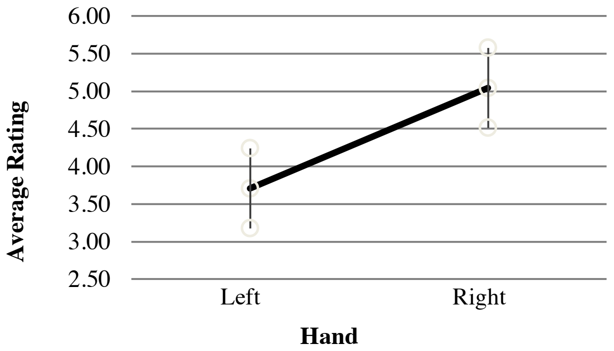


Figure 10. Average ratings of left hand chord members versus right hand chord members of both source chords.

Discussion

Experiment 2 shows that subjects gave preference to the doubled tone in each polychord. These results confirm the findings of experiment 1, in which participants also gave preference to the doubled tone. Subjects also treated non-chord tones differently than members of the polychords when determining a goodness of fit for the probe tones. Non-chord tones were given a significantly lower average rating than chord tones.

Experiment 2 results also include a right hand chord tone preference, which indicates that participants placed an emphasis on the C4–C5 register when determining goodness of fit for the probe tones. Further studies may determine if this can be attributed to subjects' primary instruments or voice types. More females than males participated in experiment 2, while in experiment 1 the ratio of females to males was almost equal. This difference might contribute to the registral emphasis on the C4–C5 register found in experiment 2.

Conclusion

Both experiment 1 and experiment 2 show a significant difference in the treatment of doubled tones to other tones in the polychords. Doubled tones, which were found in both triads in the polychords, received higher averages in both experiments, indicating that participants considered these tones to be more stable than non-doubled tones. Experiment 1 and experiment 2 also demonstrate that non-chord tones were treated differently than chord members of the polychord sonorities. The non-chord tones received significantly lower average ratings in both experiments. These findings reinforce the idea that the compositional choice of writing music with polytonality and polychords does impact the listener's perception of tonality and stability in a piece.

This further supports the analytical observation that Milhaud was making a purposeful compositional choice with these three polychord combinations in *Corcovado*. Chord 1 is drawn from the beginning of the piece, establishing the palette of sound for the movement. Chord 3 was treated similarly by subjects in the study; this chord was drawn from just before the return of the original material. It seems evident that Milhaud was purposefully re-establishing the same soundscape as the beginning for the conclusion of the composition. The polychord combination for Chord 2 is drawn from m. 23, and it signals a digression from the original soundscape established at the onset of the piece. The new section in the composition is further accentuated by changes in the music including the accompaniment pattern, rhythm, texture, and dynamics.

To further investigate the questions raised through these two experiments that investigated polychords, two additional experiments might create a deeper understanding of listeners' perception of polytonality and polychords. One experiment can repeat the probe tone study, similar to experiment 2, with additional stimuli. In this experiment, both triads of the polychord will be in a higher register than the original presentation, from C4 to C6. This study can further illuminate if a registral emphasis was found in experiment 2.

An additional experiment can build on these two preliminary experiments by having subjects evaluate and determine the most stable tone perceived in extended polytonal passages that are drawn directly from polytonal repertoire. This study will investigate the role that context plays in determining the most stable tone when composers use polytonality in their composition. This experiment would enable the acoustic spacing of original presentations of the polytonal passages to be preserved in the stimuli.

Acknowledgements. We thank Michigan State University for the support of this research.

References

Cook, N. D., Fujisawa, T. X., & Hiroo Konaka, H. (2007). Why not study polytonal psychophysics? Empirical Musicology Review 2, no.1, 38-44.

Hamamoto, M., Botelho, M., & Munger, M. P. (2010) Non-musicians' and musicians' perception of bitonality. Psychology of Music 38, no. 4, 423-445.

Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading for perceptual principles. Music Perception: An Interdisciplinary Journal 19, no. 1, 1-64.

Kostka, S. (1999). Materials and techniques of twentieth-century music. Upper Saddle River, New Jersey: Prentice-Hall.

Krumhansl, C. L. (1983). Perceptual structures for tonal music. Music Perception: An Interdisciplinary Journal 1, no. 10, 28-62.

Krumhansl, C. L. (1990) Cognitive foundations of musical pitch. New York: Oxford University Press.

Krumhansl, C. L., & Schmuckler, M. A. (1986). The Petroushka chord: A perceptual investigation. Music Perception: An Interdisciplinary Journal 4, no. 2, 153-184.

Meredith, D. (2003). The perception and cognition of pitch structure in tonal music. (master's thesis). City University, London, England.

Milhaud, D. (1960). Saudades do Brazil. Paris: Max Eschig.

Roig-Francolí, M. (2008). Understanding post-tonal music. New York: McGraw-Hill.

Straus, Joseph. Introduction to Post-Tonal Theory. New York: Norton, 2016.

Thompson, W. F., & Mor, S. (1992). A perceptual investigation of polytonality. Psychological Research 54, no. 2, 60-71.

Computational model of pitch detection, perceptive foundations, and application to Norwegian fiddle music

Olivier Lartillot,[1] Hans-Hinrich Thedens,[2] and Alexander Refsum Jensenius[3]

[1,3] *RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion, University of Oslo, Norway*

[2] *National Library, Norway*

[1]olivier.lartillot@imv.uio.no

Abstract

Automated detection of pitch in polyphonic music remains a difficult challenge. Implementation of perceptive/cognitive models have been so far less successful than engineering methods.

We present a model that is neither based on a machine-learning training on a given set of samples, nor explicitly relying on stylistic rules. Instead, the methodology consists in conceiving a set of rules as simple and general as possible while offering satisfying results for the chosen corpus of music. We present a new method for harmonic summation that penalizes harmonic series that are sparse, in particular when odd partials are absent, as it would indicate that the actual harmonic series is a multiple of the given pitch candidate. Besides, a multiple of a fundamental can be selected as pitch in addition to the fundamental itself if its attack phase is sufficiently distinctive. For that purpose, we introduce a concept of pitch percept that persists over the whole extent of the note, and that serves as a reference for the detection of higher pitches at harmonic intervals.

The proposed method enables to obtain transcriptions of relatively good quality, with a low ratio of false positives and false negatives. The construction of the model is under refinement. We are applying this method to the analysis of recordings of Norwegian folk music, containing a large part of Hardanger fiddle pieces and a cappella singing.

By attempting to design computer models based on general rules as simple as possible rather than on machine learning, while resulting in a behavior in terms of pitch detection that comes closer to human capabilities, we hypothesize that the underlying mechanisms thus modelled might suggest general computational capabilities that could be found in cognitive models as well. In the same time, an improvement of the model based on expertise in music perception and cognition is desired.

Cognitive models of pitch perception

Cognitive modelling of pitch perception remains an open and controversial issue (McDermott & Oxenham, 2008). The perception of the fundamental frequency (or F0) of a given pitch is explained using various competing–and complementing–mechanisms. The cochlea operates a spatial (or tonotopic) decomposition of sound along frequencies at each successive instant. But at the same time a precise pitch estimation requires the study of the time waveform for each individual critical frequency in that tonotopic decomposition. Because each critical frequency is actually the center of a frequency band–with increasing bandwidth as frequency increases–the first harmonics of a given pitch will be properly resolved while higher harmonics will interact with each other in frequency bands and will not be resolved. Those mechanisms have not been fully understood yet (Oxenham, 2012). Another mechanism possibly helping pitch perception relates to the phase locking of individual auditory nerves. There is evidence for cortical neurons beyond primary auditory cortex that are tuned to pitch (Bendor & Wang, 2006; McDermott & Oxenham, 2008).

Extensive studies have attempted to understand how multiple pitches could be perceptually combined to form chords and voices. The knowledge brought by these works is not sufficiently detailed to be directly translated into computational models.

In fact, the mere core mechanism of single note perception does not seem to be sufficiently understood in a cognitive point of view to allow computational modelling. For the simple case of individual sound event, this could be conceived as a simple detection based on detection of attack, onset and decay in the temporal representation of the energy of the signal. But when considering complex polyphonies, the notes detection seems to depend mainly on pitch perception, probably through a tracking of F0 over time and a detection of attack and decay along each F0 separately.

Computational models of pitch extraction

Computational approaches for pitch extraction can be decomposed into several types.

Machine learning approaches

The most dominant type of approach currently giving the best results is based on machine learning: the learning system is trained on particular audio recordings for which the corresponding transcriptions are given as well. "Learning", here, means that the program automatically optimizes the inference algorithm so that it predicts the correct transcription when given as input its corresponding audio recording. This approach is dependent on the transcriptions provided during the learning phase. The resulting algorithm does not generalize well on audio examples that have very different musical characteristics. If we consider for instance the transcription of fiddle music, a proper training would require to provide as learning examples detailed transcriptions. And if we aim to get detailed transcriptions showing the pitch fluctuation within each note, we would have to provide such example transcriptions beforehand.

Nowadays machine learning is implemented using neural networks and notably using deep learning techniques. Despite the name, "neural networks" are not supposed here to provide an actual cognitive modelling of pitch perception, since humans generally do not learn to perceive pitch based on supervised training.

Template-based models

Another method consists in recording individual notes played along the whole range of pitches under consideration and for various types of instruments. Those notes are then retrieved automatically on the audio recording to be analyzed, through a mathematical decomposition. Evidently, this engineering approach is not supposed to mimic human cognition. One major practical limitation is that the approach will not work properly for instruments with very different timbre that those prerecorded or with larger pitch range.

Cognitive computational models

The aforementioned cognitive models of pitch perception have been translated into detailed computational models (Medis & Hewitt, 1991). It has also been shown that some simplification can be carried out without degrading the general quality of the model (at least in the particular domain of application under consideration), while allowing to solve complex problems that the more detailed models were not able to tackle yet. The particular step of periodicity analysis of the critical bands, was not much explained in music cognition. In many reference works, this analysis has been modeled using autocorrelation function. But alternative models such as comb filter have been proposed as well. One of the most advanced computational models for pitch extraction based on cognitive theories has been proposed by Klapuri (2006a). Because pitch perception models so far focus mainly on single F0 extraction, additional engineering-based methods are developed to enable multipitch extraction, mainly based on time-domain cancellation.

Other approaches

Klapuri (2006b) has proposed another model, which offers significant improvements and has become a reference in Music Information Retrieval (MIR). Interestingly enough, this model does not follow cognitive theories as closely as before (such as tonotopic decomposition followed by autocorrelation function, as discussed above) but instead develop engineering-based strategies. This seems to indicate that current cognitive understanding of pitch perception is not mature enough to be directly implemented into a state of the art computational model. In many recent approaches, periodicity analysis is based on frequency *spectrum* representation, computed for instance using the *Fourier Transform*. This is the case in the subsequent work by Klapuri (2006b). Moreover, this approach is funded on the concept of *harmonic summation*, that we will discuss further in the next section.

Most current MIR approaches for pitch detection, including Klapuri's, search for F0s for each successive instant (or time frame more precisely) independently. The F0s detected frame by frame are then tracked over time in a second step, in order to form pitch contour based on time and frequency continuity, using heuristics based on auditory streaming cues or additional musical knowledge.

Still an open problem

Despite the significant advance in multipitch estimation, this task remains one of the main challenges in the MIR field, which needs to address many difficulties, such as masking, overlapping tones, mixture of harmonic and non-harmonic sources and the fact that the number of sources might be unknown (Schedl, Gómez & Urbano, 2014). Even on simple polyphonies, the performance obtained by multipitch estimation methods reaches moderate note accuracy for relatively simple music material, such as quartet, woodwind quintet recordings, and rendered MIDI, with a maximum polyphony of 5 notes (MIREX, 2016).

One way to compensate the limitations of current approaches in multi-pitch extraction is through the addition of music language models, which represent sequences of notes and other music cues based on knowledge from music theory or from constraints automatically derived from symbolic music data. Such approach would however not generalize well to various kinds of music genres and cultures, unless the models are updated accordingly. Besides, in a cognitive point of view, if listeners are able to detect the pitches without necessarily knowing that particular genre of music, we would surmise that a computational model should work independently on the stylistic rules.

Proposed model

Spectrum representation

Similar to many MIR methods, we first represent the audio signal in the frequency domain (or spectrum) using the Fourier Transform. By decomposing the audio signal into short parts (*frames*) and computing the spectrum representation for each frame successive, we obtain a bi-dimensional diagram, where the horizontal axis is the temporal evolution of the audio signal, and the vertical axis the different frequencies found in each frame. An example is shown in Figure 1. This spatial representation is similar (in terms of axes dimensions) to the representation of the pitch curves shown in Figure 2.

This spectral representation can be somewhat related to the basic principles of tonotopic decomposition, that is, decomposing the energy into frequencies. Another perceptive aspect that needs to be included in the spectrum representation is the frequency filtering operated by the outer ear, with an emphasis on frequencies around 3000 Hz (Terhardt, 1979). Figure 1 shows the result of this filtering, with most of the energy near 3000 Hz. It turns out that this frequency emphasis significantly improved the pitch detection performed in the subsequent steps, for instance when analyzing fiddle music (cf. next section). This illustrates the fact that the acoustic of fiddle instrument is tuned to human ears, for instance with respect to pitch clarity.

Harmonic summation

When representing a given pitch in the frequency domain, there is a peak of the frequency F0 of the fundamental and peaks as well at harmonic partials, whose frequencies are multiples of F0. A convenient method to detect F0s from the frequency domain is called *harmonic summation*: it consists in associating a score to each F0 by summing together the spectrum amplitudes at multiples of F0. For instance, for the frequency F0 = 440 Hz, we sum the spectrum amplitudes related to the frequencies 440, 880, 1320 Hz and so on. The highest scores would correspond to the F0s of the actual pitches in the signal.

In previous works, the score associated to each F0 candidate is computed through a simple summation of each partial magnitude. Klapuri (2006b) also follows this strategy, although using weighted summation.

In our view, a simple summation, even weighted, does not sufficiently describe the way harmonic sequences of partials are perceived. We have developed a new strategy, where instead of considering each partial p_i individually, we consider each partial in relation with its previous partial p_{i-1} or its two previous partials p_{i-1} and p_{i-2}. If p_i is an even harmonic, its contribution to the total score is computed by multiplying the spectrum magnitude at p_i and p_{i-1}. If p_i is an odd harmonic, its contribution is computed by take the maximum between (1) multiplying the spectrum magnitude at p_i and p_{i-1} and (2) multiplying the spectrum magnitude at p_i and p_{i-2}. For instance, a harmonic sequence where all odd harmonics of F0 are absent will give a score of 0, indicating that the actual pitch should rather be at 2*F0. This method penalizes harmonic series that are sparse, in particular when odd partials are absent, as it would indicate that the actual harmonic series is a multiple of the given pitch candidate.

Pitch detection based on dynamic evolution of partials

One major difficulty when detecting multiple pitches is that, mathematically speaking, any harmonic Fi of a given F0 present in the signal could itself appear as a possible F0: its own harmonic series is a subset of the harmonic series of the lower F0. For instance, the pitch F1 at an octave above a given pitch F0 has an harmonic series that corresponds to the even partials of the harmonic series of F0. Evidently, all harmonics Fi of a given F0 are not themselves perceived as individual pitches.

We may hypothesize that for a given harmonic Fi to be perceived as the fundamental of an additional pitch, superposed to the pitch at F0, its harmonic series being added to the harmonic series of F0, in the resulting harmonic series, the peaks corresponding to the series related to Fi should be higher. Klapuri (2006b) uses this hypothesis, and develops a method where each time a given F0 is found, its harmonic series is removed from the signal so that eventual harmonic series corresponding to other pitch at Fi can be detected as well. We also follow this hypothesis in our model, but instead of removing harmonic series for each pitch found in the signal, we search for predominant subseries in each given harmonic series.

In our experiments, we came to the conclusion that this hypothesis is too strong and does not explain all conditions for the appearance of pitches at harmonic intervals above other pitches. We have found another more subtle characterization: a pitch Fi can appear at such harmonic interval above F0 if the harmonic series starting from Fi increases over time. So even if the harmonic series of Fi is not particularly dominant compared to the series starting from F0, if there is a significant increase over time of some of its components, this suggests the detection of pitch Fi.

Adopting this strategy requires to rethink the way pitches are detected. As aforementioned, in most approaches in MIR, pitches are detected for each successive frame separately; they are then combined into notes by tracking the F0s values over time in order to form pitch contour based on time and frequency continuity. The decision concerning the selection of F0s is made in the first step, for each successive frame separately.

In our proposed model, F0s are still searched for in each successive frame, but once a F0 has been detected, a new pitch contour is created and further tracked in the subsequence frames. The pitch contour stores the dynamic evolution of the magnitude of each partial. This information can therefore been used to detect any pitch Fi at harmonic interval.

The proposed model will be presented in more details in an upcoming paper and will be integrated into MIRtoolbox (Lartillot & Toiviainen, 2007) and the MiningSuite.

Analysis of Hardanger fiddle music

This approach has been specifically developed for the analysis of traditional Norwegian folk music played on the Hardanger fiddle. The Hardanger fiddle is slightly smaller than a regular violin, with a shorter neck and a flatter bridge, which allows to play more than two strings at the same time. In addition to the bowed strings, there are four or five sympathetic strings that run under the board. The sympathetic strings resonate when the bowed strings are played, contributing to the rich sound of this fiddle and also giving the music its characteristic drone. The fiddle playing is also extensively ornamented (Haugen, 2016).

In the current state of this research, the algorithm has been tested on traditional Norwegian folk dances. Figure 2 shows the beginning of the analysis of a performance of the tune called *Gibøens bruremarsj*.

The first results of this new model are promising. We can generally detect all the notes, even if they are repeated very quickly and played on several strings at the same time. We can also get a detailed dynamic envelope of each note, that could be used for further research about rhythm and attacks. The algorithm also output false positives and false negatives sometimes.

Discussion

As we have seen in the study of the state of the art in computational models for pitch extraction, the implementation of perceptual models describing as closely as possible the way pitch is perceived has not lead so far to particularly successful solutions for polyphonic music transcription. It is telling in this respect that the author of the most detailed computational implementation of psychological modelling of multi-pitch extraction has turned later to a more engineering approach. His engineering approach proved more successful, as it became a reference model in MIR. It seems that the cochlear model and in particular the filterbank decomposition proves somewhat problematic, as it leads to a significant distortion of the audio signal without clear advantages, compared to more simple engineering approaches, such as the use of Fourier transforms. We hope new advances in cognitive understanding (or works already published unknown to us) will offer new guidance for computational improvements.

Nonetheless the strategies developed in computational approaches might suggest hypotheses concerning the cognitive modelling of pitch perception. In particular if a simple computational mechanism is shown to offer particularly good results in a large range of music styles, we may suppose that similar operations are performed in the auditory system.

In that respect, we might wonder whether (or not) the concept of harmonic summation might have some cognitive validity. It seems computationally more simple and effective than selecting individual frequencies in the spectrum. One main objection would be that harmonic summation would not work in the same way for inharmonic sounds. However inharmonic sounds are usually associated with attacked sound, whose pitch might be detected using complementary mechanisms taking benefit of the attack phase.

One particularity of our proposed model is that it introduces a concept of pitch percept that persists over the whole extent of the note, and that serves as a reference for the detection of higher pitches at harmonic intervals. The perceptual and cognitive implications of this notion of pitch percept need to be investigated.

Acknowledgements. This work was carried out in the context of the innovation project *SoundTracer* supported by the University of Oslo and developed in collaboration with the National Library of Norway.

References

Bendor, D., & Wang, X. Cortical representations of pitch in monkeys and humans. Current Opinion in Neurobiology 2006;16:391–399.

Haugen, M. R. Investigating Periodic Body Motions as a Tacit Reference Structure in Norwegian Telespringar Performance. Empirical Musicology Review 2016;11:272–294.

Klapuri, A. (2006a). Auditory Model-Based Methods for Multiple Fundamental Frequency Estimation. In A. Klapuri & M. Davy (Eds.), *Signal processing methods for music transcription* (pp. 229-265). Springer.

Klapuri, A. (2006b). Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes. In *Proceedings of the Seventh International Conference on Music Information Retrieval*. US: University of Victoria.

Lartillot, O., & Toiviainen, P. A Matlab Toolbox for Musical Feature Extraction From Audio. In *Proceedings of the International Conference on Digital Audio Effects*. France: Bordeaux, 2007.

McDermott, J. H., & Oxenham, A. J. Music perception, pitch, and the auditory system. Current Opinion in Neurobiology 2008;18:452–463.

Medis, R., & Hewitt, M. J. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. The Journal of the Acoustical Society of America 1991;89:2866–2882.

MIREX 2016: Multiple Fundamental Frequency Estimation & Tracking Results - MIREX Dataset http://www.music-ir.org/mirex/wiki/2016:Multiple_Fundamental_Frequency_Estimation_%26_Tracking_Results_-_MIREX_Dataset

Oxenham, A. J. Pitch perception. Journal of Neurosciences 2012;32:13335–13338.

Schedl, M, Gómez, E., & Urbano, J. Music Information Retrieval: Recent Developments and Applications. Foundations and Trends in Information Retrieval 2014;8:127–261.

Terhardt, E. Calculating virtual pitch. Hearing Research, 1979;1:155–182.

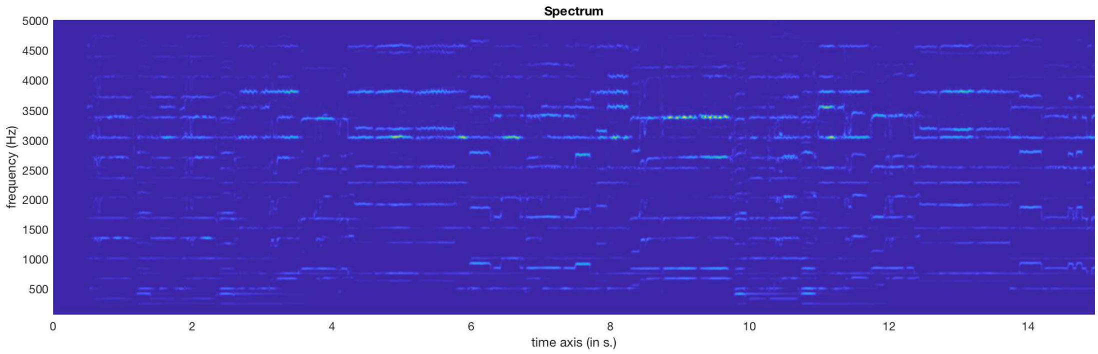


Figure 1. Frequency spectrum decomposition of the beginning of a performance of the tune called *Gibøens bruremarsj* played on Hardanger fiddle. Time in second is shown on the horizontal axis. Frequencies are decomposed along the vertical axis. The higher the magnitude, the brighter the colour.

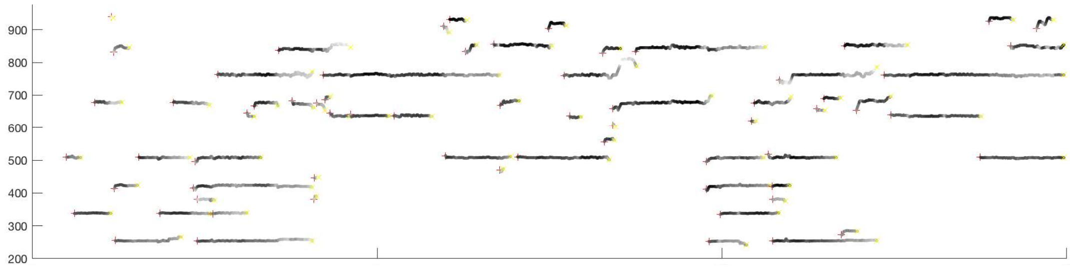


Figure 2. Pitch extracted from the audio recording shown in Figure 1. Each note is represented by a black line, starting with a red cross and ending with a yellow cross. Time is shown on the horizontal axis and frequencies on the vertical axis.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Musical Gesture and the Interpretative Construction of Contemporary Brazilian Musical Works

Tatiana Dumas Macedo

Programa de Pós-Graduação em Música da UFRJ / Cognição Musical - Processos Criativos, UFRJ, Brasil

tatianadumas@gmail.com

Abstract

This study aims to address issues that are still challenging the academic research in musical performance, especially related to the performer's gesture and the gestural conditions that determine the performance of the contemporary academic repertoire for piano. When we talk about instrumental performance, the way of doing it, that is, the way of interpreting, playing and planning the performance is relevant. The conscious development of the sequence of gestures that will constitute the instrumental performance determines in a significant way the interpretative construction. This is because the sound imagination leads the interpreter to the choice of the movements that will produce the differentiated intentional sounds, just as the gestual movements will imply imagined sonic configurations. This question arises when we deal with the interpretation of contemporary music, which contains a vast, growing repertoire that presents diverse tendencies. Because it is an interpretative practice under construction, not crystallized in consecrated aesthetic models, this practice demands from the performer interpretative decisions not foreseen in the pedagogical model that formed the interpreter. Thus, the challenges posed by the contemporary repertoire, which fuel the process of acquiring interpretative skills, emphasized in this research, contribute greatly to the pedagogy of musical performance, implying contributions to the recognition of the musical gesture within a network of elements from structures of unconventional interactions that structure the production of the musical gesture itself. The familiarity, experience and knowledge acquired by the performer about the expressive potential and the stylistic tendencies contained in the text of the work determine his musical understanding, and, consequently, the quality of the instrumental execution. The theoretical framework is based on the considerations and precepts of Robert Hatten (2004), George Kochevitsky (1967), Marc Leman (2008), Leonard Meyer (1989), among others.

Introduction

To interpret is not an isolated act that occurs only in the confrontation with a work of art. Antonio Ferraz (2013) clarifies that interpreting is something we do when we relate to the world. Therefore, it is a continuous act of dialogue and experience. Constructions of interpretations are therefore emergent in interaction and not predetermined structurally. With this in view, the musical interpretation, according to Anne Roubet (2011), is understood as a sound update of a musical piece.

Leonard Meyer (1989) explains that the interpreter is a historical being and interprets in terms of culture, inherited traditions, received education, familiarity and sensitivity to the musical style or culture in question. When confronted with the written text of the music—the score—the interpreter is faced with a work of representing the actions that will generate the musical work, in the case of prescriptive scores, and / or the effects of the musical work, in the case of descriptive scores. From then on, according to José Augusto Mannis (2012), the interpreter will proceed to identify the musical elements that will be perceived during the reading, and for its sound realization will be used a repertoire of gestures previously learned, apprehended and automated. The way the performer structures his performance, which elements he seeks to emphasize what he perceives, the strategy for the sound realization of it, reflects the way he listens to music and determines his conception of the musical work. The level of musical understanding of the performer, and consequently of the quality of performance, will depend on the experience, intimacy, and knowledge acquired by the musician over the expressive intentions and stylistic tendencies of the work (Mannis, 2012; Meyer, 1989; Silverman, 2008). Thus, the way the body embodies the conception of the work and its emotional state will constitute the singularity of its interpretation.

Contemporary academic music, and in the case of this research, Brazilian piano music contains a vast, growing repertoire encompassing diverse tendencies. Because it appears to be less homogeneous than in past periods, in relation to a set of conventions, senses, values pre-established by a consecrated practice, it demands from the performers interpretative attitudes different from those that the interpreters are accustomed to and that the traditional formation did not prepare them to it. The textural question, for example, understood as an usual procedure in composition since the Middle Ages, becomes highly significant in the contemporary compositional technique. Due to the collapse of the tonal system, texture—in the contemporary compositional technique—began to be elaborated in form of collages with the constructions of discontinuous melodies and harmonic progressions. In view of this, Souza (2004) points out that the perception of a good part of the music of the XX and XXI centuries happens through sonic blocks and the way in which they relate or are transformed in the course of the composition, befuddling the recognition of the musical gesture to be used within a network of elements of unconventional interactions that give structure to the production of the musical gesture itself.

The purpose of this study is to expose the difficulties that the text, the score imposes on the interpreter to recognize the musical gesture within a network of elements of unconventional interactions, which structure the production of the musical gesture itself. To accomplish the purpose of this study and in view of the various significations found, it is pertinent to conceptualize “musical gesture”.

The Concept of Musical Gesture

The broad concept of "gesture" proffered by authors like Marc Leman and Rolf Godøy (2010), Mannis (2012), Vilém Flusser (2014) consists of the movement of the body, representing a series of activities organized to reach a certain end, expressing intentions, feelings, ideas. The individual, therefore, incorporates the movements that are performed as a function of their interaction with the environment. This concept goes against a classic but still current and relevant reference in the field of interpretative practices, such as Gerd Kaemper (1968), which clarifies that the interaction of an individual with the environment occurs through the movements that the individual carries stored with himself from the first moments of life. In developing himself he apprehends the movements that suit him and discards those that no longer serve him.

In music, George Kochevitsky (1967) understands that the pianistic execution of a musical idea is realized by means of corporal movements, of an action determined by the will. Sound design will drive movement, and auditory and tactile sensitivity will control it. That is, the pianistic movements are subject to the sound production and, consequently, to the communication of the musical idea. Updating this understanding, Marc Leman (2008) explains that a sound can be conceived as the sound coding of an action and, in the instrumental execution, body movements are responsible for the sound production.

Irene Zavala (2012) notes that some authors use the expression "corporal gestures" as a reference to bodily movements present in musical performance and the expression "musical gestures" for compositional gestures as musical sense units, that is, musical figures resulting from writing. There are still the undifferentiated musical gestures that refer to the compositional gestures or the corporal gestures that materialize, perform the musical gestures. This type of body movement connected to the instrumental sonority is understood in this article as a musical gesture.

Elaborating on the reasoning that aims to increase and refine more and more the potentialities of the performer, Kochevitsky (1967) argues that the visual stimulation of a musical piece impels the mental elaboration of a previous sound image, which in turn will induce the choice of the appropriate gesture to produce the required sound. That is, the sound imagination, produced by the visual stimulus, will lead the interpreter to the choice of movement that will produce the differentiated sounds thought. These sounds will then be realized through the choice of the different touches pertinent to the field of pianistic technique. The gesture chosen and realized will effect the propagation of the real sound in space, which will be properly evaluated by the ear, verifying its coherence with the musical written idea. The author adds that musical writing, when it is a chain of stimuli, stimulates the cells of the visual region of the cortex and immediately pass these stimuli to the auditory region. With training, the cells of the auditory region associate physiologically with the cells of the motor area producing the reaction or motor response.

Kochevitsky (1967) points out that the stimuli come to us as signal systems (old-fashioned terminology based on the information theory of the 1950s), in which the first sign system is formed by the external stimuli that arrive directly to our receptor organs as impressions and sensations of the natural world that surround us, and the second system of signs is formed by verbal symbolism. Through it, the human being is able to perform new movements, recreate motor acts based on demonstrations and descriptions, as well as signal stimuli. Thus, the inner imagination of sound becomes the signal, which represents an action or set of actions, and which causes the motor reaction. Therefore, when we make a musical idea, our starting point is the sound image, which, because it is an auditory stimulus, should always come before the motor reaction, both in the study and in the pianistic performance. The sound resulting from this movement is quickly evaluated by the ear and, with the aid of the tactile sensation, will find the right measure to control the motor reaction according to the intention of the previous sound imagination (Kochevitsky, 1967).

In a context in which the set of pre-established conventions for the interpretative practice is laconic, as in contemporary Brazilian music, the question is in the constitution of this sound image, in the recognition and determination of the musical gesture within a network of constituted interactions between the forming elements. According to Robert Hatten (2004), the musical gesture, both from the writing and from the instrumental performance, is a form of manipulation of multiple sound parameters and inscribes itself in the sonic matter in unequivocal ways in the score, evading the musical notation. Thus, a same musical gesture can be performed by different body movements. To recognize a gesture is to recognize the configuration of this network and this search for relations confers structure to the musical gesture, being part of the cognitive process itself that seeks to understand musical composition (Hatten, 2004, Steuernagel, 2015).

The Interpretative Construction

In order to expose some difficulties of the performer in the constructions of contemporary aesthetic works—usually not contemplated in the professional-academic training cycles of instrumentalists—I present three excerpts of works, for different piano instrumentations, by distinct Brazilian composers who require diverse interpretative approaches.

The work *Os abacaxis não voam* was composed in 2001 by Guilherme Nascimento for solo piano. The piece was all written with treble clef, in the medium-high region of the piano, in which the lowest pitch is C#3, and the measure length is determined by stopwatch timing with varied indications of seconds. In general, it should be performed with a *legato* touch and should sound quite soft, with dynamics ranging from *pianissimo* (*2p*) to an extremely soft intensity (*15p*). In the middle of the piece, a section with greater density of elements is realized with a strong *subito*, connecting to other one of lower density in *4p subito*, starting the most delicate part of the work in terms of dynamics. At the end, there is a configuration that seems to synthesize the elements of the work uttering a little more movement and intensity, before ending in an ethereal way.

The melody presented in bars 15-19 (Figure 1), according to the composer's note in the score, should be performed with a *legatissimo* touch and sound slow, calm, very expressive, with fairly soft dynamics (*4p*). After the pitch Ab3, which shows a *tenuto* mark, at the end of bar 15, there is a horizontal line crossing the boundaries of this bar to the beginning of the next bar, indicating that the note in question should be extended throughout the length of the line. The same is true of the last pitch of the melody (Bb5), which should be extended to the end of bar 19. The question explicitly presented here by the composer is how the interpreter, given the timing for each bar, will define the duration of each note, avoiding rhythmic patterns, in order to constitute the melody. However, performing a melody is not only a question of choosing predetermined pitch durations. This is just one question, among many others without answers, which emerge before the score, such as: how will this dynamics, marked in the score, be realized, if the piano instrument, in general, is not able to perform soft dynamics besides *ppp*? Since the sound decay in the piano happens just after its realization, how could it be prolonged for so many seconds, in very soft dynamics, even with the use of the sustain pedal? What will the sound imagination of the composer be? At what point in time will the pitches be played? How will the melody of the excerpt from Figure 1 be performed? What relationships should be made? Is there any polarization, hierarchy? What will be the direction of the melody? Within the *legato* touch, in a fairly smooth dynamic, what kind of articulations, supports, gestures, which are not written, can be realized?

Figure 1. Bars 15-19 from the piece "Os abacaxis não voam" (2001) by Guilherme Nascimento, for piano solo.

Melodia was written in 2010 by Edson Zampronha (Figure 2) for five instruments, including a piano, and all instruments should be able to play from G#2 to F5. In the score there are six rectangles. In each one there is an entry note on the left and an exit note on the right. These entry and exit pitches are connected by other structural pitches, which are linked by lines, forming a long melody. The reading of the piece is then starting from the entry pitch and moving to the exit pitch following the lines connecting one note to the other. The sequence of pitches is to be performed at the discretion of the interpreter and backward and forward movements are allowed. The composer, in the explanatory text that indicates how to read and play the score—the general observations—, clarifies that the structural notes should be richly ornamented and these ornaments can be "chords, clusters, fast and brief melodic movements, different types of attacks, dynamics changes, timbre changes, and so on" (Zampronha, 2010, p.2-3). They can be executed in any octave and are not restricted to the pitches of the rectangle, but it is emphasized that they should address the structural notes, enhancing them. After the beginning of the piece, each instrumentalist will proceed to the note of exit independently. However, all musicians will go from one rectangle to another in a synchronized way. When a performer reaches the exit note, he must play it repeatedly or *tremolo*, almost without ornaments or very simple ornaments, indicating to the other musicians that they should go to the exit note, without, however, performing the leap movement to reach it, but through the indicated movements, that is, following the paths determined by the lines. When all the musicians reach the exit note, the rectangle is changed. The indication of dynamics, within each rectangle, is not determined. However, two types of general dynamics are suggested, which generally propose a large crescendo from *pp* to *ff*, with a return to *pp* in the last rectangle. Finally, there is an observation that the duration of the work should be approximately six minutes.

This piece presents extensive instructions, however, despite all the instructions given, many questions regarding the planning of the interpretative construction remain from two points of view: the personal and the collective. That is, how will the part of the piano be performed and how will this part be connected to the parts of the other instruments, so as to constitute a single work?

This composition is quite free in terms of execution, but within this freedom there is a basal structure that generates an organization, a logic for musical elaboration. But how will this structure be realized? How long will each rectangle be? What dynamics will be adopted in each one? Will there be any instrument that should stand out at any given moment or will the piece be performed as a sound mass, a sound block, in which all realize their ideas within the same intensity? How should this long melody, from within the rectangle, sound? How should it be organized to sound the way it was imagined? Is the form chosen the best in terms of sound ideas and execution together? How will the structural pitches be ornamented? What ornaments will be used and how will they be used? Is it possible to perform a tonal sound or should an atonal sound be evidenced? How to play each rectangle in a way they do not sound the same and repetitive? Does the composer want a pattern?

It should be noted that after the elaboration and pre-determination of possible answers to these questions, the interaction with other musicians will probably modify them, in a continuous and permanent feedback process. The performer must be attentive and open to establishing new relationships resulting from this interactivity within the group.

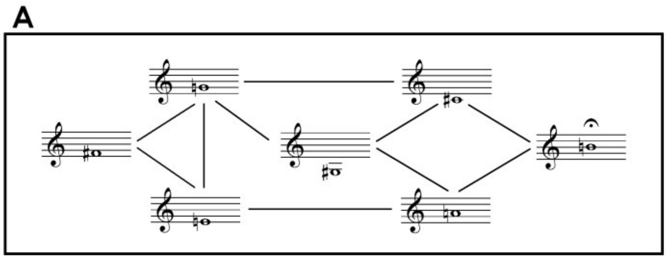


Figure 2. Excerpt / Initial rectangle of Edson Zampronha's "Melodia" (2010), for two to five instruments including piano.

Noite do Catete 14 was composed in 2016 by LC Csekö, for voice, bass clarinet, guitar, electric guitar, extended piano, percussion and light installation. The piece is written in hybrid graphic notation and measured in regular timings of *circa* 5 seconds. For this reason, each symbol used in the piece is presented at the beginning of the score, in the instructions insert, along with the specification of the material to be used and the way of playing it. The extended piano is used without preparation. For the realization of some elements / sound effects it is necessary the acquisition and manipulation of certain objects—metal and plastic bars, mallets, glasses, etc.—not usual in the pianistic practice, to produce sounds inside the piano, in the harp. The composer also inserts epigraphs in the score cover with excerpts of poetry to guide the interpretation of the piece.

The sound elements of the work are written on trigrams whose lines from bottom to top indicate the low, middle or treble regions in which they are to be performed. In the case of the piano, there are two trigrams alluding to the regions of the treble and bass clef. There are no bars. There are generally regular markings indicating timings of *circa* five seconds. The sounds demanded by the composer are written inside rectangles of improvisation *simile* and must be performed until the thick line that starts connected to the rectangle is interrupted. There is a counterpoint of dynamics, rich in detail, which must be accurately performed so that the sound blocks of the sound amalgam created by the composer can be realized.

The rectangle shown in Figure 3 exposes the last set of sound elements of the piano part, and must be played for *circa* twenty-five seconds. In this rectangle, six gestures are performed: 1) longitudinally rub the bass strings of the piano, using a hard plastic or metal blade/bar; 2) play a cluster on the piano keyboard; 3) strike, with irregular rhythm, the piano strings, with hard-headed xylophone mallets; 4) rub the strings of the piano, longitudinally, with glass or blues bottleneck); 5) play a cluster again on the piano keyboard; 6) pluck the piano strings using a hard plastic or metal blade/bar, horizontally / parallel to the keyboard. All the gestures mentioned here are previously determined to be performed with pressing sustain pedal.

The first dilemma that appears when the performer sets himself before a graphic work is the knowledge and recognition of the graphics and the automation of the musical gestures that they represent, throughout the score. How should these effects sound? Are the objects being used corresponding to the objects requested by the composer? Are the musical gestures in accordance with those intended by the composer and are they producing the sonority he devised? Are the sound effects being realized in the durations, in the correct movements and in the indicated regions? Are they rhythmically irregular and / or regular enough? In the case of the piano, most of the sound effects are performed within the harp of the instrument requiring the performer to stand up and press the pedal in an unusual and somewhat uncomfortable position for the pianist. Eventually, the pianist is going to alternate playing on the keyboard and piano harp, making it difficult to coordinate gestures and maintain internal tempo. How will the coordination of musical gestures and the manipulation of objects take effect? How to maintain the sensation of timing?

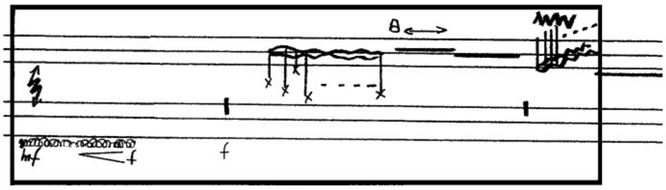


Figure 3. Last rectangle of materials from the piano part of LC Csekö's "Noite do Catete 14" (2016), for voice, bass clarinet, guitar, electric guitar, piano and percussion.

From the excerpts presented, it is perceived that the construction of musical meaning considers the body in the participation of the constitution of this sense. The experience of a musical gesture then passes through the understanding of notation, the correlation of gestures through the sensory-motor domain, the expressive dimension and the experience of the performer. Therefore evoking the cultural universe of the instrumentalist, as well as his subjective and particular identity, his affections, the intensity of his experience, his values, customs and behaviors.

Conclusion

In the light of the above considerations it can be seen that experience and imagination form creativity. At first, in the context of contemporary aesthetics, the performer does not have a cognitive plan, a repertoire of problem solving specific to the unusual issues that arise, such as: insertion of compositional elements in the work, selection and creation of materials, creation of relations between elements, participation in the structural definition of the work. To establish the formal structure of the piece, to produce the meanings that will constitute its understanding, to recognize the network of interactions between senses and gestures, to create strategies for performing and articulating the gestures that constitute the performance (which will show the intentional meanings) are qualities to be developed by the performer in the process of learning the interpretation. Thus, the process of acquiring interpretative skills becomes clearer in the face of the challenges imposed by the contemporary repertoire. This process shows therefore that the practice of contemporary music is a valuable contribution to the pedagogy of musical performance.

Interpreters learn to learn by developing strategies for their daily practice and performance. The creative process aims at solving problems that arise with each new work. This is defined cognitively as the formation of new paths and new solutions to the difficulties that present themselves for the achievement of a result, of an end product. By virtue of this, I defend the idea that the greater the experience with the demands of the interpretative construction of the current music, the more subsidies and resources the interpreter will have for the creative activity that essentialises, in a general way, his production.

References

Csekö, L.C. (2016). *Noite do Catete 14* [Catete Night 14]. Edição do autor. 1 partitura (18 p.).

Ferraz, A. M. (janeiro-junho de 2013). A liberdade na interpretação [The freedom in interpretation]. *Revista Interfaces, I*(18), 36-52.

Flusser, V. (2014). *Gestures*. (N. A. Roth, Trans.) Minneapolis, MN, USA: University of Minnesota Press.

Hatten, R. (2004). *Interpreting Musical Gestures, Topics and Tropes.* Blooming and Indianopolis: Indiana University Press.

Kaemper, G. (1968). *Techniques pianistiques* [Piano techniques]. France: Alfhonse Leduc et Cie - Éditions Musicales.

Kochevitsky, G. (1967). *The art of piano playing: a scientific approach.* Van Nuys: Summy-Birchard Inc.

Leman, M. (2008). *Embodied Music Cognition and Mediation Technology.* Cambridge, Massachusetts-London, England: The MIT Press.

Leman, M., & Godøy, R. I. (2010). Why study musical gesture? In M. Leman, & R. I. Godøy, *Musical gestures: Sound, movement, and meaning* (pp. 3-11). New York: Routledge.

Mannis, J. A. (2012). Processo criativo: do material à concretização, compreendendo interpretação-performance [Creative process: from material to concretion, comprising interpretation-performance]. *Teoria crítica e música na atualidade (Série Simpósio Internacional de Músicologia da UFRJ), 2*, 231-242.

Meyer, L. (1989). *Style and Music: theory, history and ideology.* Chicago: The University of Chicago Press.

Nascimento, G. (2001). *Abacaxis não voam* [Pineapples do not fly]. Edição do autor. 1 partitura (6 p.).

Roubet, A. (2011). Interprétation. In: C. Accaoui, *Éléments d'esthétique musicale: notions, formes et styles en musique* [Elements of musical arsthetics: notions, forms and styles in music] (pp. 257-272). Paris: Actes Sud / Cité de la musique.

Silverman, M. (2008). A performer's creative processes: implications for teaching and learning musical interpretation. *Music Education Research, 10*(2), 249-269.

Souza, A. R. (2004). *Ação e significação: Em busca de uma definição de gesto musical* [Action and meaning: in search of musical gesture's definition]. São Paulo: Instituto de Artes, Universidade Estadual Paulista (UNESP).

Steuernagel, M. S. (2015). O gesto na composição musical [The gesture in musical composition]. *Revista Vórtex, 3*(1), 146-158.

Zampronha, E. (2010). *Melodia* [Melody]. Edição do autor. 1 partitura (5 p.).

Zavala, I. P. (2012). *As inter-relações entre os gestos musicais e os gestos corporais na construção da interpretação da peça para piano solo "Sul Re" de Héctor Tosar* [The interrelationships between musical gestures and body gestures in the construction of solo piano piece "Sul Re" by Héctor Tosar]. Dissertação de mestrado em Música - UFRGS, Porto Alegre.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Metaphorical Cognition in the Discourse of Professional String Quartet Rehearsal

Su Yin Mak,[1] Hiroko Nishida,[2] Daisuke Yokomori[3]

[1]*Department of Music, The Chinese University of Hong Kong, Hong Kong*
[2]*Faculty of Design, Kyushu University, Japan*
[3]*Faculty of Languages and Cultures, Kyushu University, Japan*
[1]msy@arts.cuhk.edu.hk, [2]nishida@design.kyushu-u.ac.jp, [2]yokomori.d@gmail.com

Abstract

Current research on the role of metaphor in music analysis and cognition has focused almost exclusively on writings about music, with little experiential input from practicing musicians. This paper on the conceptual, rhetorical and social interactive roles of metaphor in rehearsal discourse is part of a long-term ethnographic project that documents and analyzes the collaborative and communicative strategies of the Hong Kong-based Romer String Quartet and the Japan-based Quartet Excelsior. Both Quartets' rehearsal sessions were recorded on a regular basis over two roughly 9-month periods to generate multiple sets of data that tracked the musicians' discourse about their creative process; interviews with the players were also conducted at various stages of the research. From the data corpus, we tagged and extracted conversational segments involving metaphorical usage for detailed analysis from four inter-related perspectives. First, using George Lakoff and Mark Johnson's theoretical framework of conceptual metaphor (1980) as a starting point, we view metaphorical language as indicative of metaphorical conceptualization, a process by which a target domain (here music) is understood in terms of the entities, properties and relations of a source domain drawn from everyday human perceptions and actions. Thus, categorization of the metaphors used by professional practicing musicians can be a means to access how they understand abstract musical concepts in terms of concrete, embodied experiences. Second, we examine agential ascriptions implied by the players' metaphorical usage, and argue that the act of performance not only realizes potentially agential elements within the musical work, but also create levels of agency beyond the score. Third, we study the metaphors as rhetorical devices and aim to discover how they serve specific discursive needs. Finally, we consider the interactions between verbal and non-verbal modes of communication in rehearsal, and hypothesize that speech and gesture serve as complementary modes of expressing conceptual metaphors about musical events and processes.

Introduction

While a number of scholars have explored the role of metaphor in the analysis and cognition of Western art music (Johnson and Larson, 2003; Larson and Johnson, 2002-3; Saslaw, 1996; Spitzer, 2004; Zbikowski, 2002), the extant research has focused almost exclusively on writings about music, with little experiential input from practicing musicians. This paper reports on the preliminary outcome of an ongoing research project on the conceptual and rhetorical roles of metaphor in rehearsal communication. Its methodology draws upon perspectives from music theory, ethnomusicology and linguistics, and is grounded in empirical interview and audiovisual rehearsal data of two professional string quartets.

The project is a follow-up to the first author's earlier ethnographic-documentary study of how the Hong Kong-based based Romer String Quartet perceived, conceptualized and communicated about musical structure in their performance preparation (Mak, 2016). The research findings of this earlier study suggest that the professional practice of performers relies upon latent processes of music analysis, and that metaphorical language plays a vital role when these processes of analysis are made manifest during the discussions and negotiations that take place during rehearsal. The first author thus decided to re-examine the Romer data with focus on uses of metaphor in the players' rehearsal discourse, and expanded the project by collaborating with Hiroko Nishida and Daisuke Yokomori, the second and third authors of the present paper, on a second case study of the Japan-based Quartet Excelsior.

Research Subjects

The Romer String Quartet is a young professional ensemble founded in 2012. The players, with ages between 30 and 40, are native Hong Kongers who received their formal music education in the United States or the U.K. Founded in 1996, the Quartet Excelsior is a more established quartet than Romer, and the players received their training in Japan and Germany. Further information on the two ensembles may be found at the following websites: www.romerstringquartet.com, www.quartet-excelsior.jp.

Data Analysis Protocol

Both Quartets' rehearsal sessions and public performances were video- and audio-recorded on a regular basis over two roughly 9-month periods to generate multiple sets of data that tracked the musicians' discourse about their creative process; interviews and conversations with the players were also conducted at various stages of the research. Tables 1 and 2 list the rehearsals we have documented for both ensembles.

Our data analysis protocol is as follows. First, all the recorded rehearsals are transcribed verbatim in their original languages by research assistants. Next, two specific research assistants with linguistic competence in Cantonese, Japanese and English would proofread for transcription errors and draft English translations to facilitate comparative study. The three project researchers would then tag and extract conversational segments involving metaphorical usage for detailed analysis, and in the process also fine-tune the English translations to ensure that verbal nuances and rhetorical intentions are accurately represented.

We currently use the annotation software ELAN developed by the Max Planck Institute for Psycholinguistics to store our data because it has the capability to align the audiovisual footage with multiple tiers of annotation. Figure 1 illustrates how the data appears in ELAN.

Table 1. List of documented Romer String Quartet rehearsals (all video recordings except No. 2)

Rehearsals	Date	Pieces
1. With theorist-observer	Sept 19, 2014	Webern, *Langsamer Satz* Beethoven, Op. 18 No. 4
2 with theorist-observer (audio only)	Dec 4, 2014	Haydn, String Quartet No. 40 in F Major, "The Dream"
3. without theorist-observer	Dec 17, 2014	Ravel, String Quartet in F Malecki, Polish Suite
4. without theorist-observer	Jan 20, 2015	Stravinsky, Double Canon Stravinsky, Three Pieces for String Quartet Haydn, String Quartet No. 40 in F Major, "The Dream" Beethoven, String Quartet No. 11 in F minor, Op. 95, "Serioso"
5. without theorist-observer	Feb 15, 2015	Haydn, String Quartet No. 40 in F Major, "The Dream" Joyce Tang, *Lineae* Rimsky-Korsakov, Allegro in B-flat
6. without theorist-observer	Feb 18, 2015	Haydn, String Quartet No. 40 in F Major, "The Dream" Rimsky-Korsakov, Allegro in B-flat

Table 2. List of documented Quartet Excelsior rehearsals (all video recordings; dress rehearsals without conversations omitted)

Rehearsals	Date	Pieces
1. without theorist-observer	Aug 3, 2016	Mozart, Divertimento K.138 Janacek, *Kreutzer* Sonata Verdi, String Quartet
2. without theorist-observer	Jan 22, 2017	Beethoven, String Quartet Hess34 Ravel, String Quartet Beethoven, String Quartet Op.130
3. without theorist-observer	Mar 15, 2017	Schubert, String Quartet D.32
4. without theorist-observer	Mar 16, 2017	Schubert, String Quartet D.32
5. without theorist-observer	Mar 20, 2017	Schubert, String Quartet D.810
6. without theorist-observer	Mar 23, 2017	Schubert, String Quartet D.32 Schubert, String Quartet D.112 Schubert, String Quartet D.810
7. without theorist-observer	Mar 26, 2017	Schubert, String Quartet D.32 Schubert, String Quartet D.810
8. without theorist-observer	May 3, 2017	Shostakovich, String Quartet Op.110
9. without theorist-observer	June 9, 2017	Beethoven, String Quartet Op.18-2 Shostakovich, String Quartet Op.110 Schubert, String Quartet D.87

Figure 1. ELAN screen shot (converted to black-and-white)

Analytical Methodology

We approach the two quartets' metaphorical usage from four inter-related perspectives: (1) conceptualization; (2) agency; (3) rhetorical function; and (4) co-expressive modes of communication. Excerpts from the rehearsal data will illustrate our analytical methodology.

Metaphorical Conceptualization

Let us begin a conversation segment from a Romer String Quartet rehearsal (Table 3). Because the present print format does not permit presentation of audiovisual recordings, for this and other excerpts discussed in this paper, we provide the English translations of the transcribed conversations.

Table 3. "Journey" Metaphor in the Romer String Quartet's Rehearsal Conversation, September 19, 2014

Va [to Vn1]:	I want to ask, at the beginning, when the three of us see you run off, we don't follow you, right? We don't try to bring you back?
Vn2:	We shouldn't.
Va:	We just stand there and wait for you to come back?
Vn1:	If I play it like just now, would that work?
Vn2:	Just now you had a slight tendency to run …
Va:	You had a slight tendency to run, and the three of us couldn't pay attention to that.
Vn1:	Okay.
Vn2:	So, we'll continue to ignore her then.
Vn1:	Don't totally ignore me!
Va:	It's not like we would scream "don't run away!" when you start running …
Vn2:	You wouldn't just fly off suddenly …
Vn1:	No, I wouldn't …
Vn2:	…you would speed up slowly, so we would slowly ignore you.
Vn1:	Okay …
Va:	If you start running and get ignored by us, you would come back on your own. But if we start following you and calling for you to come back, then the whole thing would turn into a chase.
Vn1:	Okay!

The players' discussion of pacing and coordination in terms of running and chasing exemplifies metaphorical conceptualization, a theoretical paradigm proposed by George Lakoff and Mark Johnson in their influential study *Metaphors We Live By* (1980). According to conceptual metaphor theory, metaphor is indicative of a cognitive process by which a target domain (here music) is understood in terms of the entities, properties and relations of a source domain drawn from everyday human perceptions and actions. In this instance, the metaphor in question is what we call the "journey" metaphor, by which the musical work is imagined as a "path" and the players as moving along it. Thus, following Lakoff and Johnson, we conceive of metaphor broadly as a composite category that encompasses not only metaphor in its traditional literary sense, but also other linguistic constructions that imply cross-domain cognitive mapping, such as simile, analogy, metonymy or synecdoche. We believe that analysis of the conceptual metaphors informing the players' rehearsal conversation can be a means to access how they understand abstract musical concepts in terms of concrete experiences.

Table 4. A Provisional Taxonomy of Conceptual Metaphors About Music

	Conceptual Category	Definition	Examples
1	Motion Metaphors	(a) the music or its players as moving along a "journey;" (b) musical events as moving past a stationary observer; (c) the music as an agent that "move" the players from one location to another.	(a) "We have finally arrived at C major." (b) "C major is coming up." (c) "The music is pushing us towards C major."
2	Architectural Metaphors	The musical work as a physical structure, with hierarchically-related components, that can be perceived synchronically.	thematic "boundaries," harmonic "foundation," "arch-like" melody
3	Nature /Animal Metaphors	Comparison with natural phenomenon or animals	"stormy," "ray of sunlight," "cat-like"
4	Narrative/ dramatic Metaphors	Musical components as characters or events in a story or dramatic situation	"plot of seduction," "like a movie"
5	Labour Metaphors	The music or its performance as "substantive" or requiring "work" that leads to "reward"	"struggle," "wasted" effort, "payback" moment
6	Anthropomorphic Metaphors	Ascribing human traits, emotions or intentions to music, often signaled by adjectival description of affect	"heroic," "sentimental," "purposeful"
7	Embodied Metaphors	Comparison with human bodily states	"drunken" rhythm, "sleepy" melody
8	Synaesthetic Metaphors	Comparing sound with other sensory domains, most often sight and touch	"like a painting," "goosebump-making" chord
9	Discourse Metaphors	Comparing music to speech or conversation	"question-answer" phrasing, "say it louder"

We have been using a provisional taxonomic framework of conceptual categories, shown above in Table 4, in our analysis of the data corpus. The framework disregards language which may once have had a metaphorical origin but which have become standardized musical terminology, such as "tone colour," "phrase," and "home key." The first two categories draw upon Mark Johnson and Steve Larson's work on motion and architectural metaphors in music (Larson and Johnson, 2002-3; Johnson and Larson, 2003).

Agency

Related to the notion of conceptual metaphorization is the issue of agency. In the previous example, the players clearly saw themselves, the real-life musicians, as responsible for the metaphorical actions of "running" and "chasing." But, as Seth Monahan has pointed out (2013), sentience, volition and deeds are also often ascribed to musical elements, as well as to the instrumental parts in which they occur. In both our case studies, we found that the performers frequently and habitually switch between different levels of agency in their natural conversation. Table 5, an annotated conversational excerpt from a rehearsal session by the Quartet Excelsior, illustrates.

In this conversation, the cellist begins by suggesting an affective description of the music's "character" ("should 'it' be sentimental"), which implies an anthropomorphic metaphor. The first violinist elaborates this metaphor through suggesting an imaginary persona ("as though lovesick"), who is then explicitly enacted by the cellist ("Ahhh, I'm in despair!"). The violist's response to this exchange is interesting because of its agential ambiguity: it is unclear from her utterance who is responsible for the action "as though becoming intoxicated"—whether she means that the music's persona is "drunk with love," or that the players themselves should play in this manner. Similarly, when the second violinist offers the alternative metaphor "like a film" in her attempt to clarify the cellist's meaning, it is unclear whether the agent is the music or the players. Such fluid agential ascriptions and their interactions with metaphorical usage are of great interest to us.

Table 5. Metaphor and Agency in the Quartet Excelsior's Rehearsal Conversation, March 15, 2017, Part 1

	Conversation	Analytical Annotation
Vc:	How should I put it … should it be, like, sentimental?	Adjectival description of affect, implying an anthropomorphic metaphor; Agent: the music ('it')
Va:	Yes, yes.	
Vc:	Right? I mean, as though …	
Vn1:	[as though] one is lovesick …	Anthropomorphic metaphor elaborated through personification

Vc:	Something like, 'Ahhh, I am in despair!'	Explicit enactment of imaginary persona
Va:	So … as though becoming intoxicated?	Embodied metaphor with ambiguous agency
Vc:	Well, … it's Schubert, after all … I don't know …	
Vn2:	Um … like a film?	Alternative metaphor (narrative/dramatic), but agency remains ambiguous
Vc:	Yes, but not quite … I guess there are some parts that are beautiful, but … I mean, playing something like this, more *piano*	Verbal expression fails. Mode of communication switches to musicking.
	[Demonstrates on instrument]	
Vc:	I mean … hmm … how should I put it? The overall condition, yes, the condition, could be a bit more … how should I put it?	Searching for the right word
Va:	You mean, the shape?	Lexis becomes increasingly technical
Vc:	Yes, yes, yes. Tone colour as well.	
	[Demonstrates on instrument. Vn1 and Vn2 join in.]	Musicking
Vc:	If it's *pianissimo*, it would be something like this.	
Va:	I think it's better to keep it simple, with the feeling of an accompaniment.	
Vn2:	Like a guitar or something.	Simile without cross-domain cognitive mapping
Vn1:	[Nods] Yes, to create the atmosphere only.	
Va:	Yes.	

Rhetorical Function

The previous two examples also illustrate the different rhetorical functions metaphors can assume in the rehearsal context. In the first Romer excerpt (Table 3), a single extended metaphor, "journey," serves to clarify meaning. In the Excelsior excerpt (Table 5), the rhetorical role of metaphor is considerably more complex. Recall how the players repeatedly prefaced their utterances with "how should I put it": they seem to have found the music ineffable, and their different metaphorical "takes" on the cellist's notion of sentimentality suggest that they were using metaphor to grasp at and articulate that ineffability as they negotiated different interpretative opinions.

For the cellist in this rehearsal segment, metaphorical expression ultimately failed. Lost for words, he instead demonstrated his interpretation through playing. This exemplifies what Amanda Bayley calls "musicking": "where players use their instruments or singing rather than words to explain what they mean." (2013, 409, footnote 11). Note that Bayley's usage of the term "musicking" differs from Christopher Small's original definition, according to which the verb "to music" means "to take part, in any capacity, in a music performance, whether by performing, by listening, by rehearsing or practising, by providing material for the performance (what is called composing), or by dancing." (1988, 8).

When the quartet members resumed their conversation, their mode of communication was initially non-metaphorical, as evident in the reliance on technical music vocabulary ("*pianissimo,*" "tone colour"); although the second violinist resorted to a simile ("like a guitar"), both terms of the comparison are musical and there was no cross-domain cognitive mapping. However, at a later point in the conversation, when the players' attention switched from performance execution back to the consideration of affect, metaphorical language returned (see Table 6). The players negotiated their interpretative opinions through responding, both verbally and gesturally, to the cellist's original anthropomorphic description ("sentimental") and a new embodied metaphor ("healthy") introduced by the first violinist. The pervasive presence of humour in this segment also illustrates the ethnomusicological model of "vocal anthropology," which approaches the intertwining of speaking and musicking in musicians' discourse as a site of their social interaction (Feld et. al., 2006).

Table 6. Metaphor and Agency in the Quartet Excelsior's Rehearsal Conversation, March 15, 2017, Part 2

	Conversation	Analytical Annotation
Vn1:	Right now, our playing is far from being sentimental.	Return of affective description
Vc:	Oh, is that so?	
Va:	[Laughing at Vn1's comment] "Far from being that …" …	
Vn1:	It sounds very healthy.	New embodied metaphor
Vc:	"Healthy"…	
Vn1:	I mean, because you were playing in strict rhythm.	Offers possible technical explanation
Vc:	Then, why don't we try playing in the opposite way?	
Vn1:	[laughs]	
	[All play. Vn1 makes an exaggeratedly desperate facial expression, and Va imitates her.]	Humour through physical gesture
Va:	[While playing] Only you can make that face.	
	[Vn 1 stops the rehearsal.]	
Vn1:	No, no, no, not like that. Nooo!	
Vn2:	[Laughs]	
Va:	We were going in the wrong direction.	New journey metaphor
Vn1:	Yes, not like that! I mean, I think playing less expressively is fine, but, um, the tone colour, we have to be a bit more careful ...	Offers alternative explanation with direct verbal description
Vn2:	Maybe we were using our bows too much.	Offers technical explanation
Vn1:	…as long as it's not healthy ….	Picks up embodied metaphor again
Vc:	"Healthy" …	
	[All start playing.]	

Va:	We should sound "unhealthy" then.[laughs]	Joke: deliberately skews Vn1's metaphor
Vn1:	No, no …Okay, fine. Let's just play sentimentally then.	Reverts to original affective description

Co-expressive Modes of Communication

The constant alternations between musical, gestural and verbal modes of communication in the above example is characteristic of the rehearsal practice of both quartets in general. Moreover, we have found that metaphors are often used in tandem with two other discursive strategies:

1. Onomatopoeia: phonetic imitations of sound gestures through spoken/sung vocalizations (e.g. "ta-ta-TA"), and use of onomatopoeic words (e.g. "boom") or descriptions ("sound of boiling water");
2. Intertextual reference: citations of other works, composers, style periods, etc. (e.g. "like Tchaikovsky")

Two further examples illustrate. The first excerpt is from an Excelsior rehearsal of the Verdi Quartet (Table 7), and second excerpt is from a Romer rehearsal of Rimsky-Korsakov's Allegro in B-flat (Table 8).

Table 7. Interaction of Talking, Playing, Singing, and Use of Physical Gestures in the Quartet Excelsior's Rehearsal, August 3, 2016.

	Conversation	Analytical Annotation
Vn1:	Um… I wanted to go through bit by bit ...	
Vc:	So …?	Proposes to resume playing
Vn1:	[sings] Paaa—rappa–rappa–ra	Musicking using onomatopoeia
Vn2	Um, we slow down a bit here, right?	Responding to Vn1's singing, Vn2 raises a new question about pacing
Vn1:	Yes. Well, perhaps, (you can) play it a bit more expressively, like "tarata—taratata—tararara—*ra*—rara". [sings with hand gesture]	Musicking using onomatopoeia + gestural visualization of melodic shape
Vn2:	Yes, yes.	
Vn1:	With this … [pauses, gesticulates while searching for right word] …harmony? [plays]	Co-speech-gesture + technical description; confirmed by musicking

Table 8. Mixed Metaphors and Interaction of Talking, Playing, Singing, and Use of Physical Gestures in the Romer String Quartet's Rehearsal, February 15, 2015.

	Conversation	Analytical Annotation
Vn1:	Like that whole patch in *forte,* even here [plays on violin], we should keep that …that "princess" character	Anthropomorphic metaphor
	[Everyone laughs]	
Vn2:	Where's the ring then?	Joke: deliberately skews Vn1's metaphor
Vn1:	What I mean is, it's a bit like *Scheherazade*	Elaboration via intertextual reference
Va:	Also, at this moment, there are no obstacles yet	New metaphor (journey)
Vn1:	Indeed not! We see *forte* and forget we have to come down later …	
Va:	In my score at bar 39 there is a *forte* but it doesn't last …	Technical description
Vn1:	…it doesn't last …	
Va:	… and it falls all the way to *pianissimo*	
Vn1:	Yes, it's only a [sings]"Waa"	Musicking using onomatopoeia
Va:	I guess we need that broadness, but not that strongness – is "strongness" even a word?	Words fail
Vn1:	Yes, we don't want this [Vn1 and Vn2 make flexed-bicep gesture]	Gestural visualization
Va:	You mean, not Hercules	New anthropomorphic metaphor
Vn1:	No, not Hercules but *Scheherazade*	Confirmation by reverting to earlier intertextual reference

In both examples, the players move seamlessly between talking, playing, singing, and the use of physical gestures. In the Excelsior Quartet excerpt, the first violinist's hand gesture visualized the shape of the musical phrase she was demonstrating through singing and onomatopoeia (see Figure 2).

Figure 2. Gestural Visualization of Musical Phrase in the Quartet Excelsior's Rehearsal, August 3, 2016.

In the Romer excerpt, the violist was unable to find the right words for what he meant ("we want broadness but not strongness … is 'strongness' even a word?"), so the two violinists used a flexed-bicep gesture to help him clarify his meaning (see Figure 3). Interestingly, the violist then responded with a metaphor ("you mean, not Hercules") and the first violinist confirmed by way of an intertextual reference ("yes, not Hercules but *Scheherazade*"). From these and other examples from the data corpus, we hypothesize that speech, musicking and physical gesture serve as co-expressive and complementary means of conceptually metaphorizing musical events and structures in the players' discourse.

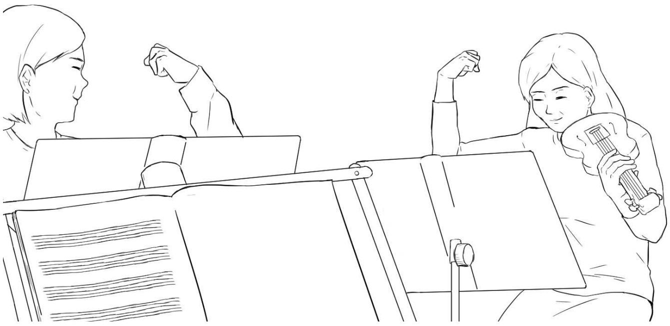


Figure 3. Flexed-bicep Gesture in the Romer String Quartet's Rehearsal, February 15, 2015.

Conclusion

In conclusion, our collaborative research analyzes metaphorical usage in professional string quartet rehearsal from conceptual, rhetorical and social interactive perspectives. Although our work is still in progress, the multiple interactions between musical, verbal and gestural communication we have been able to observe in the rehearsal data already demonstrate the multivalent and processual nature of metaphorical cognition in the music-making process. Indeed, as both conceptual and rhetorical device, metaphor mediates between abstract structure and tangible experience, between intellectual understanding and sensory perception, between the self-referentiality of music and its meaning for the performer; and in the process, one can no longer—or no longer needs to—distinguish the dancer from the dance, the musician from the music. We hope that, in proposing a new interdisciplinary methodology that integrates ethnographic-documentary and theoretical-analytical approaches for the field of performance studies in music, our research may facilitate a better understanding of the structural processes implicit in performance, provide empirical support for assessing and refining extant theoretical models, and lend fresh insight to the conventional view of the string quartet genre as a site for conversation.

Acknowledgements. We are grateful to the Romer String Quartet and the Quartet Excelsior for participating in our research, and to the Faculty of Arts, The Chinese University of Hong Kong and the Inamori Foundation for providing funding support. We also thank Yukino Tagawa for her help in preparing Figures 2 and 3.

References

Bayley, Amanda. (2011). Ethnographic Research into Contemporary String Quartet Rehearsal. *Ethnomusicology Forum,* 20, 385-411.

Feld, Steven and Aaron A. Fox, Thomas Porcello and David Samuels. (2006). Vocal Anthropology: From the Music of Language to the Language of Song. In Alessandro Duranti (Ed.), *A Companion to Linguistic Anthropology*, (pp. 321-345). Oxford: Blackwell.

Johnson, Mark, and Steve Larson. (2003). Something in the Way She Moves: Metaphors of Musical Motion. *Metaphor and Symbol* 18 (2): 63-84.

Lakoff, George, and Mark Johnson. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Larson, Steve and Mark Johnson. (2002-03). Architectural Metaphors in Music Discourse and Music Experience. *Yearbook of Comparative and General Literature*, 50, 141-54.

Mak, Su Yin. (2016). String Theory: An Ethnographic Study of a Professional Quartet in Hong Kong. *Intégral,* 30, 53-65.

Monahan, Seth. (2013). Action and Agency Revisited. *Journal of Music Theory,* 57, 321-371.

Small, Christopher. (1998). *Musicking: The Meanings of Performing and Listening*. Middletown, CT: Wesleyan University Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Reconceptualising the Functions of Listening in Everyday Life: A Domain-Based Aggregate Thematic Framework and Comparative ESM Study

Liam Thomas Maloney[1]

[1]*Music Science and Technology Research Cluster, Department of Music., University of York, UK*

[1]lm1182@york.ac.uk

Abstract

Merriam (1964) suggested the "functions of music" pertain to the underlying driver for music listening. Numerous researchers have contributed potential functions to the field of research. Yet, there exists no agreed taxonomy of suggested functions. The aims of the project were, to develop an exhaustive domain-based framework of the functions of music listening drawn from existing literature, and to confirm the validity of the proposed framework through comparison with real-world Experience Sampling Methodology (ESM) data. An aggregate qualitative thematic analysis was performed on 67 datasets. 713 functions were identified, thematically coded, and then grouped by domain of action. The resulting analysis was termed the Aggregate Thematic Functions Framework (ATFF). Following this, Consenting adults (n=76; mean age: 28, SD: 7.91; female: 34) participated in a 7-day ESM study gathering 575 qualitative descriptive logs examining potential drivers for listening. The functions identified within the logs were then thematically coded and compared with the findings of the ATFF to create a final Extended Functions Framework (EFF).

The ATFF identified 45 functions. Functions were sorted into 5 domains (*Cognitive*: 8 functions, *Emotional*: 3 functions, *Physiological*: 10 functions, *Social Group*: 9 functions and *Social Individual*: 5 functions). 1 sub-domain concerning emotional regulation within the *Emotional* domain was identified (5 functions), and 1 larger *Meta-Domain* (5 functions) encapsulating functions that augment several different aspects of the self simultaneously. The EFF identified 54 functions, with only 35 ATFF functions present in the ESM study. There was noticeable disparity between the *Social* domains (*Individual* and *Group*) in the ATFF and ESM data; the ESM data found little evidence of the functions from these domains. The remaining domains and functions approached similarity or matched. Several functions were then re-categorized based on the ESM study evidence. The study found 3 functions not proposed within the functions literature.

Introduction

Music psychology research is often perceived as somewhat removed from the realities of musicking (Clarke, Dibben, & Pitts, 2010). Indeed, it may not be possible to give a full account of the potential functions of music without exploring *why* music is drawn into action (Sloboda, 2005b) and the consequences of listening (Konecni, 1982; Krause, North, & Hewitt, 2013; Sloboda, 2005b). Music that appears to be *function*-less, such as western art music listened to as a focused exercise, has the potential to effect listeners (Clarke et al., 2010). Even passive or unintentional listening holds the potential to influence or interact with those listening (Greasley & Lamont, 2011). It is only through developing an understanding of music listening that accounts for social, cognitive, physiological, and emotional interplay that we can truly interpret the relevance of listening (Konecni, 1982).

Given the pervasiveness of the concept of "music as a resource" (DeNora, 2003; North et al. 2004; Kusek & Leonhard, 2005) it is not inconceivable that this resource is drawn into action daily for commonplace functions, and functions not merely limited to the psychological or emotional aspects of everyday life are employed. There is a large corpus of evidence that suggests music can act beyond the confines of these realms. The "functions literature" encompasses many other forms of music study, allowing us access to physiology through sports sciences, the social aspects of musicking through sociologically focused studies, and cognition from a wide range of psychological sub-fields.

The Functions of Music

Alan P. Merriam was the first to propose music as having distinct *uses* and *functions*. The notion that music was a means of conveying messages or as a means of portraying rhetorical constructs was not new, dating back to at least the latter part of the Baroque period (cf. Mattheson & Lenneberg, 1958). However, in opposition to the perspectives of rhetoric and phenomenology inquiry that had gone before, Merriam situated the importance and meaning of music not within the musical materials themselves, but rather within the employment of those materials within a socially constructed context. Merriam posited music had two socially constructed facets; *use* and *function*, and that the difference between these two interconnected notions is a significant one (Merriam, 1964).

""Use" then, refers to the situation in which music is employed in human action; "function" concerns the reasons for its employment and particularly the broader purpose which it serves." (Merriam, 1964)

Merriam highlights 10 specific functions of music in his inceptive list (emotional expression, aesthetic appreciation, entertainment, communication, symbolic representation, physical response, enforcing conformity, validating social institutions, contributing to the continuity and stability of culture, and contributing to the integration of society). Whilst these functions are limited when compared to other contemporary models, they do represent a significant and lasting contribution to the body of research in the functions literature, and arguably inspired the field's creation. Although the functions in Merriam's study are limited in number, their scope is rather comprehensive and wide-ranging, and could potentially represent larger categories of functions rather than more nuanced, precise functions.

Since Merriam's preliminary publication many scholars have approached the field of research from varying perspectives to attempt to further define and refine the functions of music in contemporary society (e.g. Bull, 2000; DeNora, 2000; Hargreaves & North, 1999; Sloboda, 2005a; Sloboda, Lamont, & Greasley, 2012; Williams, 2006). However, this abundance of models presents researchers with two problems. Firstly, Merriam's initial founding definitions have often been ignored. Many researchers include references to aesthetic judgements and points that would be considered *use* within their models. Clearly, this muddies the waters of investigation. The second, and certainly more pressing issue, is that there currently exists no consensus amongst researchers.

Yet, the continued exploration has created as much conceptual noise as it has clarity. Researchers have been presented with a potentially fruitful seam of exploration, but it is being obfuscated by sporadic study and non-unified approaches. As researchers, we find ourselves faced with one key question: is there an exhaustive or comprehensive model of music function that would fulfil both Merriam's original tenets and provide support for the work already performed into the subject? I.e. is there a possible consensus of research that can be attained? And furthermore, if such a consensus can be achieved is it possible to confirm a sense of parity with scholarly activity and real-world behaviours of listeners?

To approach the question of a consensus of the functions of music two steps were required. Firstly, an analysis of the current understanding of the functions of music. This was then followed by a real-world study of the functions of music "in action". These differing perspectives could then be compared and allow for a possible consensus. The resulting analysis and possible consensus therefore used a heavily qualitative inductive approach, rather than a metrical or ordinal approach.

Study 1: Aggregate Thematic Functions Framework

Literature Search

The search for salient literature was conducted using electronic academic databases. Using multiple keywords, the aim was to gather the broadest range of possible literature dealing with the functions of music. Keywords included: function, music, use, regulation, strategy, listening (also including plurals). The term *use* was specifically included to overcome the misidentification of function by many researchers as previously alluded to. The disciplines from which results came were not limited. A date restriction of 1964 was enforced, excluding literature from before Merriam's original publication.

This study draws together previous research into the potential functions of music by amassing 48 individual studies (11 containing multiple datasets) to glean the most expansive picture of musical functionality in everyday life, and to attempt to begin mapping out the current state of music functions employed by listeners. The studies were selected based partially on the work of Schäfer et al. (2013), and from a range of other studies that explore the field from varying disciplines. Studies came from a broad range of disciplines; music psychology, sociology, musicology, music in everyday life, music and emotion and sports science. The analysis aims to offer depth and breadth with respect to collecting and qualifying all the potential functions of music. Following Schäfer et al. this review does include studies that examine the use of other media types, but filters out reference to any media type that is not music.

Results were then filtered to those that presented novel models of function based on primary research, and to papers that presented augmentations of previous work. Studies that faithfully replicated the work of prior research were excluded to remove redundancy. 48 studies were identified and included within the analysis.

Procedure

Across the 48 studies, and the 67 datasets presented within them, it has been possible to identify 732 references to the functions of music. All papers and functions were considered equally weighted regardless of specialism or impact factor of published papers (or the journal from which they originated) or citation count.

A small number of factors (19 findings) were labelled as *incorrect* or *problematic* as they did not match Merriam's description of functions (most were strictly speaking *use*, and others were descriptions of musical aesthetics i.e. "it is beautiful"). The *incorrect* or *problematic* functions were not sorted, were excluded from the final analysis, and are not included in the data presented here. Some papers and models made specific reference to stages of life or ages, however the following does not compartmentalise any finding as a function of age.

The 713 valid referenced functions (excluding the 19 factors labelled *incorrect* or *problematic*) were sorted into one or more of 5 commonly cited domains of music function (cognitive, emotional, physiological, individualistic and social) depending on the focus of the original study and explanations or extrapolations of the function itself from within the documentation. The identified functions within each domain were thematically grouped based on their description in the original source publication and semantic coherence to form a "function" with an appropriate title and a citation count. This sorting was performed with qualitative data analysis software (QSR's NVivo).

Results

The Aggregate Thematic Functions Framework (ATFF) combines the findings of the 67 datasets to create the ATF framework presenting 45 different functions of music sorted by domain. The frequency of each function appearing within the aggregate dataset is included to show commonality of findings in the data, although this does not equate to commonality of incidences in real-world listening situations.

The most function-rich unique domain was the *physiological* domain with 10 distinct functions (and 149 references within the aggregate body). The least rich unique area was that of *social individual* with only 5 functions (and only 50 references within the aggregate body). Some functions within the findings crossed various domains of functionality. These functions do not represent multiple functions but rather small differences in how functions operate. Hence, the *multi-domain* area contains only 5 functions, but these functions have been referenced in numerous domains from multiple perspectives and represent 12 functions.

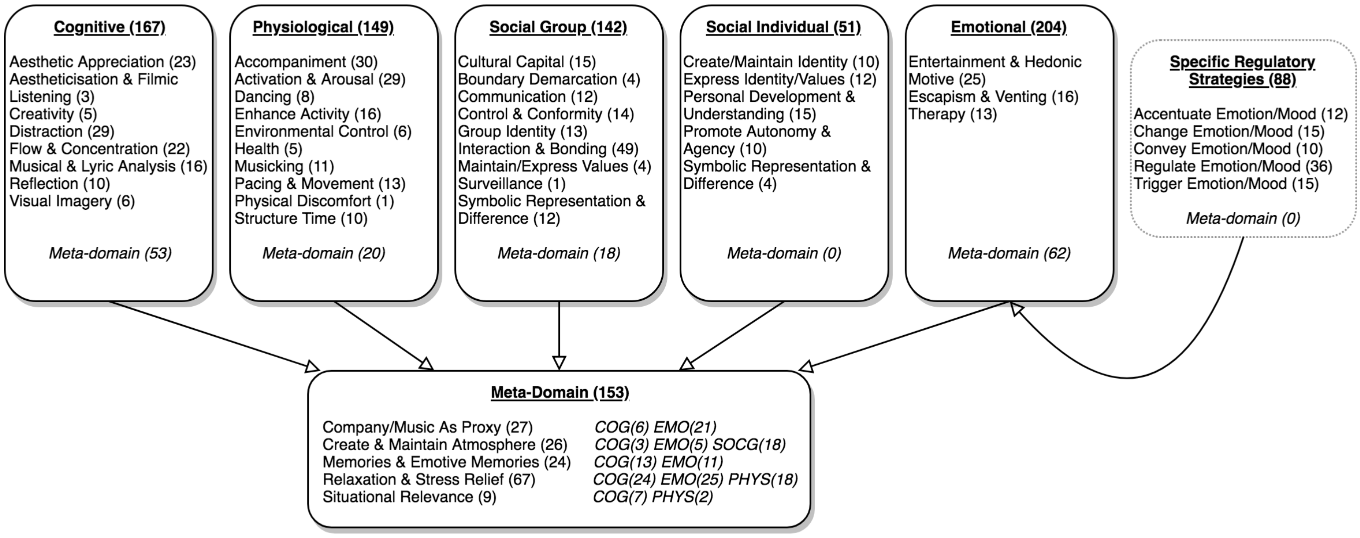


Figure 1. Visualisation of proposed Aggregate Thematic Functions Framework (*n*=45)

The *emotional* domain is the only grouping of functions that presents a particularly striking point of delineation allowing for a grouping or sub-domain to develop. Within the literature, numerous references were made to functions that in some way manipulated emotions, but in vague terms, i.e. "to influence my feelings" (Juslin & Västfjäll, 2008) and "mood management" (Moebius & Michel-Annen, 1994; North, Hargreaves, & O'Neill, 2000; Williams, 2006). These functions strongly reference the emotional domain of function, but their specific meaning remains opaque. When this is combined with other very specific emotional functions that suggest a directionality in the shift in arousal or an emotional reflex, it is possible to conceptualise a particular subset of emotional functions that are concerned with mood or emotional management that attempts to describe the possible directionality of feelings (increasing, accentuating, maintaining, changing, or triggering). These *Specific Regulatory Strategies* stand apart from other emotional functions as they deal with the constant varying state of an individual, rather than other functions under the domain heading that deal with very specific episodes concerning emotion.

Study 2: Extended Functions Framework

Procedure

To confirm and explore the parity between the ATFF and real-world listening habits, an experience sampling methodology (ESM) study was designed. The ESM study required participants to report various aspects of their everyday musical engagement over the course of 7 days. For this analysis, only data collected concerning situational goals, music selection considerations, and reported influences/effects were used to identify the possible functions of music employed in each listening episode. Additional data was gathered for further research.

The "Functions of Listening" ESM study drew on 76 individuals (mean age: 28, SD: 7.91; female: 34) who rated themselves as "highly engaged listeners". Participants were identified in public spaces by their use of headphones or earphones and approached. Consenting participants were sent 3 SMS messages requesting them to fill in a listening log at random intervals throughout the day for 7 days, with at least 2 hours between log requests. No messages were sent between 10:30pm and 7:30am. Over the course of the 7-day study 19 participants were removed either at the request of participants or after 2 days of non-activity. 57 participants (75%) completed the study.

872 total logs were gathered, of which 574 logs (65.82%) contained listening events. The salient questions concerning drivers for listening and effects were then analysed using the same methodology as Procedure 1. 1039 unique references to functions were identified; many logs included multiple simultaneous functions. Here a divergent analysis methodology was applied. Firstly, a deductive qualitative analysis was conducted, coding all functions within the literature to their corresponding domain and function from the previous ATFF analysis. Any functions that did not correspond appropriately with a function from the AFTF were coded as "unknown". The second portion of the analysis saw the data in each function examined for further sub-themes within functions, and the unknown functions were coded inductively to new possible functions. This sorting was also performed exclusively within NVivo.

10 functions appearing in the ATFF were not found in the 7-day ESM study (Health, Group Identity, Maintain/Express Values in *social group*, Surveillance, Symbolic Representation in *social group,* Create & Maintain Identity, Express Identity/Values, Symbolic Representation in *social individual*, and Therapy). These functions are included in the EFF even though they were not identified in the short study; the justification of which is to further the exhaustiveness of the framework rather than simply replicate a small-scale study. The results were then visualised in a similar manner to the ATFF. This is referred to as the Extended Functions Framework (EFF).

Results

In the EFF the number of functions has increased to 54 and shows a significant refining of some concepts. 9 functions were identified in the ESM study that had little or no evidence in the ATFF literature: consequently, some functions have been significantly expanded with "sub-functions", and functions that had been grouped in the previous analysis were

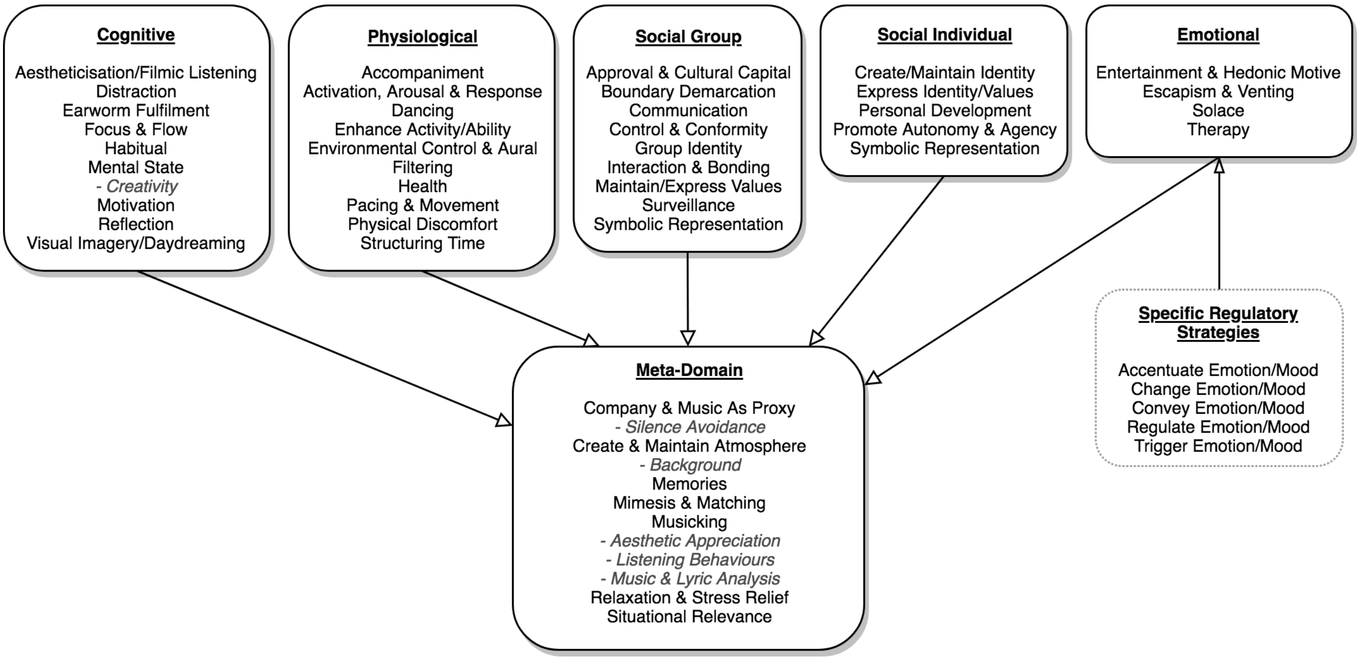


Figure 2. Visualisation of proposed Extended Functions Framework (*n*=54)

separated. The reapportionment of functions in the EFF has seen an increase from 5 to 12 functions in the *multi-domain*, and sees the *multi-domain* become the most function-rich domain. A frequency count has purposely been excluded here as this would only represent one cohort rather than any guide as to real-world mean employment of functions.

Concerning sub-functions, several functions were re-categorised: to "Company & Music as Proxy" was attributed the sub-function of "Silence Avoidance", to "Create & Maintain Atmosphere" was attributed the sub-function of "Background", and "Musicking" was reconceptualised to include "Music & Lyric Analysis", "Aesthetic Appreciation" and "Listening Behaviours". Furthermore, the "Creativity" function is apportioned as a sub-function under the wider function of "Mental State". Functions previously apportioned elsewhere in the ATFF were reconceptualised in the EFF; "Motivation" in the *cognitive* domain, and "Solace" in the *emotional* domain. Most interestingly, 3 functions that are not discussed within the literature underpinning the ATFF were found within the study data. These being "Earworm Fulfilment", "Habitual", and "Mimesis & Matching".

Discussion

There does exist significant parity between the proposed academic consensus and the real-world application thereof. 35 functions (77.77%) from the ATFF were found in the ESM study. Further to which, 3 new functions were identified. There was a significant conceptual re-arrangement of the existing research when compared with the study evidence. The iterative nature of both the ATFF and EFF suggests the current final function total across all studies and the ESM study is 54 functions.

The reapportionment of various functions and sub-functions was a necessity to more accurately represent the framework's approach to function. Interestingly, a "domain jump" was present for Musicking and its sub-functions. This was necessitated by the nature of Musicking (Small, 2011) and the possible functions and effects of musical engagement (not simply performance, but the various forms of listening practice that musicking also encompasses). Further functions also required reapportionment upon completion of the ESM study, owing to the relatively oblique nature of the descriptions in the source literature used to construct the ATFF.

Whilst the possible consensus the EFF represents is the key aim of these studies, the identification and proposal of newly identified functions is worthy of note. The new functions (Earworm Fulfilment, Habitual Listening and Mimesis & Matching) have no evidence in the 48 publications used as sources for the ATFF. However, there is supporting evidence beyond the functions literature. Whilst earworms and their prevalence is an area of current interest to many researchers, no contemporary models of function include the concept of earworms. Participants reported the need to "clear" an earworm or "get the song out of my head"; phrases that closely mirror the typical descriptions of earworms in the related research.

The remaining two newly identified functions (Habitual, and Mimesis & Matching) require further investigation and description. Habitual functions appear surrounded by descriptions such as "just because it's what I usually do" or "I normally have music on". The driver appears to be a conditioned or operant response rather than based on anything connected to the specific musical materials that are engaged with. With respect to the Mimesis & Matching function, the, albeit scant, evidence appears to point to some interplay or internal sense of mimesis between the location and the musical materials. Listeners report using music to mirror aspects of their surroundings such as geography, architecture, weather, and even time of day. These two functions require further exploration and analysis.

Finally, the 10 functions that were not found in the ESM study require reassessment. It is feasible these functions are simply relatively rare "in the wild" or that listeners have a lack of self-awareness concerning many of these functions, particularly those concerning identity work (DeNora, 2000). It is possible that these functions are occurring concurrently with other functions, but their presence is obfuscated behind more recognisable or easily comprehended functions such as Relaxation or Accompaniment. However, as much of the research that suggests these functions is predicated on adolescents and the process of physical and emotional maturation, there is a possibility that these functions were not present due to the relatively mature age of participants.

Conclusion

The studies presented here attempt to compile a consensus as to the functions of music. The work presents a prototype domain-based taxonomy of the functions of music working from an aggregate dataset combined with real-world data. The 54 suggested functions, and their domain allocations, represent as exhaustive a model as possible. The studies blend previous scholarly activity, both theoretical and experimental, with new study data to provide a possible consensus. However, the study found a lack of similitude between the literature and the real-world listening of participants in some respects. The study found several functions not proposed within the functions literature. Further study of these functions is required to approach an exhaustive taxonomy of the functions of music.

References

Bull, M. (2000). *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. Oxford: Berg.

Clarke, E., Dibben, N., & Pitts, S. (2010). *Music and Mind in Everyday Life*. Oxford: Oxford University Press.

DeNora, T. (2000). Music and self identity. In *Music In Everyday Life* (pp. 141–147). Cambridge: Cambridge University Press.

DeNora, T. (2003). Music sociology: getting the music into the action. *British Journal of Music Education*, *20*(2), 165–177.

Greasley, A. E., & Lamont, A. (2011). Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae*, *15*(1), 45–71.

Hargreaves, D. J., & North, A. C. (1999). The Functions of Music in Everyday Life: Redefining the Social in Music Psychology. *Psychology of Music*, *27*(1), 71–83.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *The Behavioral and Brain Sciences*, *31*, 559-575; discussion 575-621.

Konecni, V. J. (1982). Social Interaction and Musical Preference. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 497–516). Cambridge, MA: Elsevier.

Krause, A., North, A. C., & Hewitt, L. Y. (2013). Music-listening in everyday life: Devices and choice. *Psychology of Music*, (August), 1–16.

Kusek, D., & Leonhard, G. (2005). *The Future of Music: Manifesto For The Digital Music Revolution*. Boston: Berklee Press.

Mattheson, J., & Lenneberg, H. (1958). Johann Mattheson on Affect and Rhetoric in Music. *Journal of Music Theory*, *2*(1), 47–84.

Merriam, A. P. (1964). *The Anthropology of Music*. Evanston, IL: Northwestern University Press.

Moebius, H., & Michel-Annen, B. (1994). Colouring the grey everyday: the psychology of the Walkman. *Free Associations*, *4*(32), 570–576.

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of Music in Everyday Life. *Music Perception*, *22*(l), 41–77.

North, A. C., Hargreaves, D. J., & O'Neill, S. A. (2000). The importance of music to adolescents. *British Journal of Educational Psychology*, *70*(2), 255–272.

Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. *Frontiers in Psychology*, *4*(August), 511.

Sloboda, J. A. (2005a). Everyday Uses of Music Listening: A Preliminary Study. In *exploring the musical mind: COGNITION EMOTION ABILITY FUNCTION* (pp. 317–331). Oxford: Oxford University Press.

Sloboda, J. A. (2005b). *Exploring the musical mind: cognition, emotion, ability, function*. Oxford: Oxford University Press.

Sloboda, J. A., Lamont, A., & Greasley, A. (2012). Choosing to hear music. In *Oxford Handbook of Music Psychology* (pp. 1–11). Oxford: Oxford University Press.

Small, C. (2011). *Musicking: The Meanings of Performing and Listening*. Wesleyan University Press.

Williams, A. (2006). *Portable Music & Its Functions*. New York: Peter Lang Publishing.

N.B. source publication references for ATFF available on request.

Expressive alignment with timbre: changes of sound-kinetic patterns during the break routine of an electronic dance music set

María Marchiano[1] and Isabel Cecilia Martínez[1]

[1]*Universidad Nacional de La Plata, Argentina*

[1]marchiano.maria@gmail.com, [2]isabelmartinez@fba.unlp.edu.ar

Abstract

In this work, we describe the changes of sound-kinetic patterns during break routine's sequences (break-down, build-up, and drop sequence) of an electronic dance music (EDM) party, resulting from the expressive alignment between human movement patterns and sonic patterns of music. Expressive alignment refers to synchronization of these two patterns. On one side, the study of EDM's dance has shown that dancers share a movement pattern at the drop moment in some parties, and on the other side, expressive alignment with timbral patterns have not been studied yet. In order to analyze the sound patterns, (i) we registered and described all break routines of an EDM party video, (ii) processed its audio signal to extract acoustical features related to timbre, rhythm and pitch, and (iii) compared these two set of data. To study the human movement, (i) we identified and registered the movement of arms, head and shoulders of 37 kinetic patterns during 15 break routines, (ii) described them with the effort-shape elements from Laban Movement Analysis, and (iii) compared the kinetic patterns. With regard to sound analysis, we identified 65 break routines and observed that they are defined by sound patterns. The drop and the break-down sections are specifically linked to acoustical changes related to timbre. With regard to movement analysis, we identified changes in movement patterns during each step of the break routine, and noted that people develop personal movement patterns, differentiated from those of others by the organization of effort-shape elements. From a musical point of view, acoustical changes of the break routine –strongly based on timbral changes- modify the sonic environmental conditions, and the consequent expressive alignment of the dancers. Although people keep their personal styles of movement, they all change their movement patterns in phase with the sonic changes of music, producing a shared sound-kinetic pattern.

Background

The nature of the relation between music and movement has concerned researchers for a long time. Does the music move us or do we move with music?

The idea that music moves us is strongly based on the theory of affordances. Affordances are relations between an organism and their environment, in which "the environment affords a behavior to the organism" (Chemero, 2003, p. 187). In a musical context, this means that a sound environment affords some possibilities of action to people, and that these affordances change when the sound environment changes. From this perspective, music takes the shape of moving sonic forms (Leman, 2007; Hanslick, 1854), since it is the physical movement of sound which makes us move.

The empirical study of musical affordances has focused on the search for common patterns of music and human movement. The process of synchronizing these patterns is called expressive alignment (Leman, 2016), and is defined as a continuous transition between the environment processing and the intentional and affective states of the person. The result of the expressive alignment of a musical and a human movement pattern is a sound-kinetic pattern. Several studies have shown that the kinetic patterns made with music reflect some aspects of the sound (Kohn and Eitan, 2009; Maes et al., 2014; Krueger, 2013). This leads to the assumption that people move in similar ways with the same music, since they share expressive affordances with its moving sonic forms (Leman, 2016; Leman et al., 2009; Maes et al., 2014). According to Leman, there are two types of expressive alignment of music and human movement patterns: (i) the alignment of the salient time markers of body and musical rhythms -studied in terms of timing framework-, and (ii) the alignment of the continuous sound and body movements – normally studied with motion capture technologies. Most of the studies of kinetic pattern's formation with music are based in the analysis of non-natural situations, mainly of laboratory experiments in which people are asked to move spontaneously to a wide range of musical extracts.

Musical patterns have been studied from two main approaches. One is auditory analysis, which is the traditional and most common way to study musical characteristics, and the other is the computational analysis of the acoustical features of recorded sounds. The role of timbre in the formation of musical patterns has been little studied from both approaches. On the one hand, their auditory analysis is problematic, because we have very few categories to think of our perception of timbre (for a musicological perspective, see Mastropietro, 2014). On the other hand, timbre is studied from an acoustic point of view, looking at the acoustic dimensions which determine our auditory perception of timbre. In monophonic timbres, the spectral centroid, attack-time, spectral flux and roughness seem to be central (McAdams, 1999), and in polyphonic timbres these are the sub-band flux, spectral entropy, zero-crossing rate, and –although less relevant- MFCCs (Alluri, 2012). These dimensions have been successful in the global characterization of musical styles (Acouturier, Pachet and Sandler, 2005; Alluri, 2012), but their power to create musical patterns has been little assessed. Burger et al. (2013) studied the expressive alignment between acoustical patterns related to timbre (sub-band flux and percussiveness) and the continuous body movements that people made with the music.

We think that an electronic dance music (EDM) party is an optimal and natural environment to study the expressive alignment of body movements and timbral patterns of music, as much for the non-choreographed characteristic of its dance (Gallo, 2014) as for the perceptual relevance of its timbral features (Anzil, 2016; Marchiano, Martínez and Damesón, in press). The most important musical pattern defined by the

EDM producers and DJs is the *break routine*, formed by the *breakdown - build up - drop* sequence, which constitutes a formal articulation of the EDM set. At the breakdown section, the bass drum –and normally the bass too- is removed, breaking down "the groove and intensity of the track. The build-up section builds it up to a peak which is symbolized by dropping down the bass and bass drum" (Solberg, 2014, p. 67). With respect to the expressive alignment in EDM parties, Solberg and Jensenius (2017) found that dancers share a movement pattern at the drop moment, raising arms together.

In this work we studied the sonic patterns of break routines formed by acoustical features, aiming to determine their timbral status and their power to afford kinetic patterns in EDM's dance.

Method

We designed a mixed method consisting firstly of a qualitative analysis of sound, followed by two quantitative evaluations of sound. Movement was analyzed qualitatively.

Stimulus

The stimuli used was an EDM party audiovisual record of the techno DJ Len Faki, 1 hour and 32 minutes long (Len Faki, 2014). We looked for an EDM music genre with a lot of break routines, and a record with audiovisual characteristics that allowed us to analyze people movements. The record was selected from among the 10 most viewed videos of the Boiler Room Youtube channel (5.1 millions). It is a common practice for DJ's to be at the side of the dance floor, and that people dance around them. The Boiler Room's records normally show the DJ, the people near him/her and the crowd behind them, from the waist up (see Figure 3).

Sound Analysis

Three methodologies for sound analysis were used: aural analysis, computational modelling, and a comparison between both. In all cases, the analysis focuses on the moments of musical changes.

Aural analysis. Firstly, we identified all the break routines of the video and registered them on Elan 5.0 timeline. Secondly, the break routines were divided into breakdowns, build ups, and drops. Thirdly, we made a formal and timbral description. Finally, we identified differences between the break routines, registering five main break routine types defined by their steps sequence.

Computational acoustic analysis. The acoustic signal of the video was processed with MIRToolbox 1.7 on MatLab v2015a. We extracted data of several features related to timbre, rhythm and pitch perception. The set of timbral features extracted was sub-band flux, spectral entropy, MFCCs, zero crossing-rate, roughness and general spectral information. The rhythm feature used was the fluctuation patterns and the pitch feature was the chroma. We decided not to include the percussiveness in the timbre set of features because its relevance in polyphonic timbres has not yet been widely studied and because we thought that it can also be related to rhythm aspects, especially in a computational modelling that search for sudden changes.

The novelty curves of each feature data were then computed, in which the peaks represent the moments of changes in the sound dimension defined by the feature (Lartillot et al., 2013; Hartmann, Lartillot and Toivianen, 2017).

Comparative analysis of aural and computational data.

Temporal coincidences between changes in each acoustical feature and the steps of all break routines were identified by extracting segments of computational analysis corresponding to all aurally identified break routines. Ten seconds after and before them were added, in order to obtain information on the first and the last step of the break routines. Temporal coincidences were searched with a 1 second window around the aural temporal location of each break routine step. The percentage of coincidences for each feature and each step was calculated, and three sets of data were made: (i) general percentages by feature (the percentage of coincidences between the beginning of all break routine steps and the novelty peaks of the feature), and (ii) differentiated percentages by feature and steps.

Movement Analysis

15 break routines from the video were selected, showing people at least from the waist up throughout the segment. We then analyzed the kinetic patterns formed by arms, head and shoulders. In total, 37 sequences of movements were analyzed, from 10 persons dancing at different moments in time at the party.

The movement analysis was done in Elan 5.0 software. The music was silenced during observations. Firstly, we detected the kinetic patterns by real-time annotations and then we reviewed and temporarily adjusted the identified units. Secondly, we characterized each unit -in the context of the sequence- with the effort-shape elements from the Laban Movement Analysis (Laban, 1950). In most cases, the outcome of this process was a new and more global kinetic pattern, formed by the clustering of the local units previously identified.

Movements were described with respect to three main aspects of the Laban theory. One is the motion factors of the effort elements defined by weight (strong/light), time (sudden/sustained), and space (direct/indirect), and the other is the shape of the movement, defined by the vertical (up/down or rising/sinking), horizontal (left/right or widening/narrowing), and sagittal (forward/back or advancing/retreating) spatial axes (Broughton and Davidson, 2016). The third is the movement quantity, which is determined by movement velocity and the quantity of bodily articulations involved. Not all of these factors are relevant in all movements (Laban, 1950). We registered those factors that better describe each kinetic unit and its relation to the changes from the previous and to the following units of the pattern. After these descriptions, we did a comparative analysis between each person's patterns of movement.

Results

Sound Analysis

From the aural analysis, 65 break routines were identified.

Firstly, during the registering of the break routine sequences and in addition to the break-down, build-up and drop steps, we found that in the selected EDM set the drop was usually preceded by a break-beat –a short percussive musical fragment-, apparently working as an anticipation (14 break routines with break-beats; 20.6% of the total).

The formal analysis of all break routines showed marked differences in the order of appearance of the steps and the quantity of steps. For the purpose of this study, we categorized the break routines in five types: basic, basic with breakbeat, minimum non-directional, minimum directional, and complex break routines (Figure 1). We describe the *basic* type in the background section, because it is the most well-known and it seems to represent the basic structure of break routines. It is formed by a break-down, build-up and drop sequence, and its structure is the most directed to the drop. However, this was not the most frequent type in the analyzed set (15.38%). Sometimes, the DJ introduces a break-beat before the drop in the basic sequences, so we named it *basic with break-beat* type. The next two break routines present only two steps, occasionally with a break-beat between them: the *minimum non-directional* type is formed by a break-down and drop sequence and it is the most frequently played in our set (50.76%), and the *minimum directional* type is composed of a build-up and drop sequence. These two break routines are normally short and seem to have less structural significance in the complete EDM set, while the basics types work as formal articulators. The last type is the *complex* break routine, and it has many steps, ordered differently from the well-known basic break routine sequence. This type tends to be temporarily longer than the others.

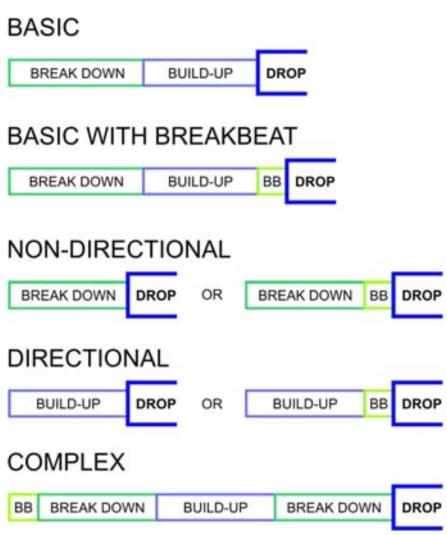


Figure 1. Types of breaks routines.

Finally, we identified two main types of break-downs, defined by their timbral identity: one is defined by the removal of low frequencies, and the other removed the higher zone. The break-downs were then divided into low-cut and high-cut procedures for the statistical analysis.

Computational analysis shows that the break routine is defined by acoustical patterns. From the general percentages by feature, we observed that some of the analyzed features are temporally aligned with the break routine sequence (Table 1). These are some timbral features, chroma and percussiveness, while fluctuation patterns –the feature related to rhythm- does not show relevant temporal coincidences. The Sub-Band Flux is the feature with the highest percentage of coincidences (68.15%).

Table 1. General percentage by features: percentage of total coincidences between all break routines steps and features changes (values without build-up coincidences).

	Features	Coincidences (%)
Timbre	Spectrum	43.31
	Sub-Band Flux	68.15
	MFCCs	28.02
	Roughness	31.84
	Zero-crossing rate	48.4
	Entropy	30.57
Pitch	Chroma	40.12
Rhythm	Fluctuation patterns	17.19
	Percussiveness	43.31

In a more specific analysis of the break routine sequence, we found significant differences in the temporal coincidences for each step (Figure 2). The drop and -in second place- the low-cut break-down moments have the highest quantity of temporal coincidences. The acoustical pattern that defines the drop is formed by Sub-Band Flux and Percussiveness changes, while the one which define the low-cut break-down is formed by Sub-Band Flux and zero-crossing rate changes

The drop and the low-cut break down steps are characterized by a change in the low-zone frequency –the first introducing the lowest layers, and the second removing them. If we assume that percussiveness plays a role in our perception of timbre –as some studies claim (Burger et al., 2013)-, it seems that the timbral features have more marked relevance in the definition of low frequencies changes in EDM than other features. If not, and if this feature also plays a role in some rhythmical aspect, then the sound pattern of the drop moment would be timbral and rhythmical, and this is intimately related to the drum bass re-introduction at the drop moment. This feature needs to be specifically studied for further analysis. In any case, the actual relevance of the acoustical features related to timbre in the definition of the drop moment and the low-cut break-down beginning does not discard any change in other musical or sonic dimensions, as shown by the descriptions of rhythm, texture, intensity, and general spectral characteristics of Anzil (2016) and Solberg (2014, 2016).

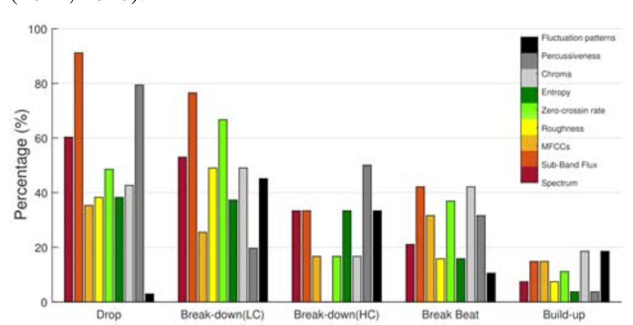


Figure 2. Temporal coincidences between all break routine's steps and features' novelty

We suggest that the reason for the low percentage of the build-up steps is that they never have a sudden onset. As inferred by its name, the build-up develops a slow and progressive transformation of some textural layers, and there is not a perceptual or an acoustical event that determines its beginning. Given the methodological failure to grasp the acoustic background of this step, we removed it from the general percentages (Table 1).

We did not observe any relevant or new links between the acoustical features and the break routine types.

Movement Analysis

With regard to the Laban Movement Analysis of each kinetic pattern, we firstly observed that dance movements with EDM can be described with effort-shape elements. People organize their movements emphasizing some elements of effort (weight, time, space) or/and shape (vertical, horizontal, sagittal). Their specific combination and sustainment over time form a kinetic unit. For example, the unit can be made of a simple rise, direct and slow movement of arms and a sinking movement of head (Figure 3, 3.2 build-up moment, man), or of a more complex movement like that of the girl on the left before the break routine (Figure 3, 1.1 and 1.2). Secondly, we observed that people shift these units of movements, forming kinetic patterns in the form of a chain (if the units and their changes are temporarily and constructively clear) or a perpetual stream (if their units connect with others in a continuous transition).

Based on this analysis, we compared the kinetic patterns of the same person at different moments in time at the party, and of different persons at the same moment. The first result showed that people in EDM parties do not share kinetic patterns or even their units of movements. As it can be seen in Figure 3, in the break-down segment (2.1), the girl on the left and the man make really different movements, and after the drop moment (3.1 and 3.2) while these two use an important amount of space moving their arms, the girl on the right barely moves them. At this point it is necessary to clarify, that in this specific break routine most of the dancers raise their arms at the drop moment, as Solberg (2017) has already observed, producing a shared kinetic pattern. However, this generalized movement was observed only in this break routine, which constitutes one *basic* and the most directional type of movement. Moreover, it is possible that the upward movement of the DJ's arms may also have led to this behavior. The role of the DJ's movements in the crowd's shared kinetic patterns would need to be studied for further analysis. The differences found in this study compared to that of Solberg may be related to EDM genre differences (house vs. techno; Solberg, 2016), to differences between set-ups (a club-environment recreated in a laboratory and an audiovisual record of some dancers in a real party), and to some probable cultural distance between the dancers of the two studies.

The second result is that people develop personal movement patterns, differentiated from those of others by the organization of effort-shape elements. On the one hand, we noted that each person frequently repeated certain movements that were not shared by others. For example, the marked directions of movement of the right girl and the man (Figure 3, 1.1 and 1.2) define one of the movements of each of these persons that give identity to their personal dance styles, at least in the sense that they repeat them a lot throughout the entire party. The arm, head and shoulder of the girl move from left to right, while the limbs of the man move in different directions. On the other hand, both the kind of temporal organization and the level of variation of the movements are key aspects in the definition of personal styles. Some people repeat a basic movement for a long time, changing only details (as the men in the picture, who repeats the described unit for 30 seconds, with little changes in his arm movements), and others change rapidly from one movement to another (the girl's unit take only 5 seconds).

We noted that underlying this, there is a common background of movements shared by EDM dancers. We identified two: one is the swaying in sync with some pulsation, and the other is the sagittal movement of arms from the chest to the front. But even these basic movements, mostly defined by shape elements, take different effort and timing factors depending on the person.

Figure 3. Laban Movement Analysis of some movements shapes during a break routine (originally made on Elan).

The examples given above are based in the shape of the movements because is the only aspect that we can appreciate in a photograph, but this analysis was done for all effort-shape elements, and the illustrative examples are a synthesis of the results of all 37 analyzed kinetic patterns.

With regard to the alignment of the kinetic and sound patterns, we found that people change their movements when the music changes during the break routine, which means that there is also a temporal alignment between people movements. Although they do not share the same effort-shape characteristics of movement, at certain moments all or most people make a change in some aspect of their ongoing kinetic pattern. The shared thing is the change.

The sound-kinetic pattern in each step of the break routine is analyzed showing that the most relevant expressive alignment happens at the drop moment. In 35 of the 37 kinetic patterns analyzed (94.5%) an important change occurs. Four of these 35 persons change their movements some instants before the drop, all during a break-beat step, and maintain the new unit after it. The rest of the people shift their pattern at exactly the same moment or until one second after the drop. The accuracy of this change on the kinetic patterns seems to be clearly related to the strong acoustical change that defines the drop moment. The kinetic change can take two forms: the occurrence of a new or previously little used movement that looks like a kind of kinetic accent, or just the change to a common personal movement that sometimes is played with a higher amount of movement quantity or a stronger weight effort. We did not find any link between these two types of changes and the sound characteristics, and neither did we find logic in the kinetic possibilities of each person.

Movements do not change immediately at the start of the break-down section. In a high percentage of cases, the previous kinetic pattern remains for a few seconds, and then slowly begins to decrease the quantity of movement. Sometimes the fight attitude of the motion factors (strong, sudden, direct) gradually changes to light, sustained and indirect, while maintaining the shape of the movement. Other times the dancer simply stops moving. This is visible in Figure 3, where the girl on the left side and the man are dancing at the beginning of the break-down (2.1), and seconds later stop moving, appearing distracted from the music (2.2). Although the aural and the acoustic analysis showed that this section has a clear beginning, we consider that their non-temporarily accurate movement's alignments may be due to an unpredictable start.

By contrast, in the build-up section, most people gradually increase the movement quantity of their kinetic patterns, and this slow change is clearly linked with the slow changes in the music. The bodily behaviors during break-downs and build-ups seem almost a direct embodied interpretation of the music: the movement disappears with the lower layers in the break-down step, and it reappears and takes strength when the layers build up again until the drop.

In summary, the different nature of all these kinetic alignments maintains a close relationship with the acoustic characteristics of the beak routine's steps. But the temporal organization of break routine's steps also plays a role in kinetic affordances. So finally, we reviewed our movement's analysis grouping the types of break routines, searching for similarities and differences. Firstly we noted that in both the minimum directional and the non-directional type, the kinetic patterns have less variability. Secondly, we noted that during the complex break routines, sound and kinetic patterns are not so clearly aligned (the non-change of the movements of the two persons at the drop moment happens in a complex break routine). The behavior of the people is quite different to the previous description: in some break-downs people dance more than in build-ups. Its musical structure seems to be bodily unpredictable, but we also consider that what looks like a simple shift of the order of the steps, in fact implies a deeper change of the function of the steps. For example, some break-downs before the drop seem to work almost us a build-up, as if the lack of sounds builds up greater tension.

Conclusion

In this work, we found that there are sound-kinetic patterns that emerge as outcomes of the expressive alignment between musical patterns and dance movements. This alignments consists of the temporal coincidence of changes in sonic and kinetic patterns. The sonic change at the drop moment affords a more generalized and accurate change in human movements, while in the break-down and build-up sections the kinetic patterns change more slowly and gradually. This structure of the movement pattern is related to the acoustical kind of change of each break routine's steps: the drop is generated by a sudden change of some timbral features, and the break-down –especially those that only eliminate the lowest frequencies- also happen as a sudden but less strong change. In fact the beginning of the build-up is neither acoustically nor auditorily clear.

We identified some formal differences between break routines. However, we did not find particularly relevant links between break routines and the acoustical features and kinetic patterns of dance. Both these links and their status in the EDM culture require further study.

The general idea that people move similarly with the same music has no support in EDM dance. Each person makes movements which are defined differently in terms of their effort-shape elements from those made by other people at the same moment in time. The EDM dancers develop personal styles, which seem to be constructed over a limited repertoire of movements that they combine in several and creative ways. However, beyond these personal styles, people shared a kinetic pattern based on the change of the movement unit at the same time. These common changes are aligned with –and afford to- the acoustical changes during the break routine's steps.

References

Acouturier, J-J., Pachet, F., & Sandler, M. (2005). The Way It Sounds. Timbre Models for Analysis and Retrieval of Music Signals

Alluri, V. (2012). Acoustic, Neural, and Perceptual Correlates of Polyphonic Timbre (Doctoral Thesis). University of Jyväskylä: Jyväskylä.

Anzil, I. (2016). Sensación táctil y audiotáctil en la música. El caso de las músicas electrónicas utilizadas para el baile social en locales de baile de la Ciudad Autónoma de Buenos Aires y alrededores (Doctoral Thesis) [Tactil and Audiotactil sensation in music. The case of the electronic music used for social dance in Buenos Aires city]. Universidad Nacional de La Plata, La Plata.

Broughton, M. C. & Davidson, J. (2016). An Expressive Bodily Movement Repertoire for Marimba Performance, Revealed through Observers' Laban Effort-Shape Analyses, and Allied Musical Features: Two Case Studies. *Frontiers in Psychology, 7.*

Burger, B., Thompson, M., Luck, G., Saarikallio, S., & Toivianen, P. (2013). Influences of rhythm- and timbre- related musical features on characteristics of music-induced movement. *Frontiers in Psychology, 4*.

Chemero, A. (2003). An Outline of a Theory of Affordances. *Ecological psychology, 15*(2), 181-95.

Gallo, G. (2014). Libertades coreografiadas: palabras habladas, comunicaciones corporales y códigos en pistas dance de la ciudad de Buenos Aires. [Choreographed liberties: spoken words,

corporal communications and codes in dancefloors of Buenos Aires city]. *Estudios sociológicos 34*(100).
Hanslick, Eduard (1854). *The beautiful in music.* Londres: Novello and company.
Hartmann, M., Lartillot, O. & Toivianen, P. (2017). Interaction features for prediction of perceptual segmentation: Effects of musicianship and experimental task. *Journal of New Music Research, 46*(2), 156-164.
Kohn, D. & Eitan, Z. (2009). Musical Parameters and Children's Movement Responses. *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM)* (pp. 233-241). Jyväskylä, Finland
Krueger, J. (2013). Affordances and the musically extended mind. *Frontiers in Psychology, 4*.
Laban, R. (1950). *The Mastery of Movement*. Binsted: Dance Books.
Lartillot, O., Cereghetti, D., Eliard, K. & Grandjean, D. (2013). A Simple, high-yield Method for Assessing Structural Novelty. *Proceedings of the 3rd International Conference on Music & Emotion (ICME3)*. Jyväskylä, Finland.
Leman, M. (2007). *Embodied Music Cognition and Mediation Technology.* Cambridge, MA: MIT Press.
Leman, M. (2016). *The Expressive Moment. How Interaction (with Music) Shapes Human Empowerment*. Massachusetts: MIT Press.
Leman, M., Desmet, F., Styns, F., Van Noorden, L., & Moelants, D. (2009). Sharing Musical Expression Through Embodied Listening: A Case Study Based on Chinese Guqin Music. *Music Perception, 26(3)*, 263-278.
Len Faki [Boiler Room]. (2014, May 14). Len Faki Boiler Room Berlin DJ Set [Archivo de video]. Recuperado de https://www.youtube.com/watch?v=jQRI3b2SX8c
Maes, P-J., Dyck, E., Lesaffre, M., Leman, M., & Kroonenberg, P. (2014). The Coupling of Action and Perception in Musical Meaning Formation. *Music Perception. 32*(1), 67-84.
Marchiano, M., Martínez, I. C., & Damesón, J. (in press). Textural Layers and Polyphonic Timbre in Electronic Dance Music. *Proceedings of the International Conference of Students of Systematic Musicology (SysMus18)*.
Mastropietro, C. (2014). Música y timbre. El estudio de la Instrumentación desde los fenómenos tímbricos. [Music and timbre. The study of Instrumentation from Timbral Phenomena]. La Plata, Argentina: Ediciones Al Margen.
McAdams, S. (1999). Perspectives on the contribution of the timbre to musical structure. *Computer Music Journal, 23*(3), 85-102.
Solberg, R. T. (2014). "Waiting for the Bass to Drop": Correlations between Intense Emotional Experiences and Production Techniques in Build-up and Drop Sections of Electronic Dance Music. *Dancecult: Journal of Electronic Dance Music Culture 6*(1), 61–82.
Solberg, R. T. & Jensenius, A. R. (2016). Pleasurable and Intersubjectively Embodied Experiences of Electronic Dance Music. *Empirical Musicology Review, 11*(3-4), 301-318
Solberg, R. T. & Jensenius, A. R. (2017). Arm and head movements to musical passages of electronic dance music. *Book of Abstracts of 25th ESCOM* (pp. 175-176).

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Roll Over Beethoven: Uniform Information Density in Rock Music, Folk, and German Art Song

Peter Martens,[1] Joshua Albrecht,[2]

[1] *School of Music, Texas Tech University, United States*
[2] *Music Department, The University of Mary Hardin-Baylor, United States*

[1]peter.martens@ttu.edu, [2]jalbrecht@umhb.edu

Abstract

The Uniform Information Density hypothesis (UID) suggests that a presentational strategy of speakers and composers will be to maintain a consistent level of information entropy throughout a message. Temperley (2014) has recently suggested that composers may try to compensate for low information in one musical domain with high information in another domain to maintain listener interest. In this paper, we formally test this hypothesis in two studies in which we examine the information entropy of pitch and rhythmic successions in three corpora, including 95 rock songs, 1,408 folksongs, and 50 19th-century art songs by Schubert, Schumann, and Mendelssohn. Specifically, we tested the hypothesis that rhythmic and pitch entropies would be inversely correlated for event onsets as a way to maintain UID in these corpora. Contrary to our hypothesis, we found negative results, suggesting that composers intentionally punctuate a predictable musical background with highly unexpected events.

Introduction

In the field of information theory, "information" is a technical term, reflecting the level of probability of a message. Information entropy, measured in bits (binary digits), is the negative log of the probability of the message. Put more simply, the less likely a message is, the higher its information entropy. Completely certain messages have a probability of 1 and carry 0 bits of information, whereas completely impossible messages would (theoretically) have a probability of 0 and carry infinite bits.

Developed originally by Bell Labs to aid in the construction of the telecommunications infrastructure (Shannon 1948), information theory has proved useful in modeling various aspects of communication. Recently, Levy & Jaeger (2007) proposed a psycholinguistic theory of information entropy in which speakers wanting to maximize communicative effectiveness will tend to use a strategy of maintaining an even level of complexity throughout the process of their speech. Formally, the notion is one of *uniform information density* (UID), in which speakers maximize the amount of information communicated while simultaneously minimizing the cognitive load for listeners processing that information. According to Levy and Jaeger, this is accomplished by avoiding peaks or troughs in the amount of information per unit comprising the utterance. Because unlikely, low probability events take longer to process (Levy 2008), one way this manifests itself is in lengthened durations for words or syllables that are less predictable so as to decrease the listener's cognitive load of text processing for unpredictable events.

Several studies have provided evidence consistent with the UID theory. Bell *et. al* (2003) found an inverse correlation between word predictability and duration, and Aylett & Turk (2004) found the same relationship with syllables. Similarly, Jaeger (2010) found that speakers tend to add the low-information complementary word "that" preceding lower probability words. Genzel & Charniak (2002) found that less common words and word combinations occur more often later in a text, consistent with the notion that preceding context increases the probability of what follows, and that speakers counteract this tendency by using lower probability words.

Similarly, recent music cognition studies suggest that music may also be consistent with a UID strategy. Bartlette (2007) found that performers use longer performed durations on less expected harmonies as opposed to more expected harmonies. In a corpus study of thousands of classical music themes, Temperley (2014) found that repetitions of intervallic patterns used either larger intervals or more chromaticism than the earlier version of the pattern, consistent with the theory that lower probability events are added to counteract the lower information of repeated passages. These results are consistent with the notion that low probability in one domain will be compensated with high probability in another domain to maintain UID.

Hypothesis

This paper uses a corpus method to investigate the UID hypothesis that the information entropy of rhythmic elements of a melody will be inversely related to the information entropy of pitch elements. We predict a significant negative correlation between rhythm and melody. For example, we predict that composers will compensate for very complex, high entropy rhythmic successions by pairing them with low entropy pitch successions, and low probability pitch successions (like large leaps) will be paired with simpler, higher probability rhythms (like quarter notes). Formally, we test the following two hypotheses:

H1a: First-order pitch-class transition probabilities within a key (i.e. scale-degree successions) are negatively correlated with first-order transition probabilities for event onsets, relative to the metric grid.

H1b: First-order pitch transition probabilities without a key context (i.e. directed pitch intervals) are negatively correlated with first-order transition probabilities for event onsets, relative to the metric grid.

Study 1

Method

Corpora. If UID is a reflection of human cognitive constraints and not style dependent, then it should be observable in any collection of music. To test our hypothesis using different musical styles, three separate corpora were examined:

a) 200 rock songs from Rolling Stones "500 Greatest Songs of All Time" (de Clerq & Temperley 2011)

b) 6,215 folk songs from the Essen Folksong Collection (Schaffrath 1995)

c) 301 19th-century German art songs composed by Schubert, Schumann, and Mendelssohn (VanHandel 2005)

Population estimates of information entropy. For this study, we assumed that each of the three corpora would consist of different treatments of pitches and rhythms. For example, we assumed complex rhythmic syncopations would likely be more frequent in the Rolling Stone corpus than in the other two, while we assumed chromaticism would likely be more frequent in the 19th-c. art song corpus than the others. Assuming that listeners and composers of a style are familiar with and competent in that style, then they should incorporate appropriate population estimates of likelihood into their listening perceptions when they become aware of the style.

We therefore established population estimates of the likelihood of various rhythm and pitch successions for each separate database. We calculated three domains of probability:

a) First-order scale degree successions

b) Zeroth-order intervals

c) First-order event onsets, relative to the metric grid

We chose to use metric position rather than simple duration or inter-onset intervals for rhythmic events, because metric position incorporates inter-onset interval and duration, and is a cognitively salient element of listening. Two quarter-note durations are much more likely on beats one and two in a 4/4 measure than they are on the second eighth-note of beat four followed by the second eighth-note of beat one, and incorporating metric position allows us to differentiate these rhythms.

Sampling method. Different meters have different patterns of strong and weak beats, so for the purpose of this study, we chose to limit the collection to only those songs that were composed in 4/4. For any songs that used multiple time signatures, sections that were not in 4/4 were excluded. Some compound meters masqueraded as simple meters through the use of triplets. For the sake of simplicity, any onset pairs in which at least one of the onsets was not on one of the 16 sixteenth notes in a 4/4 measure were excluded. This simplification also eliminated thirty-second notes, resulting in 4 onsets per beat (16 total onsets per measure).

Another complication is the mode of the song. Different modes feature different patterns of stable and unstable scale degrees (Albrecht & Huron 2014). It is likely that listeners and composers are able to easily infer the mode of a song and so expect different scale degrees. Again, for the sake of simplicity, we elected to eliminate any minor key songs from the corpora. For the Rolling Stone corpus, several songs were written in different modes (e.g. Aeolian or Mixolydian), and several songs mixed different modes, perhaps with major tonic triads but minor mediant scale degrees in the melody (see de Clerq & Temperley 2011, Temperley 2012). For the Rolling Stone corpus, the two authors independently listened to each song and determined whether they were better categorized as 'basically major' or 'basically minor.' The lists were compared and any songs that were in the minor mode or in which there was disagreement between the authors were eliminated from further consideration. It should be noted that songs that modulated (much more common in 19) were taken in entirety with scale degrees related to the home key (see Martens & Albrecht 2018).

Once the reduction of songs was accomplished, there remained 95 songs from the Rolling Stones corpus (RS), 1,408 songs from the Essen Corpus (ES), and 50 songs from the 19th-c. German art song corpus (19). For each remaining song, each onset moment for the scale-degree, interval, and rhythmic onset was assigned an information entropy measured in bits based on the population estimates derived from the entire sample they were taken from.

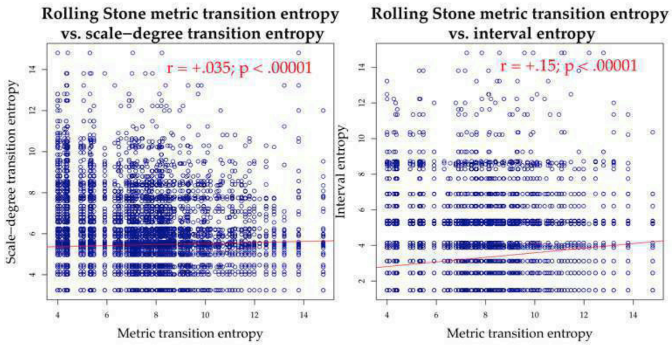


Figure 1. Correlation between metric entropy and scale-degree entropy (left), and rhythm entropy and interval entropy (right) for the Rolling Stone corpus.

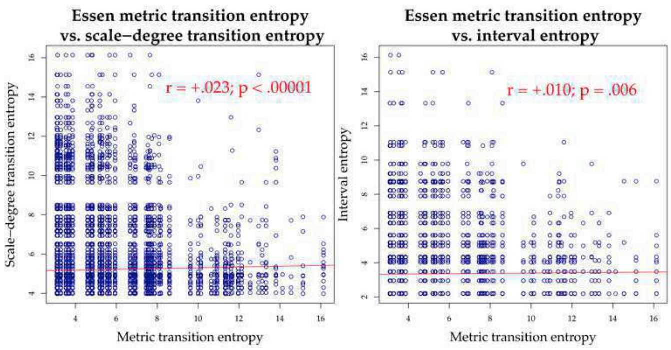


Figure 2. Correlation between metric entropy and scale-degree entropy (left), and rhythm entropy and interval entropy (right) for the Essen corpus.

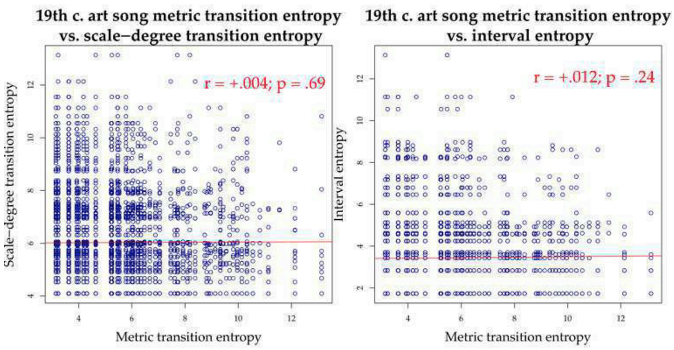


Figure 3. Correlation between metric entropy and scale-degree entropy (left), and rhythm entropy and interval entropy (right) for the 19th-century song corpus.

Results

Before examining the correlation between rhythm entropy and pitch entropy, as a check we first examined the correlation between interval and scale degree entropies. As expected, the correlations for all three corpora were significantly positive at

$p < .0001$ (RS, r = +.47; ES, r = +.41; 19, r = +.52), indicating that the scale-degree and interval approaches are not substantively different ways of looking at pitch movement.

Contrary to our hypothesis, the correlations between pitch entropy and rhythm entropy were *positive* for both the scale-degree and interval metrics for all three corpora (see Figures 1-3). It is important to stress that the effect sizes were very small (r of between +.004 to +.15), but given the sheer number of onset pairs, most of them were significant. The correlations for the 19th-century corpus were not significant, partly due to the small effect sizes and partly due to the much smaller sample sizes.

We failed to reject the null hypothesis for both H1a and H1b. These results are not consistent with a UID hypothesis. Rather than the hypothesized negative correlation, there is in most cases (with the exception of 19th-c.), a significant *positive* correlation, the direct opposite of our prediction. That is, rather than compensating for high entropy in the pitch domain with low entropy in the rhythm domain, composers across all three corpora seem to be positively correlating entropy between rhythm and pitch, albeit with a small effect. In short, Study 1 produced negative results.

Discussion

Interestingly, the correlation close to 0 by corpus obscures a wide range of correlations when broken down by song. A histogram of the correlations between scale-degree entropy and rhythm entropy broken down by song are displayed in Figure 4. Three plots are displayed, presented by corpus. Although a majority of songs employ entropy correlations between -.1 and +.1, a number of songs reveal correlations up to +.6 and down to -.6. These results may suggest that composers may somehow be choosing an information entropy strategy for particular compositions.

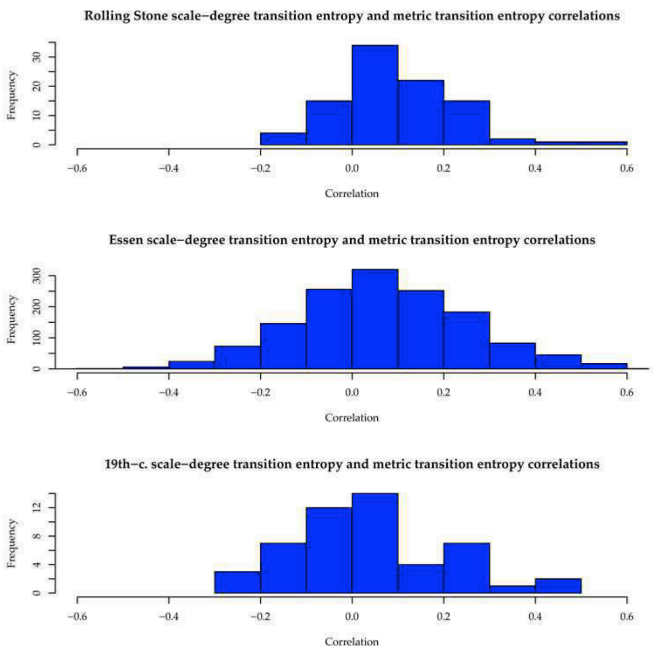


Figure 4. Correlations between scale-degree entropy and metric entropy by song, divided by corpus.

Another possible interpretation of the results is suggested by a close inspection of Figures 1-3 (especially Figures 2-3). Although there are many high entropy moments for pitch and rhythm individually, the upper-right corners of these figures are particularly sparse, meaning that there are very few moments that are high in entropy for pitch and rhythm simultaneously. On the contrary, most of the songs' onset moments are low in both pitch and rhythm entropy. The significant positive correlations of the overall plot may be a reflection of the large number of points in the lower left quadrants.

Our original hypothesis, that there would be an overall negative correlation between rhythm and pitch entropy, actually consists of two assumptions:

1. As melodies use *less expected* pitch successions (like big leaps or dissonances), they will use *simpler* rhythms to *enhance perceptibility*
2. As melodies use *more predictable* pitch successions, they will use *more complex* rhythms to *increase interest*

It is possible that these two assumptions are not equally likely. Of the two, the second assumption seems more problematic. After all, because we only took into consideration pure melody, we neglected to consider other domains that might increase interest. It is possible that compensatory interest can be attained from the text, texture, instrumentation, dynamics, and other domains. It is also possible that a large number of low entropy moments may be acceptable without the danger of boring listeners as long as there are occasional high entropy moments to maintain interest. It is possible that the large number of low entropy moments is suggesting a positive correlation that may be spurious to those high entropy moments. To test just assumption 1, a follow-up study was conducted that eliminated low entropy moments, examining only the high entropy moments.

Study 2

Method

Sampling method. We used the same sample from Study 1, limited to only 'high entropy' moments. We therefore examined only onset moments in which information entropy was more than 2 standard deviations above the mean in any of the three domains (interval, scale degree, or metric position) in each corpus.

Hypothesis. For this follow-up study, we used modified versions of our original hypotheses:

H2a: First-order pitch-class transition probabilities within a key (i.e. scale-degree successions) in which information entropy is more than 2 sd above the mean for its corpus are negatively correlated with first-order transition probabilities for event onsets, relative to the metric grid.

H2b: First-order pitch transition probabilities without a key context (i.e. directed pitch intervals) in which information entropy is more than 2 sd above the mean for its corpus are negatively correlated with first-order transition probabilities for event onsets, relative to the metric grid.

H2c: First-order transition probabilities for event onsets, relative to the metric grid, in which information entropy is more than 2 sd above the mean for its corpus are negatively correlated with first-order pitch transition probabilities within a key (i.e. scale degree successions).

H2d: First-order transition probabilities for event onsets, relative to the metric grid, in which information entropy is more than 2 sd above the mean for its corpus are negatively correlated with first-order pitch transition probabilities without a key context (i.e. directed pitch intervals)

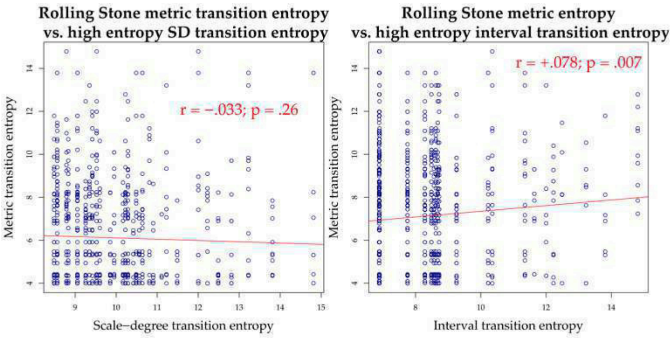


Figure 5. Correlation between high entropy scale-degree entropy and metric entropy (left), and high entropy interval entropy and metric entropy (right) for the Rolling Stone corpus.

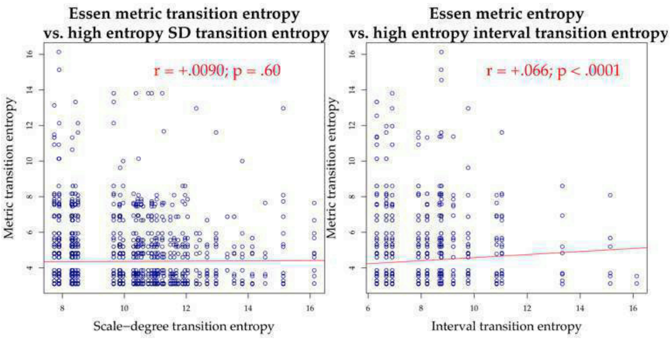


Figure 6. Correlation between high entropy scale-degree entropy and metric entropy (left), and high entropy interval entropy and metric entropy (right) for the Essen corpus.

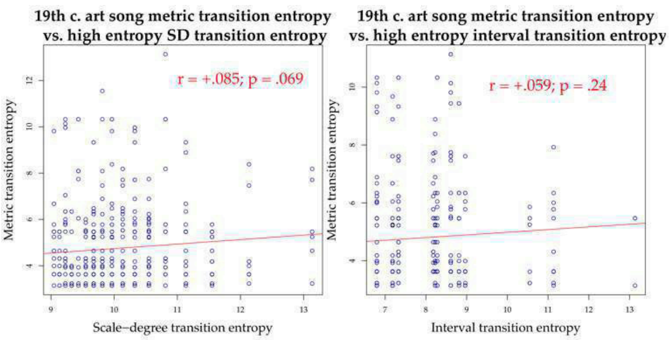


Figure 7. Correlation between high entropy scale-degree entropy and metric entropy (left), and high entropy interval entropy and metric entropy (right) for the 19th-century corpus.

Results

Correlations between high entropy pitch moments and their matched metric onset entropy moments for each of the three corpora appear in Figures 5-7 (RS, ES, & 19) respectively. High entropy scale-degree transitions appear on the left of each figure and high entropy interval transitions appear on the right. All pitch information entropies are above 6 bits. Again, contrary to our hypotheses H2a and H2b, none of the correlations are significantly negatively correlated with metric entropy. In fact, as with H1a and H1b, some of the correlations (2 out of 6) are significantly *positively* correlated, contrary to our hypothesis.

Finally, correlations between high entropy metric onsets and their matched pitch entropy moments for each of the three corpora appear in Figures 8-10 (RS, ES, & 19) respectively. Correlations with high entropy scale-degree transitions appear on the left of each figure and correlations with high entropy interval transitions appear on the right. All metric information entropies are above 7 bits. These results are a little more complicated than earlier results. The Rolling Stone corpus demonstrates significant negative correlations, finally *consistent with* our hypothesis, although the effect sizes are still small (-.098 and -.11). One other correlation was negative, though not significant, contrary to our hypothesis. There is, however, one significantly positive correlation and two other positive correlation that is not significant.

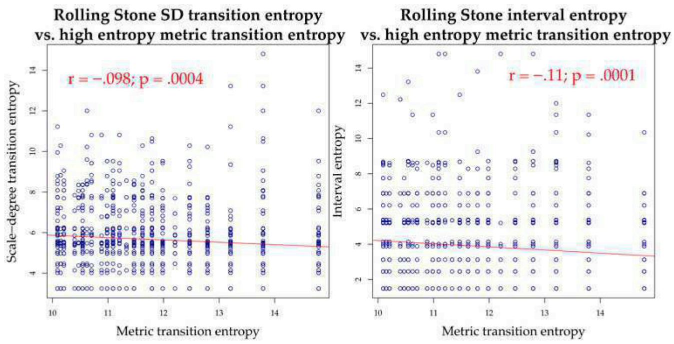


Figure 8. Correlation between high entropy metric entropy and scale-degree entropy (left), and high entropy metric entropy and interval entropy (right) for the Rolling Stone corpus.

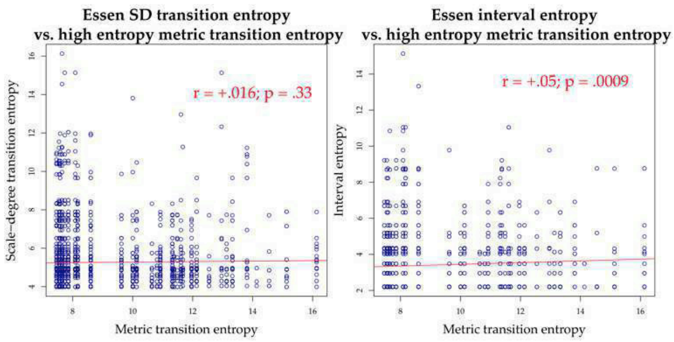


Figure 9. Correlation between high entropy metric entropy and scale-degree entropy (left), and high entropy metric entropy and interval entropy (right) for the Essen corpus.

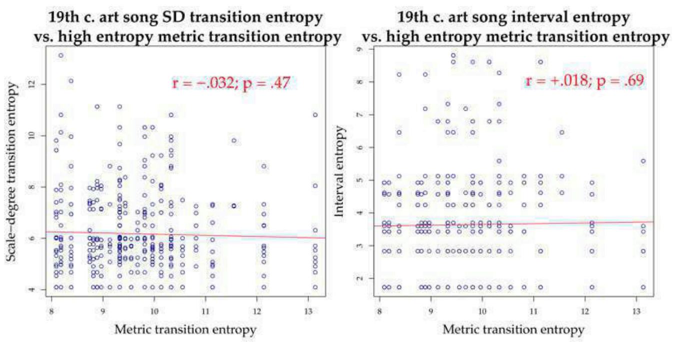


Figure 10. Correlation between high entropy metric entropy and scale-degree entropy (left), and high entropy metric entropy and interval entropy (right) for the 19th-century corpus.

Discussion

Like Study 1, Study 2 did not produce consistent results that aligned with the UID hypothesis. H2a and H2b were all skewed in a contrary direction to the hypothesis, with two of six correlations being significantly positive with small positive correlations. H2c and H2d results were mixed, with 25% of cases consistent with these hypotheses, 25% significantly contrary to them, and 50% of cases not significant either direction. Importantly, even the significant correlations were all small (-.11, -.098, +.016, and +.05). It is possible that the Rolling Stone corpus is actually structured differently than the other databases, with highly unlikely rhythmic events paired with significantly more likely pitch events. Of note is that all of the rhythmic onsets in the RS database were above 10 bits of information, a fairly high degree of entropy. However, with the tiny effect sizes and mixed results, we feel compelled to report overall negative results for Study 2. These results are consistent with spurious effects caused by noise in the data.

Conclusion and General Discussion

Contrary to the UID hypothesis, there appears to be a weak, though statistically significant *positive* correlation between rhythmic and pitch information entropy. This seems to be a small but real effect. In other words, in these melodies composers seem to be correlating entropies, at least in the pitch and rhythm domains. Rather than a consistent density of information, composers seem to be punctuating an otherwise mostly predictable musical background with short, highly unpredictable foregrounded moments.

One possible reason this compositional strategy may be different from the UID seen in speech could be differences between the modes of communication. Rather than communicating information, music is mostly expressive. Huron (2006) has documented the way that composers design their music to avoid habituation by grabbing listeners' attention through orienting responses. It is possible that these short, unpredictable events against a background of predictable music may serve as a hook to promote orienting responses in listeners.

It is also possible that methodological issues with our research design have produced false negative results. One possible difficulty is that we only examined pitch and rhythm domains. Of course, there is much more to these works, especially the piano part of the 19th-century art song repertoire and the instrumentation, texture, dynamics, and production effects of the Rolling Stone corpus. A more consistent UID strategy might be observable with these factors. Perhaps the operationalization of rhythmic succession was also an issue, as we did not consider phrasing boundaries or rests between onsets. Finally, rather than considering events as probabilistic within a large collection of onsets in a corpus, we might find different results if we were look at conditional probabilities. That is, although 3.75-4 is relatively improbable in the Essen database, and so would receive a high entropy in our approach, when hearing an attack on beat 3.75, a listener can be sure that a folksong will proceed to an attack on beat 4. In fact, it is the only possible consequent in this corpus, leading a very low probability if looking at dynamic expectations.

Future research plans include looking at the difference between conditional probabilities and schematic probabilities. We also plan to look at complete phrases information density, rather than individual moments. We also plan to look for differences in entropy levels between verses and choruses in the RS database. We also plan to look at the 'rhythm' of high entropy moments, or the speed with which these events occur. It is possible this may map onto an orienting response strategy, in which we might be able to predict the next high entropy moment based on how long it has been since the last one.

Acknowledgements. We would like to thank Daniel Dudney for his early work on this project and for his insightful analyses of entropy levels in rhythms and pitches in popular music. We would also like to thank our respective institutions for their support of travel to ICMPC via a UMHB faculty development grant and a TTU Talkington College of Visual and Performing Arts international travel grant.

References

Albrecht, J., & Huron, D. (2014). A statistical approach to tracing the historical development of major and minor pitch distributions, 1400-1750. *Music Perception, 31* (3), 223-243.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47* (1), 31-56.

Bartlette, C. (2007). A study of harmonic distance and its role in musical performance. (Unpublished doctoral dissertation). University of Rochester, Rochester, N.Y.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girard, C., Gregory, M., and Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113* (2), 1001-1024.

de Clerq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music, 30*, 47-70.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 199-206). Stroudsburg, PA: Association for Computational Linguistics.

Huron, D., (2006). *Sweet anticipation*: *Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Jaeger, F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61* (1), 23-62.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106* (3), 1126-1177.

Levy, R., & Jaeger, F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölhopf, J. Platt, & T. Hofmann (Eds.), *Advances in Neural Information Processing Systems* (pp. 849-856). Cambridge, MA: MIT Press.

Martens, P., & Albrecht, J. (2018). Pitch and rhythmic succession in German folksong, 19th-century art song, and Classic Rock melodies. In R. Parncutt (Ed.), *Proceedings of the Fifteenth International Conference on Music Perception and Cognition*. Graz, Austria: University of Graz.

Schaffrath, H. (1995). *The Essen folksong collection.* D. Huron (Ed.). Stanford, CA: Center for Computer Assisted Research in the Humanities.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27* (3), 379-423.

Temperley, D. (2014). Information flow and repetition in music. *Journal of Music Theory, 58,* 155-178.

Temperley, D. (2012). Scalar shift in popular music. *Music Theory Online, 17*.4.

Temperley, D., & Gildea, D. (2015). Information density and syntactic repetition. *Cognitive Science, 39,* 1802-1823.

VanHandel, L. Setting a menu to music: Prosody and melody in 19th c. art song. (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Pitch and Rhythmic Succession in German folksong, 19th-century art song, and classic rock melodies

Peter Martens,[1] Joshua Albrecht,[2]

[1] *School of Music, Texas Tech University, United States*
[2] *Music Department, The University of Mary Hardin-Baylor, United States*

[1]peter.martens@ttu.edu, [2]jalbrecht@umhb.edu

Abstract

As an increasing amount of music from diverse genres and epochs is encoded, the opportunity arises to identify and compare salient melodic characteristics in a variety of styles across time and place. Using European folksong from the Essen collection (Schaffrath, 1995), French and German art song from the 19th century (VanHandel, 2005), and rock melodies from the Rolling Stone 500 Greatest Songs of All Time (de Clerq & Temperley 2011), we hypothesized that melodies from these historically distinct corpora would correlate significantly within the pitch-based parameters of interval content and first-order scale degree succession, but not in rhythmic content. Further, we hypothesized that more detailed comparison between the corpora would reveal salient stylistic differences that might reflect more complex relationships between melodic styles due to chronological evolution, high art vs. low art, the relationship of music to text, and the influence of non-western musical elements. While each melodic variable did show a significant main effect of corpus membership, pairwise comparisons were illustrative. For example, the rhythmic content of Essen and 19th-c. melodies correlated significantly with each other but not with Rolling Stone melodies, while Rolling Stone and 19th-c. melodies correlated more strongly with respect to intervallic content, possibly reflecting a common text-setting strategy less used in folksong.

Introduction

Foundational work in encoding melodic corpora took a significant leap forward with the Essen Associative Code and Folksong Database (Schaffrath, 1995), which provided researchers with thousands of examples from European and other song traditions. Since that time, numerous other melodic databases have been created, generally based on epoch and genre. Two collections of note are the database of European art song from the 19th century (VanHandel, 2005), and rock melodies from the Rolling Stone 500 Greatest Songs of All Time (de Clerq & Temperley, 2011). All three collections have already given rise to respective scholarly insight (e.g. Huron 2006, VanHandel 2009, Temperley & de Clerq 2013); this study combines these three corpora to look for insights arising from emergent properties of all three as a whole, while also comparing the datasets against one another in order to discover possible low-level statistical properties that define or set off the corpora as distinct from one another.

Hypothesis

This paper uses a corpus method to compare pitch- and rhythm-based melodic properties between corpora. Two pitch-based approaches were used, 1) first-order pitch-class transition probabilities within a key, i.e. scale-degree successions, and 2) first-order pitch transition probabilities without a key context, i.e. directed pitch intervals. In the domain of rhythm, first-order transition probabilities for event onsets were used, measured relative to the underlying metric grid. We explored the following general hypotheses.

H1: Given that melodies from all three historically distinct corpora are diatonic in nature, they will correlate significantly within the pitch-based parameters.

H2: Given that rock from the 20th century incorporates rhythms from cultures outside the western tradition of the Essen and 19th-c. melodies, the rhythmic content of Rolling Stone melodies will not correlate strongly with the rhythms of the earlier styles.

Study 1

Method

Corpora. Three separate corpora were examined:

a) 200 rock songs from Rolling Stones “500 Greatest Songs of All Time” (de Clerq & Temperley 2011)

b) 6,215 folk songs from the Essen Folksong Collection (Schaffrath 1995)

c) 301 19th-century European art songs (VanHandel 2005)

We established population estimates of the likelihood of various rhythm and pitch successions for each separate database, calculating three domains of probability:

a) First-order scale degree successions
b) Zeroth-order intervals
c) First-order event onsets, relative to the metric grid

We chose to use metric position rather than simple duration or inter-onset intervals for rhythmic events, because metric position incorporates inter-onset interval and duration, and is a cognitively salient element of listening. Two quarter-note durations are much more likely on beats one and two in a 4/4 measure than they are on the second eighth-note of beat four followed by the second eighth-note of beat one, and incorporating metric position allows us to differentiate these rhythms.

Sampling method. Different meters have different patterns of strong and weak beats, so for the purpose of this study, we chose to limit the collection to only those songs that were composed in 4/4. Any songs that used multiple time signatures were excluded. Some compound meters masqueraded as simple meters through the use of triplets. For the sake of simplicity, any onset pairs in which at least one of the onsets was not on one of the 16 sixteenth notes in a 4/4 measure were excluded. This simplification also eliminated thirty-second notes, resulting in 4 onsets per beat (16 total onsets per measure), referred to below as metric positions 1, 1.25, 1.5, 1.75, 2, 2.25, etc.

Another complication is the mode of the melody. Different modes feature different patterns of stable and unstable scale degrees (Albrecht & Huron 2014). It is likely that listeners and composers are able to easily infer the mode of a song and so expect different scale degrees or intervals. Again, for the sake of simplicity, we elected to eliminate any minor key songs from the corpora. For the Rolling Stone corpus, several songs were written in different modes (e.g. Aeolian or Mixolydian), and several songs mixed different modes, perhaps with major tonic triads but minor mediant scale degrees in the melody (see de Clerq & Temperley 2011, Temperley 2012). The two authors independently listened to each song and determined whether they were better categorized as 'basically major' or 'basically minor.' The lists were compared and any songs that were in the minor mode or in which there was disagreement were eliminated from further consideration in the present study. It should be noted that songs that modulated (much more common in 19th-c. melodies) were taken in entirety with scale degrees related to the home key, and may figure prominently in some of the results reported below.

Once the reduction of songs was accomplished, there remained 95 songs from the Rolling Stones corpus (hereafter RS), 1,408 songs from the Essen Corpus (hereafter ES), and 50 songs from the 19th-c. corpus, all by Mendelssohn, Schubert, or Schumann (hereafter 19). For each remaining song, each onset moment for the scale-degree, interval, and rhythmic onset was assigned a probability based on the population estimates derived from the entire sample from which they were taken.

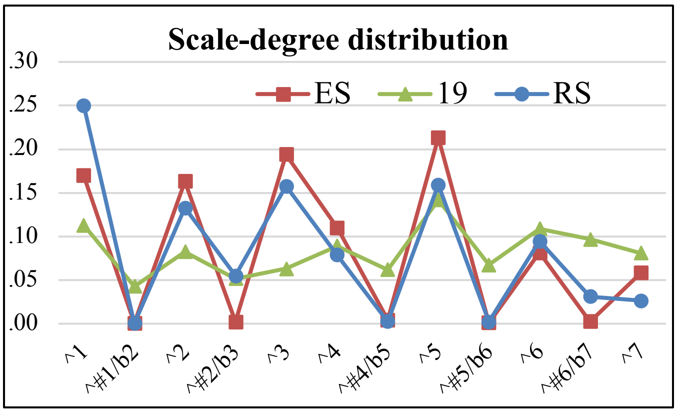


Figure 1. Zeroth-order scale-degree distribution in the three corpora.

Results

Scale Degree. Figure 1 shows the overall zeroth-order distribution of scale degrees in the three corpora. Melodies from 19 show a rather flat distribution, while ES and RS melodies are strikingly similar. Even in this view, however, RS melodies show traces of Dorian relative to ES, with more ^b3 and ^b7 and less ^7 (all $p < .0002$). We might attribute 19's flat distribution to greater use of surface chromaticism in these melodies. But consider Table 1, which lists the 16 most common scale degree successions from each corpus. Looking at first-order scale degree successions in these 19th-c. melodies shown in the middle columns of the Table, we see that a repetition of ^6 is the second most common succession; though not in the top 16, the successions that make up ^3-^#4-^#5-^6 are also surprisingly common. These successions are well at home in the melodies' relative minor key areas, in which they function as the typical scale degree successions ^1-^1 and ^5-^6-^7-^8. Other examples of this phenomenon in 19th-c. melodies from Table 1 include scale degree repetitions and connections indigenous to the key of the subdominant (^1-^b7, ^6-^5-^4). Thus the seeming chromatic surface of 19th-c. melodies conveyed by Figure 1 are more likely due to a chromatic middleground of tonicizations and key changes, rather than to the chromatic ornamentation of a single-key melody.

Two other characteristics provide interesting insights into the typical sounds of the melodies from each corpus. Orientation on the tonic (^1) varies markedly, with 20% of 19 transitions involving the tonic pitch while in ES it is 31% and RS 40%. Note repetition also varies widely, with RS melodies being most repetitive (31% of transitions) and not significantly different than 19 (28% of transitions). ES melodies contain significantly less note repetition than both 19 and RS, at 22% (both $p < .0002$). This feature will be revisited below with respect to interval usage.

Table 1. The 16 most probable scale-degree transitions in Essen (ES), 19th-century (19) and Rolling Stone (RS) corpora.

ES		19		RS	
trans.	prob.	trans.	prob.	trans.	prob.
^3-^2	0.0628	^5-^5	0.0458	^1-^1	0.1036
^5-^5	0.0626	^6-^6	0.0309	^2-^1	0.0609
^2-^1	0.0526	^1-^1	0.0305	^5-^5	0.0588
^3-^4	0.0524	^b7-^b7	0.0272	^3-^2	0.0467
^5-^4	0.0458	^7-^7	0.0257	^3-^3	0.0467
^1-^1	0.0429	^4-^4	0.0257	^1-^2	0.0320
^6-^5	0.0422	^5-^4	0.0219	^6-^5	0.0314
^3-^3	0.0397	^b6-^b6	0.0199	^2-^2	0.0306
^2-^2	0.0356	^1-^b7	0.0189	^1-^6	0.0287
^1-^2	0.0347	^2-^2	0.0186	^4-^3	0.0266
^5-^3	0.0344	^6-^5	0.0185	^5-^6	0.0234
^3-^5	0.0342	^2-^1	0.0181	^2-^3	0.0244
^5-^1	0.0333	^3-^3	0.0177	^5-^3	0.0241
^1-^7	0.0305	^#4-#^4	0.0176	^1-^3	0.0224
^3-^4	0.0296	^4-^5	0.0157	^4-^4	0.0221
^1-^3	0.0253	^1-^5	0.0157	^5-^4	0.0221

Despite the note-repetition link between RS and 19, overall pairwise correlations are uniformly positive and robust but with ES and RS showing the strongest similarity (ES-19, r = +.53; ES-RS, r = +.82; 19-RS, r = +.63). Further, the differences between these correlations bear out that ES-19 and 19-RS correlations are significantly less strong than the ES-RS correlation, and they do not differ significantly from one another (ES-19 v. ES-RS, $p < .0001$; 19-RS v. ES-RS, $p = .0003$, ES-19 v. 19-RS, ns). Thus we see a picture of RS and

ES melodies being broadly similar in terms of scale degree succession, with 19 melodies behaving differently.

Table 2. Probabilities of directed interval occurrence in Essen (ES), 19th-century (19) and Rolling Stone (RS) corpora.

int size	ES	19	RS
-12	0.0014	0.0021	0.0026
-11	0.0000	0.0001	0.0002
-10	0.0004	0.0016	0.0008
-9	0.0022	0.0026	0.0027
-8	0.0030	0.0056	0.0017
-7	0.0094	0.0171	0.0081
-6	0.0011	0.0036	0.0027
-5	0.0201	0.0307	0.0248
-4	0.0277	0.0323	0.0239
-3	0.0584	0.0418	0.0667
-2	0.2206	0.1526	0.1887
-1	0.0880	0.0866	0.0575
0	0.2162	0.2968	0.3037
1	0.0595	0.0773	0.0364
2	0.1293	0.0938	0.1144
3	0.0511	0.0414	0.0621
4	0.0324	0.0280	0.0273
5	0.0495	0.0501	0.0260
6	0.0005	0.0028	0.0022
7	0.0126	0.0114	0.0146
8	0.0039	0.0061	0.0030
9	0.0081	0.0095	0.0089
10	0.0021	0.0020	0.0025
11	0.0001	0.0006	0.0004
12	0.0023	0.0024	0.0051

Directed Interval. One of the reasons we approached pitch also with directed intervals was to eliminate the influence of global tonic just discussed. That is, if melodies in 19 are generally diatonic but contain shifting key centers, then their intervallic usage should resemble that of the Essen and Rolling Stone melodies more closely than they did in scale degree usage. Of course, shifts to minor keys within a globally major melody would presumably include commensurate differences in interval usage, but this should be a relatively small effect (cf. Huron 2006, Ch. 9).

Table 2 shows interval probabilities in the three corpora from descending octave (-12 semitones) to ascending octave (+12 semitones). Intervals outside this range did exist but were exceedingly rare (no more than .01% in any corpus) and so are omitted from the figure and subsequent analysis for clarity. Given that orientation to an overall tonic is not a factor in this analysis, pairwise correlations were not surprisingly even stronger and more uniform than in the scale degree comparison (ES-19, $r = +.87$; ES-RS, $r = +.91$; 19-RS, $r = +.95$).

Note the higher probability of unisons in RS and 19 relative to ES, highlighted in the Table (both differences $p < .0001$, and $d > .70$), which is one basis for the strong interval use correlation between 19 and RS and which mirrors their higher incidence of scale degree repetition discussed above. A question must be asked, however: do we as listeners apprehend first-order melodic unisons, i.e. note repetitions, in the same way we do other intervals, or might they be essentially ignored since they contain no new pitch information? Are they perceptual non-intervals?

Speculating that unisons are less salient perceptually than other intervals, we further speculated that the reason for this difference in note repetition between corpora arises from different priorities in text setting, rather than from intentional interval usage on the part of the composers. We therefore recalculated interval use ignoring unisons (Table 3, below).

Table 3. Probabilities of directed interval occurrence in three corpora, unisons omitted.

int size	ES	19	RS
-12	0.0018	0.0030	0.0039
-11	0.0000	0.0000	0.0003
-10	0.0005	0.0022	0.0012
-9	0.0028	0.0036	0.0039
-8	0.0038	0.0082	0.0024
-7	0.0120	0.0266	0.0118
-6	0.0014	0.0051	0.0039
-5	0.0257	0.0439	0.0363
-4	0.0354	0.0456	0.0350
-3	0.0745	0.0593	0.0976
-2	0.2815	0.2141	0.2762
-1	0.1122	0.1242	0.0842
1	0.0759	0.1088	0.0533
2	0.1650	0.1327	0.1674
3	0.0652	0.0596	0.0909
4	0.0413	0.0409	0.0399
5	0.0632	0.0733	0.0381
6	0.0006	0.0042	0.0033
7	0.0161	0.0165	0.0213
8	0.0049	0.0087	0.0043
9	0.0103	0.0128	0.0130
10	0.0027	0.0027	0.0036
11	0.0001	0.0008	0.0006
12	0.0029	0.0034	0.0075

Without unisons, what comes to the fore in this Table is the plurality of descending steps (intervals -2 and -1) in all three corpora, comprising over a third of all intervals used in each. Further emphases emerge, including the chromatic character of 19 melodies, which can be seen in its significantly greater usage of ascending half steps than ES melodies (+1, $p = .009$), with RS melodies providing an even greater half-step contrast (both +1 and -1, $p < .0001$). In a similar vein, the pentatonic tendencies of RS relative to 19 melodies are clear from significantly greater probability in RS of minor 3rds (-3 and +3, $p < .0001$ and $p = .0004$) and major 2nds (-2 and +2, $p = .0016$ and $p = .015$).

Recall that, with unisons considered, RS and 19 were most strongly correlated. Multiple regression analysis shows that interval usage in RS predicted interval usage in 19 at $p < .0001$ and ES at $p = .006$, with 19 and ES being poor predictors of each other's intervals. Without unisons, however, ES becomes the central node, predicting interval usage in both 19 and RS at $p < .0001$, with RS and 19 being significant *negative* predictors of each other's intervals at $p = .0037$.

Table 4. Probabilities of rhythmic transitions in three corpora beginning on a beat (onsets 1, 2, 3, or 4), on a beat division (onsets *x*.5), or on a beat subdivision (onsets *x*.25 or *x*.75).

transition starts on	ES	19	RS
beat	0.76	0.67	0.45
division	0.22	0.27	0.42
subdivision	0.02	0.06	0.12

Rhythmic transitions. A clear stylistic difference emerges here, with significantly stronger correlations between ES and 19 than between either and RS. (ES-19 v. ES-RS, $p < .0001$; 19-RS v. ES-RS, $p = .0003$, ES-19 v. 19-RS, ns). A simple way to account for these differences is that RS melodies contain more activity at the 8th- and 16th-note levels. As shown in Table 4, RS melodies are indeed nearly as likely to have a transition begin on a beat division as on a beat, and far more likely to begin on a beat subdivision. Only one pairwise comparison between RS and the other two corpora that is not significant at $p \leq .01$ is the RS and 19 beat comparison. By contrast, no difference between ES and 19 is significant. Unlike the pitch parameters, rhythmic usage appears to trend historically in one direction, toward greater emphasis on beat divisions and subdivisions. Of course, Table 4 could simply capture the evolution of notational use (actual or assumed) across the centuries, but these data do support commonly-held ideas about the increase in syncopation, etc. in Western music during the 19th and 20th centuries.

Figure 2 below shows the relationship of rhythmic usage across the three corpora in more detail. For each possible event that could begin a rhythmic transition (the *antecedent* event), pairwise comparisons of the immediately subsequent event (the *consequent* event) were made. These data points answer questions such as "given an antecedent on beat 1.5, how closely do 19 and RS correspond with respect to the consequents?"

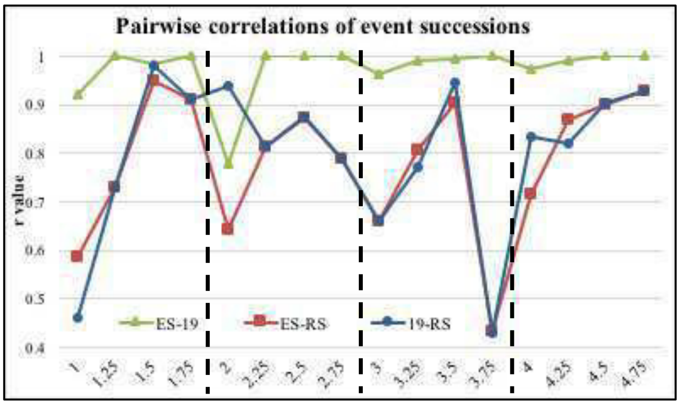


Figure 2. Pairwise correlations by event successions across the three corpora.

The general correlation between ES and 19 evident from Table 3 can be seen in Figure 2 in that their *r*-values are at or near 1.0 for every attack point except beat 2. Rhythmic usage in RS diverges from ES and 19 most notably on beats 1, 3, and 3.75. Beat 3.75 is particularly interesting in RS melodies, since both 4.25 and 4.5 are more probable consequent attacks than beat 4. In ES and 19, 3.75-4.25 occurs zero times, while 3.75-4.5 occurs once (in a folksong). That this moment in Rock melodies is so markedly prone to a syncopated continuation warrants further study.

Returning to beat 2, it is intriguing as the sole instance of RS correlating with 19 more closely than does Essen. Where the typical ES melody follows an attack on beat 2 with an attack on beat 3, 19 and RS melodies follow 2 with 2.5 much more commonly. This difference also emerges from the overall transition probabilities, from which we derive the prototypical 4/4 rhythms in each corpus:

ES ||: ♩♩♩♩ :|| 19 ||: ♩ ♫♩ ♩ :|| RS ||: ♫♫♫♫ :||

Conclusion

These three musical parameters paint a rich and varied picture of these three significant melodic styles in western music. In the pitch-based parameters, traditional European folksong and 20th-century Rock melodies seem most closely related, with the one area of notable similarity between 19 and RS likely pointing to similar strategies in the setting of text, in contrast to those used in composing folk melodies. These pitch-based findings thus provide mixed support for our initial hypothesis. In the domain of rhythm, Rock melodies do indeed stand out, providing clear support for our second hypothesis.

Acknowledgements. We would like to thank our respective institutions for their support of travel to ICMPC via a UMHB faculty development grant and a TTU Talkington College of Visual and Performing Arts international travel grant.

References

Albrecht, J., & Huron, D. (2014). A statistical approach to tracing the historical development of major and minor pitch distributions, 1400-1750. *Music Perception, 31* (3), 223-243.

de Clerq, T., & Temperley, D. (2011). A corpus analysis of rock harmony. *Popular Music, 30*, 47-70.

Huron, D., (2006). *Sweet anticipation*: *Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Schaffrath, H. (1995). In: D. Huron (Ed.), *The Essen Folksong Collection*. Stanford, CA: Center for Computer-Assisted Research in the Humanities.

Temperley, D. (2012). Scalar shift in popular music. *Music Theory Online, 17*.4.

Temperley, D., & de Clerq, T. (2013). Statistical Analysis of Harmony and Melody in Rock Music. *Journal of New Music Research, 42,* 187-204.

VanHandel, L. (2005). Setting a Menu to Music: Prosody and Melody in 19th c. Art Song. (Unpublished doctoral dissertation). Stanford University, Standford, California.

VanHandel, L. (2009). National Metrical Types in Nineteenth Century Art Song. Empirical Musicology Review, 4, 4, 134-145

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Relationship of Musical Life and Music Studies: Science or Fiction?

Irene Martínez Cantero[1]

[1] *Professional Conservatory of Music, Pilar de la Horadada, Alacant, Spain*
Social Psychology and Methodology, Autónoma University, Madrid, Spain

[1]irenemc2000@yahoo.com

Abstract

Perception and cognition of music are two aspects closely related to emotion and motivation towards the music activity. Their importance has been verified by numerous disciplines that provide a part of their global knowledge: Aesthetics (Collier, 2007), Musicology (Shaw, 2001), Morals (Fairlay, 2006), Politics and Economics (Fouce, & Pecourt, 2008), or Sexuality (Trotta, 2009). However, Social Psychology and Neuroscience are, perhaps, the nearest (Hargreaves, Hargreaves, & North, 2012; Sloboda, 2012; Ockelford, 2012). In this communication is presented an introduction of state of the matter in Music Neuro-Psychology. From these disciplines, a previous research and empirical data in different student cases on listening and music activity is also provided. These cases are different from each other case studies, but present similar results in relation to student musical real life. Main objective analyzes and compares the role of music in different age students in six case studies of different educational systems from southeastern Spain. The results make it possible to discern common areas in which the informal environment in perception, emotion, motivation and musical cognition is revealing. From them is extracted some general conclusions and contributions to the improvement of the musical educational practice in any of its levels, in which student's musical life should have a greater presence.

Introduction

Functional magnetic resonance (fMRI) has been prove pleasant music has more capacity to produce the entrainment effect (Trost et al., 2014). In addition, movement regularity and its relations with emotion (Ferrucci & Prior, 2014) are linked with emotional reactions to music and musical instruments interpretation (Levitin, 2008). Although basic emotions are related to movement, the most complex emotions emerge through emotional and motivational systems activation (Levitin, 2008).

Pleasure experienced with music causes release of neurotransmitters, dopamine and serotonin, which influence learning, reward (Altenmüller & Schlaug, 2013) and satisfaction feelings with expected results (Evers, & Suhr, 2002). In addition, these neurotransmitters are capable of modifying or modeling human behavior (Levitin, 2008). Blood and Zatorre (2001) demonstrated through PET intense emotion in musical activities are related to reward process, motivation and excitement. Menon and Levitin (2005), likewise, that a higher dopamine levels, causes a greater positive mood and affectivity, so it influences reward and reinforcement.

Significant live event, such as listening to music or making music, is associated with certain emotions and are all remembered together (Ohberg et al., 2004). However, music in the children lives and music in their specialized studies often do not respond to equally intense channels, since the events and memories associated with them may be different. Thus, Boal-Palheiros & Hargreaves (2011) found differences in children and adolescents perceptions of music in children and adolescents in different situations, such as school and family. From all this, it turns out that excitement and motivation in the memory of musical studies will depend to a large extent on the importance that music has in the same student in their different musical environments (Martínez, 2017).

Background: previous researches

Study presented is part of a larger research that began in 2004 and is part of a doctoral thesis focused on motivations and environments of extracurricular music students. Several articles (Martínez, & Jauset-Berrocal, 2017; Jauset-Berrocal, Martínez, & Añaños, 2017), communications in congresses (Martínez, Casas-Mas, & Montero, 2016; Martínez, Montero, & Casas-Mas, 2016; Martínez, 2016, Martínez, 2017) have partially explored results linked to this proposal in relation to initial motivation of students, teacher preconceptions about their pupils, cognition, emotion, etc. However, both mode of analysis and proposed objectives, participants or the contents and information analysis modes are different in this text.

Objectives

In this research, the three participants groups with 10 children from 8 to 16 years old are opposites in performance. The first group consists of 10 students considered to be of higher performance because they belong to the Youth Orchestra of Alicante. The second group, with 10 students considered average for not having made with them any previous selection in relation to their performance. And the third group, are 10 students that dropped out extracurricular music studies. Analysis has focused on two different contexts: musical studies and students life. Likewise, these groups have been differentiated and studied by age, and has also been linked to the following objectives or thematic areas of study: 1. Expand knowledge about perception, cognition, motivation and emotion of music and musical studies in the three groups, ages and contexts. 2. To analyze partial results and linkages, differences or similarities in all the items raised: ages, contexts and student profiles. 3. Arrive at comparable conclusions with literature found closest to these topics.

Methodology

Research was carried out from March 2016 to May 2017, following a qualitative methodology. The answers have been coded from a structured interview in three levels of perception, cognition, emotion and motivation, and in three contexts: musical studies, music in the student's life, and the listening to music (Table 1). Selection of three participating groups has sought homogenization among them in age, gender and instrument.

Table 1. Contexts, Themes, Asks and Codes.

Educative center	**Asks**: What made you decide study music? $_1$What instruments do you know?$_2$ What instrument/s do you study? Why?$_3$ Would you study another instrument? Why?$_4$
Themes	**Codes**
Perception$_1$	Simple/Average/Complex
Cognition$_2$	Basic/Average/Elaborated
Emotion$_3$	Basic/Average/Complex
Motivation$_4$	Negative/Average/Positive
Life	**Asks**: Do your parents listen to music? Is the same music that you listen to?$_5$ Did you listen to your instrument before you started play it?$_6$ Where, how did it happen, what did you think?$_7$ Do you like instruments of your friends or family? Why?$_8$
Themes	**Codes**
Perception$_5$	Simple/Average/Complex
Cognition$_6$	Basic/Average/Elaborated
Emotion$_7$	Basic/Average/Complex
Motivation$_8$	Negative/Average/Positive
Listening to music	**Asks**: What kind of music do you listen to? What do you like about this music? Why?$_{10}$
Themes	**Codes**
Perception$_{10}$	Simple/Average/Complex
Cognition$_{10}$	Basic/Average/Elaborated
Emotion$_{10}$	Basic/Average/Complex
Motivation$_{10}$	Negative/Average/Positive

Perception analyzes way in which student conceives music inside and outside educational center. Codes (simple, average and complex) depend on if they are simple descriptions, elaborated, or in a more profound way, explaining moods or experiences of their own. Cognition shows the information or the previous knowledge of the students. It's basic, if student has little knowledge and previous learning situations; average, in case that are sufficient, but not high; and elaborated, if presents knowledge and broad situations. Emotion values student emotional states. Depending on emotion type, are codified in basic: joy or sadness; intermediate: fear or anger; and complex: combination of both. And motivation deepens in music student involvement, musical studies, and musical environment surrounding attitude. The codes are positive: if is flattering and provide extensive details about it; average: if details are not broad; and negative, if student hasn't got any implication.

Results

Results show differences between student typologies, especially in perception and cognition in the three contexts. In addition, average and dropout students tend to follow more similar line, which contrasts with those of success. For both: perception and cognition, success group always shows a more complex level than the others. However, results are more homogeneous when it analyzed in real contexts or listening to music than in educational contexts (Figures 1 to 6).

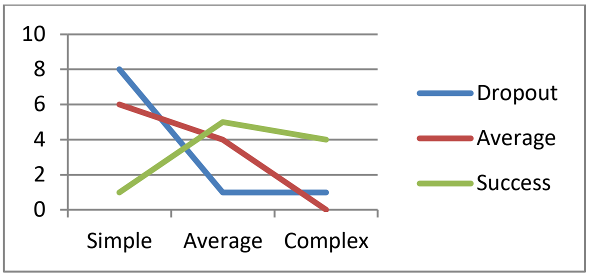


Figure 1. Educational perception.

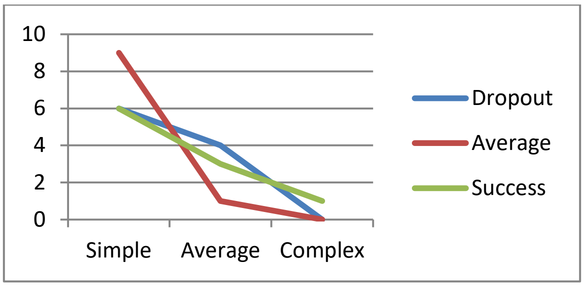


Figure 2. Perception in their lifes.

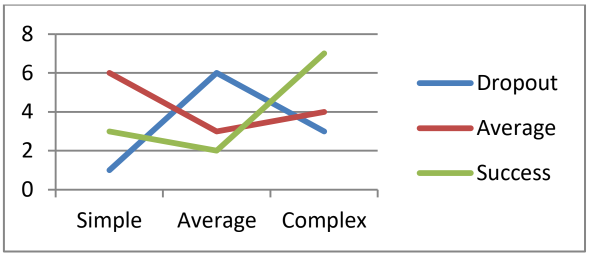


Figure 3. Listeing to music perception.

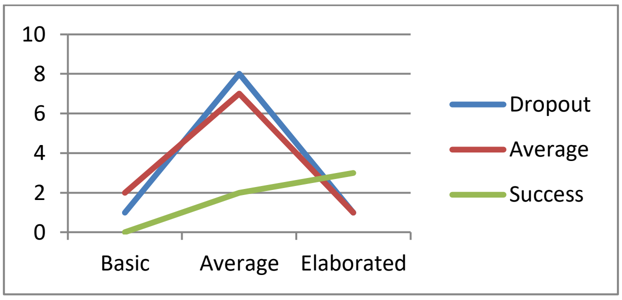


Figure 4. Educational cognition.

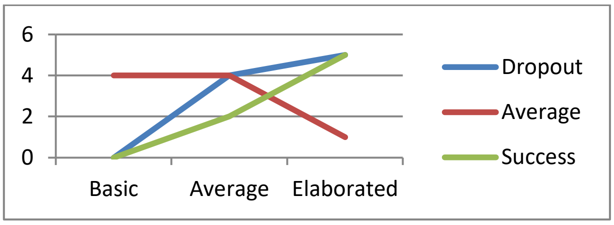


Figure 5. Cognition in their lifes.

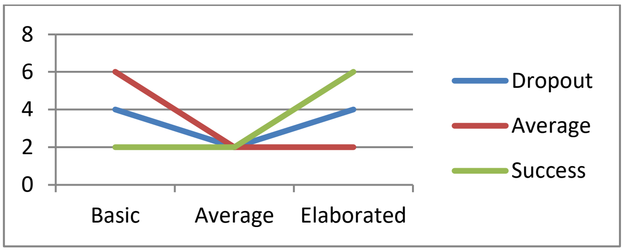


Figure 6. Listening to music cognition.

In emotion, majority of students don't use terminology specific to this construct, although there is a greater presence of these terms when they allude to listening to music. In the figure presented below (Figure 7), total count of responses about emotions, whatever their type, can been observed, in the three contexts. It's worth mentioning, however, that complex emotions are usually expressed by success group, being lesser by dropout and average groups.

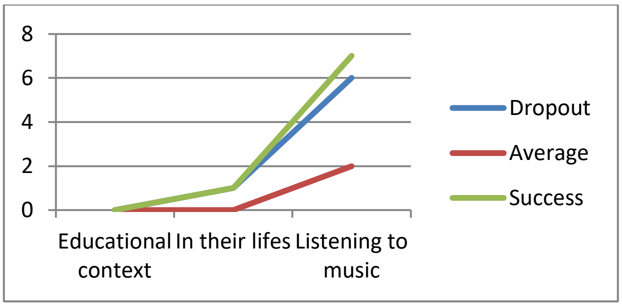


Figure 7. Emotional answers.

Same relationship with context is found in motivation, not finding large differences among student types (Figure 8), so a figure is added only with the total of responses for each type of motivation and context.

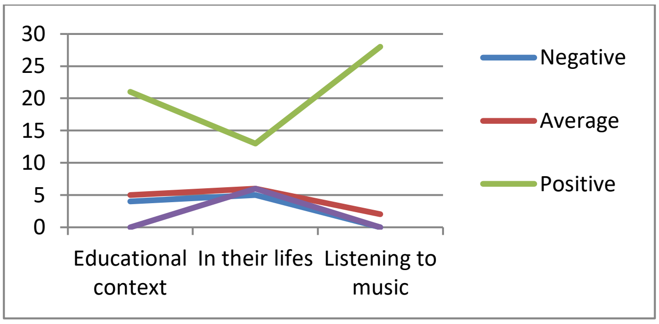


Figure 8. Motivational answers.

Regarding age, have been established two groups: 8 to 11 years old and 12 or more years old, each of which have a total of 5 students for each performance type. The major differences are found in educational center for motivation, in music in their lives for cognition and in listening to music for perception, as shown in the following table (Table 2).

Table 2. Differences by age.

Theme	Codes	8-11 years old	12 or more years old
In educational center			
Motivation	Negative	0.3	1
	Average	0.3	1.3
	Positive	4.3	2.6
In their lifes			
Cognition	Basic	0.6	0.6
	Average	2.6	0.6
	Elaborated	1	2.6
Listening to music			
Perception	Simple	1.6	0.3
	Average	3	0.6
	Complex	0.6	4

Conclusion

A study of three groups of students with differences in performance has been exposed to be able to analyze in them aspects so strongly intertwined as perception, cognition, emotion and motivation. The data shows in all of them results that are favorable for the group of higher performance and unfavorable for the group of dropout and average. The similarities include, not only the typologies already exhibited of students, but they even go more far and relate to age. In addition, it has also been presented in this research their relationships in different contexts in which musical activity is understood. Nevertheless, it is still possible to try to combine all these results into a conclusion that explains them together, as well as to relate it to literature on the subject. To do this, we will proceed by analyzing each topic differentially and finally, reach intended conclusion.

Emotion and motivation towards music are elements prior to musical learning and influence its quality to a great extent. However, the learning itself also modifies the previous structures to the same by the memories that these trigger. The differences in the profiles of the students analyzed in this research show it, being more favorable the cognitive results in students with higher performance. These students also present a higher motivational and emotional level, especially in formal learning situations. However, when we delves into students lifes and their listening to music, the differences found in all the themes discussed are minor, which leads to a reflection on the way in which music is treated in formalized educational environments. Undoubtedly, there should be a change towards an education that really starts from what the student is, in which the non-formal contexts have greater presence. This becomes especially necessary to change established preconceptions, especially in the adolescence, at which time music acquires a role of greater importance to the student. It is paradoxical that the greatest differences in relation to age have been found, precisely, in motivational and perceptive aspects at these ages.

It also draws attention to the fact of the little value given to emotion in educational contexts, in which the student reluctance to it. However, in the listening to music, the results show emotion as more relevant. In some way, playing an instrument is left out of the emotion it triggers, when this should even be prior to motivation. It can hardly be enjoyed the musical activity and get to integrate it as a way of life that combines the formal and informal, if it doesn't been start from this premise for any learning.

Literature shows that, without emotion, it is difficult to learn. Therefore, it is concluded that it is necessary to start from it also in the learning of a musical instrument. Research has also demystified the concept of talent since quite some time (Richert, Alvino, & McDonnell, 1982; Richert, 1991; Feldhusen, & Jarwan, 1993; Treffinger, & Feldhusen, 1996). On the other hand, this should also address factors such as the socio-cultural complexity of perception, attending to the motivational and understanding the interests, needs, behaviors and values of children (McPherson, 2006), which we sometimes determine as adults educated in a society and values acquired from years ago.

References

Altenmüller, E., Schlaug G. (2013). Neurobiological aspects of neurologic music therapy. *Music Med*, 5, 210-216.

Blood A. J., Zatorre R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences. 98*(20), 11818-11823.

Boal-Palheiros, G., & Hargreaves, D. J. (2011). Listening to music at home and at school. *British Journal of Music Education, 18*(2),103-118.

Collier, G. L. (2007).Beyond valence and activity in the emotional connotations of music. *Psychology of Music, 35*(1), 110-131.

Evers, E., Suhr, B. (2000). Changes of the neurotransmitter serotonin but not of hormones during short time music perception. *Eur Arch Psychiatry Clin Neurosci*, 250,144-147.

Fairley, J. (2006). Dancing back to front: regeton, sexuality, gender and transnationalism in Cuba. In *Popular music.* EUA: Cambridge University Press,

Feldhusen, F. J., & Jarwan, F. A. (1993). Identification of Gifted and Talented Youth for Educational Programs. En K. A. Heller, F. J. Mönks & A. H. Pasow (Eds.). *International Handbook of Research and Development of Giftedness and Talent.* Oxford: Pergamon Press.

Ferrucci R, & Priori A. (2014). Transcraneal cerebellar direct current estimulation. tcDCS: Motor control, cognition, learning and emotions. *NeuroImage, 85*(3), 918-923.

Fouce, H., & Pecourt, J. (2008). Emociones en lugar de soluciones. Música popular, intelectuales y cambio político en la España de la transición. *Transcultural Music Review*, 19.

Hargreaves, D. J. Hargreaves, J., & North, A. (2012). *Imagination and creativity in music listening.* Oxford, UK: Oxford University Press.

Jauset-Berrocal, J. A., Martínez, I. & Añaños, E. (2017). Aprendizaje musical y educación: aportaciones desde la neurociencia [Music learning and education: contributions from neuroscience]. *Cultura y Educación, 29*(2), 833-847.

Levitin D. (2008). *Tu cerebro y la música. El estudio científico de una obsesión humana*. Barcelona: RBA.

Martínez, I. & Jauset-Berrocal, J. A. (2017). Why do they choose their instrument? *British Journal of Music Education, 34*(2), 203-215.

Martínez, I. (2017). *Rock y Entornos Mediáticos como Recursos Educativos. Aproximaciones Psicológicas para Buenas Prácticas de Innovación e Investigación en Música.* [Rock and Media Environments as Educational Resources. Psychological Approaches for Good Practices of Innovation and Research in Music]. Alemania: Editorial Académica Española.

Martínez, I. (noviembre, 2016). *¿Por qué estudian música en su tiempo libre?* [Why do they study music in their free time?]. Poster presented in III Congreso Nacional y I Internacional de Sem-ee, realizado en Musikene, San Sebastián.

Martínez, I. (octubre, 2017). *Música, Emoción y Motivación. Aportes desde el campo neurocientífico.* Trabajo presentado en COMPSIMÚSICA. AEPMIM realizado en UNED. Madrid

Martínez, I. Montero, I. & Casas-Mas, A. (noviembre, 2016). *Aprendices de brujos. La autorregulación del aprendizaje musical* [Sorcerers apprentices. The self-regulation of musical learning.]. Paper presented in III Congreso Nacional y I Internacional de Sem-ee, realizado en Musikene, San Sebastián.

Martínez, I., Casas-Mas, A., & Montero, I. (noviembre, 2016). *Concepciones del profesorado sobre los inicios de la actividad musical extraescolar* [Teachers' conceptions about the beginnings of the extracurricular musical activity.]. Paper presented in III Congreso Nacional y I Internacional de Sem-ee, realizado en Musikene, San Sebastián.

McPherson, G. E. (2006). Developing Motvation. En G. E. McPherson (Ed.), *The Child as Musician: A Handbook of Musical Development* (pp. 213-238). Oxford: Oxford University Press.

Menon, V., & Levitin, D. (2005). The rewards of music listening: Response and physiological connectivity of the mesolimbic system*. NeuroImage, 28*(1), 175-184.

Ockelford, A. (septiembre, 2012*). Memory and creativity: Different sides of the same coin in musical improvisation?* Perspectives on Musical Improvisation Conference, Oxford.

Ohberg, L., Lorentzon, R., Alfredson, H., & Maffulli, N. (2004). Eccentric training in patients with chronic Achilles tendinosis: normalised tendon structure and decreased thickness at follow up. *British Journal of Sports Medicine, 38*(1), 8-11.

Richert, E. S. (1991). Rampant Problems and Promising Practices in Identification. En N. Colangelo y G.A. Davis (Eds.), *Handbook of Gifted Education.* Boston: Allyn and Bacon.

Richert, E. S., Alvino, J. J., & McDonnel, R. C. (1982*). National Report on Identification: Assesment and Recommendations for Comprehensive Identification of Gifted and Talented Youth.* New Jersey: A Product of Educational Information and Resource Center.

Shaw, J. (2001). Turn thrifty lectures into rich jazz recitals. *The Times Higher*, 24.

Sloboda, J. A. (2012). *La mente musical: La psicología cognitiva de la Música.* Madrid: Machado Ediciones.

Treffinger, D. J., & Feldhusen, J. F. (1996). Talented Recognition and Development. Succesor to gifted Education. *Journal for the Education of the Gifted, 19*(2), 181-193.

Trost, W., Frühloz, S., Schön, D., Labbé, C., Pichon, S, Gradjean, D., & Vuilleumier, P. (2014). Getting the beat: Entrainment of brain activity by musical rhythm and pleasantness. *NeuroImage,* 103, 55-64.

Trotta, F. (2009). Música popular, moral e sexualidade: Reflexões sobre o forró contemporáneo. [Popular music, moral and sexuality: Reflections on contemporary forró]. *Contracampo,* 20, 132-146.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Towards a New Multidisciplinary and Inter-processual Vision in the Aesthetics of Music: *Aesthetic Globalization of Musical Listening*

Irene Martínez Cantero[1]

[1] *Professional Music Conservatory, Pilar de la Horadada, Alacant, Spain*
Social Psychology and Methodology, Autónoma University, Madrid, Spain

[1]irenemc2000@yahoo.com

Abstract

Although philosophical themes have remained almost constant since the Enlightenment (Fubini, 1994), we consider that "In art there is nothing but the new" (Amount, 2001, p.163). Therefore, knowledge should not been limited what it already know, or the ways of knowing to what it was thought been known about art, with independence of whatever it wanted to know. The analysis of music is conditioned and limited by the conceptual models used (Martínez, 2005), but music research has presented a progressive advance towards multi-disciplinarity. However, none of the analyzes has attended to diverse agents simultaniusly. For example, Kivy (2005) always understands production, interpretation and perception as different entities. But there must be new proposals to contemplate music that goes "beyond the recurrent explanation to the experience of music as a type of communication that always says and does not say something deep and full of meaning" (Vilar, 2010, p 162). We propose two main topics of study of emotion in music: inter-disciplinarity and inter-processing, independently and related. From a perspective that we have called *Aesthetic Globalization of Musical Listening*, we intend a new organization of knowledge accumulated in other areas and of the different agents and situations that can be included in it. From the support of authors such as Calle (2006), Jiménez (2010) and Gianetti, (2002), we trying to concretize their study topics and possible relationship with other disciplines. The breadth of thematic approaches with which to share musical emotion from aesthetics, as well as the novelty of contemplating processes, agents, contexts and systems in it, could allow a work that responds to musical emotion in parallel with other truths.

Introduction

If the Music has changed and continues to change its concept regarding its artistic dimension nowadays, and if the Musicology is increasingly addressing its importance in a given social context; if the own emission and reception of the Music is more and more studied in the Psychology, it seems evident that mult-idisciplinarity as a base of knowledge is substituting the areas since which it is known, that offer only brushstrokes of it. That is why we can't close our ears to all these new ways of knowing, we must not only hear what we say or we are told; but also to listen to it. Nor can we fractionate the processes in which Music is immersed as if they were independent of each other, when it increasingly seems evident that it is the relationship and the implication of all of them that is closest to the truth that we have always wonder as people.

And it is for this, that we consider different philosophers who have contributed very highly to aesthetic musical knowledge. We will focus on emotion in music, moving since the old quarrels between formalists and expressionists, to analyze in detail the Kivy theory, which will be criticized under postulates perhaps, at least, closer to current system.

These premises, which will be endorsed in a small theoretical framework with specialists from various areas, are the beginning of this work in which an approach to the analysis of some contributions defending the need for change in aesthetics is sought, as any discipline, it seems to be losing its validity as isolated knowledge. It's what we want to know, and everything related to what we want to know what really matters, not so much the lenses we can use to look it. Although what is known also depends on the point of view from which one looks, we can look at Music from the different agents and contexts that intervene in its process of creation, emission and perception, which, without doubt, should enrich and improve the knowledge.

Purposes and Objectives

From these brief premises we want to contemplate two approaches: interdisciplinarity and interprocessuality of music, independently and related to the study of emotion in music. Although other authors have already opted for a new orientation of knowledge, this text offers a perspective that we have called the *Aesthetic Globalization of Musical Listening*, as a new organization of knowledge accumulated in these new currents; that is, we simply want to know and propose a new way of seeing the same fact.

In relation to this, the objectives that we set ourselves are the following:

1. To analyze the musical emotion in the Aesthetics of Music and focus it on the theories of Kivy.
2. Advancing towards the knowledge of this same aspect in other disciplines.
3. Find theoretical bases that allow to support the defense of a multidisciplinary, rather than aesthetic knowledge only, about emotion in music.
4. Develop a small theory that could support the inter-procesuality of musical emotion.
5. To find common aspects in the previous objectives that facilitate the denomination of *Aesthetic Globalization of the Musical Listening* that we propose, as well as to expose some of the purposes that could be included.

Methodology

The methodology that will be used, will be based on a bibliographic review of diverse disciplines that study musical emotion. Since the subject we want to find out is wide, in these pages we will only to see a brief approach to the most relevant or current findings of the same. Due precisely to this, it is considered a deepening in a particular author that has laid many of the foundations of the Aesthetics of Music and that is still a topical: Peter Kivy, with the intention to carry out a critical analysis a little more large. Before to it, will not be obviated other authors that can help to the proposal presented.

In addition, we will take as themes the most important items present in studies of Aesthetics of Music since Antiquity: the expressiveness of music, the reasons for which it is affirmed, the emotion that can arouse in the listener and the various agents who can intervene in this process taken widely: composer, music sheet, interpreter and audience.

Research Analysis

An approach to musical emotion in Kivy's aesthetics

Focusing on the Theory of Kivy (2005), the denial of emotion in absolute music is based on two assumptions. One is the undesirable character of negative emotions. In this sense, he considers that if music expresses common emotions in people, we would not choose to listen to a music that provokes in us emotions that are unpleasant to us. The other is that absolute music does not necessarily provoke a behavior.

For the first argument, it can be considered that it is scarce because the explanation or response is quite reductionist; that is, there could be other explanations to the problem posed by Kivy or simply negative emotions could be another problem to investigate, with multiple answers that could allude to causing less intense emotions than emotions themselves, which provoke a memory of emotions or that this can even be positive if it is taken as a moment of confrontation, for give some examples. The denial that music provokes emotions seems, therefore, quite radical and simplistic as a solution to the problem that he exposes as an argument.

As for the second, it could be said more of the same. Why there are no behaviors that respond to an emotion should be an important question of study, not a defense in favor of the denial of emotion in music. And both could be contemplated and in fact, this has been done, in other disciplines. We know that in classical music there is a protocol of its own in which the movement of the public is not possible except to applaud at certain times. Frith (1996) speaks of a classical concert culture, where the audience listens in silence, repressing all movement or expression. It is this music that Kivy mainly refers to, if we understand pure music in this way.

In addition, what he understand by pure music can lead to other genres and styles that do not use the word, such as instrumental music in jazz or flamenco, in which are allowed and favored behaviors. Nor has pure music presented this characteristic as something historical. At other times, pure music had well-defined purposes in which behavior mediated more freely. This is how the Theory of Affections in the Baroque arises, which takes as its central theme the expression of passions, states of excitement and emotions (Michels, 1977). Only after Romanticism and the instrumental music's protagonism, the great symphonies become platforms on which the composer carries out a discourse (Cook, & Dibben, 2001).

This author also analyzes music from the point of view of morality, which could affect us by enlightening us, motivating in us certain moral behavior or improving our character in general. He himself rejects the first and the second option, since he assumes that music can provoke emotions or motivations, something that he does not consider to be proven. As for the third, it seems to be the best accepted and could prompt philosophers to seek an explanation about their meaning, their [emotional] expressiveness or their moral character (Kivy, 2008). Nevertheless, the criticism to this position are many (for example, Davies, 2001) and they are based in the importance that the music acquires in certain occasions like totally amoral vehicle (Nazi programs, for example).

In fact, some authors have found specifically musical characteristics that serve in a certain way as a form of domain (Baccigalupo, & Montaño, 2008), since they are capable of uniting masses of people, of provoking community emotions, of creating inescapable links, in which usually involves ideological programs (Cohen, 2001) or a monetary purpose in a consumer market (Gértrudix, 2000). In any case, with it, in a certain way, Kivy affirms that music expresses something and affects us, which doesn't seem to fit very well with its defense at all costs of the denial of emotion in music.

Finally, in relation to the existence of concept of genius in Kivy (2011), he emphasize the role he gives to the context by opting for Haëndel, Mozart and Beethoven as prototypes of this idea, due to the role that historically they was awarded his contemporaries. However, the very concept of genius that Kivy defends is that equivalent to qualities such as inspiration or breaking the established rules to innovate.

In this part, Kivy does not seem to take into account the proper of his times or what common characteristics makes today they be considered as bearers of that special something in music. If his music has moved in his times and continues to do so at present, it would be, at least, a reason, in any case, to be able to give it the ability to provoke emotions in the musical experience. Could be this and not their historical or current role what is outstanding in them. For this, they use different musical resources and diverse experiences, for which this reason could also be considered, as well as many other points of view and defenses, as the reasons why they have endured in history.

In fact, even the concept of genius has been widely criticized and some authors establish other factors that demystify the contributions of Kivy in this regard. To give an example, in the History of Western Music, there are numerous examples of almost complete families of musicians (Bach, Haydn, Mozart ...). This perspective does not seem to take into account that, precisely for this reason, all these musicians had a particularly stimulating environment [that triggers emotions favorable to music] and the help of very close people in their music education (Jorquera, 2010).

On the other hand, returning to the musical experience, it is considered as a process in itself in which different agents and varied contexts intervene. But what is striking about this process is that, in most studies of the Aesthetics of Music, it is at the same time conceptually independent of the specific agents and contexts that are or are not present, but dependent of them empirically (agents and contexts). Although Kivy (2005) usually sees it exclusively from this point of view, and perhaps it has evolved from the material view to something more immaterial, always in other authors and always in Kivy, it has been analyzed in isolation; that is, production, interpretation and perception as distinct entities.

However, there must also be new proposals to contemplate music that go "beyond the recurrent explanation to the experience of music as a kind of communication that always says and does not simultaneously say something deep and full of meaning" (Vilar, 2010 , p 162).

This type of approach in which music is considered as a common whole for composer, performer, audience, time, concrete contexts, etc., it would allow us to respond to theories such as Kivy's and, perhaps, go a little beyond the recurring dilemma between formalists and expressionists about emotion in music. On the other hand, and also in reference to this author, it would be possible not to consider music as an isolated entity, from which the little consideration that this art may have in relation to other areas may derive (Kivy, 1983).

In other cultures, music is not that, but it is linked to dance and ritual, for example. In places like South America and Africa, music is part of daily life in a much more natural way, it is integrated into work, society, religion ... (Hagreaves, 2011, Martínez, 2014). Emotion is something that belongs to all human beings and music is also this: both are universal. Music has been present in all societies and cultures with different ends over time (Nettl, 1956, Merriam, 1964, Gregory, 1997, Casas-Mas, 2013).

In a way, therefore, it can be considered that Kivy's analysis manipulates a certain music and form of music: pure Western music, because in us there is a rooted class element around it. It makes us consider it superior to any other type of music, which should not be legitimate either, because it hides, like Kivy, not only a clear split between knowledge and emotion, reason and meaning; but also a commonly assumed hierarchy whereby reason must always prevail over emotion, knowledge about sensibility, form over content and, ultimately, mind over body (Torrado, Casas, & Pozo, 2005).

Sometimes doubt the assumptions that are analyzed, can lead to new ways of dealing with problems that are commonly assumed. Although we do not doubt Kivy's ability in some of his publications, a deep analysis from different points of view make up the grosso on which to conceptualize his theories (see, for example, Kivy, 2007), and should be the final point under which conceptualize and analyze any problem as a confrontation of disciplinary explanations towards the convergence of areas, and not so much the defense of a specific position or of a single theory, which we could always strike as relative.

Brief theoretical framework since other disciplines

Since time immemorial, the greatest thinkers have postulated eloquence about the ability of music to express powerful and subtle emotions. Is this then an illusion? (Collier, 2007). Hanslick (1957, referenced in Kivy, 1990, p.46) made a harsh criticism of this romantic and confused vision of emotion in music:

How can we speak of a clear sense of what is represented when nobody really knows what is represented? Probably everyone will agree about the beauty of a composition, but they will all be different depending on people (Hanslick, 1957, in Kivy, 1990, p. 46).

If we simply go a little deeper into the subject of listening to music and the emotion that it arouses in the listener, we know that the works carried out have been numerous. These works deal with the social (Hargreaves, 2012, Hargreaves, Hargreaves, & North, 2012, Sloboda, 2012), the neurobiological (Cross, 2004, Dietrich, 2004, Juslin and Västfjall, 2008, Ockelford, 2012), the emotional (Kivy, 1990; Collier, 2007), the musicological (Obelkevitch, 1989, Stockfelt, 1994, Shaw, 2001), the political (Attali, 1995, Fouce, & Pecourt, 2008) or even on its moral use (Foucault, 2001; Fairley, 2006) or the sexuality (Matos, 2007; Trotta, 2009).

Thus, the meaning in general and the musical in particular, should be understood as founded on humanity: our capacity to construct meaning, based on the imaginative structures by which we apprehend reality (Johnson, 1987), whatever the adopted approach or the discipline that is taken as a paradigm. We now know that musical cognition is not only a consequence of mere abstractions, but also the result of the use of the imaginative and bodily structures that emerge from our sensory and motor experience, and contribute to our understanding and guiding reasoning about the world (Martínez, 2005).

Although it is true that musical expression has limits and difficulties to complex emotions such as pride, admiration, shyness, envy or jealousy (Collier, 2002), this does not mean that there are no subtleties that can be expressed or that can't reducing to other basic emotions or categories (Plutchik, 1980, Roberts, & Wedell, 1994, Russell, 1980, Juslin, 1997a, 1997b, 1997c, Shaver, Schwartz, Kirson, & O'Connor, 1987). The actual question is whether musical expressiveness can be reduced to a few dimensions or emotions (Collier, 2007).

For the exact knowledge of it, we can no longer simply respond from a perspective, because advances in other areas of knowledge would even make some questions and affirmations ridiculous. Neither since the simple listener, and nor since the interpreter, as a carrier of the musical expression that transmits to the listener.

The different analyzes of the same piece of music are conditioned and constrained by the conceptual models that are used. Thus understood, it could be said, for example, that the musical motif pre-assigns values to the attributes that define a category and guides the categorization of different versions, whose variations will be an anchor point in the process of making the work. Each particular conceptual model is in turn outlined by a global conceptual model that informs, for example, about the theme and that is developed by abstraction of local models built on the broad knowledge base of a given culture (Martínez, 2005).

Conclusion

Some conclusions can be derived from the study that has been proposed and they could be a starting point for a greater and more specific study of the proposed objectives.

On the one hand, it has been proven that listening, interpreting and receiving music from an emotional point of view is something that has been repeatedly studied by various disciplines. Although all contribute brushstrokes to the same framework, they also find points in common that seem to give partial answers to questions that the Aesthetics poses at present and that, however, are not usually taken into account. On the other hand, it has also been seen that this division is a constant for the of what intervenes in the musical process. These disciplines usually refer to the emotional capacity of music, but they do not study it completely; that is, what they analyze is what emotional characteristics can be found in a work created by a composer, what reasons could have led the composer to determine the procedures and tools necessary to arouse these emotions, what is the role of the interpreter and how contributes to the emotion in it, and, above all, what is the emotion or emotions that can arouse in the listener.

From all this we can conclude that not everything is said and that, despite the criticisms that have been made against formalists and specifically for the Kivy theory, an aesthetic vision of music is necessary, but so is the need in it of new orientations that may enable to reduce the shortcomings that seem to be present in their current role as a discipline. We therefore consider that it should no longer be disregard the contributions of other areas, but rather seek a more complete vision of the musical process in a broad sense of the term, not divided either for the agents that intervene in it. That is why we now humbly want to contribute what is what we call the Aesthetic Globalization of Musical Listening and some ideas for the change of topics of study.

Aesthetic Globalization of Musical Listening

We define Aesthetic Globalization of Musical Listening as an attribute that allows aesthetics a general change in its current orientation. In it there are two aspects that are of special relevance. One of them is the need to contemplate in his study and in his postulates and affirmations the rest of the disciplines that study the same themes that an aesthetic of music is proposed, whether they are humanistic or scientific. The other, results from the need to contemplate the musical process in a broad sense, seeking for it those substrates that are common to the different agents, systems, processes and contexts that participate in the same event. The composer, the interpreter and the audience will be understood as agents, and it will be sought what exist in common in all of them in relation to listening to in the same musical work. Only in this way can we find out what is here true about it. In addition, agents can't be studied if they are not analyzed in a specific context in which all, or part of them are located. To give an example, a work by a Baroque composer can be heard at home in a recording with performers who were recording live in a church. Given the dilemma of the difference in contexts, we must analyze what is kept common in all of them.

The third part is the system in which the music is produced: sound or writing. For the first case, the possibilities are wide: open air, concert halls, church, cassette, vinyl, cd, dvd, contain or not images, etc. On the other hand, written music can be read or interpreted, created, it can have a paper or computer support, even be in the network, it can have to be analyzed to understand itself or it can simply not be understood, etc. Finally: what is the path that music follows since the composer to the listener?, how the creative idea arises?, how it materializes?, how it changes, varies, etc.?, what the interpreter recognizes and how he recognizes it,? what different points of sight there can be in different interpreters?, if there are historical recordings: how they resemble or differ the conception of the composer himself or other interpreters?, and what of it is capable of capturing the audience?, in a generalized way, in relation with his creation, is for practical utility or not heard? And, is this achieved?, etc.

Figure 1. Aesthetic Musical Globalization (Irene Martínez Cantero).

OTHER HUMAN DISCIPLINES

OTHER SCIENTIFIC DISCIPLINES

However, some theorists and philosophers have already studied and reflected on this (Jiménez, 2010, Gianetti, 2002), standing out Román de la Calle (2006):

> Is it not, after all, culture itself a kind of second nature? Is it generate culture, nature? Or perhaps, rather, the very claim of nature has been transformed… into a kind of second culture, into a modality… of cultural alternative? (de la Calle, 2006, p.86)

One should not look from the aesthetics more to the work of art, but rather to the natural objects, in which the contextual, the cultural and the sensitive should be immersed: "The silence of nature is clearly confronted with the loquacity of a diffuse aesthetic, which persistently tries to invade and colonize our environment and -from him- also our daily existence"(de la Calle, 2006, p.92). The rigorous examination of the new worlds of aesthetics seems to be seen as something true.

References

Amount, J. (2001). *La estética hoy* [The Aesthetics today]. Madrid: Cátedra.

Attali, J. (1995). *Ruidos: ensayo sobre la economía política de la música* [Noises: Essay on the Political Economy of Music]. Madrid: Siglo XXI.

Baccigalupo, C., & Montaño, V. (2008). Modelo de incorporación de tic en el proceso de innovación docente para la implementación de un b-learning [Model of incorporation of tic in the teaching innovation process for the implementation of a b-learning]. *Didáctica, Innovación y Multimedia,* 11, 1-7.

Casas-Mas, A. (2013). *Culturas de aprendizaje musical: concepciones, procesos y prácticas de aprendizaje en Clásico, Flamenco y Jazz* [Cultures of Musical Learning: conceptions, processes and learning practices in Classical, Flamenco and Jazz]. [Tesis Doctoral]. Universidad Autónoma de Madrid. Facultad de Psicología, Madrid.

Cohen, A. J. (2001). Music as a Source of Emotion in Film. En P.N. Juslin, y J.A.Sloboda (Eds.). *Music and Emotion, Theory and Research* (pp. 249-274). New York: Oxford University Press.

Collier, G. L. (2002). Why Does Music Express Only Some Emotions? A Test of a Philosophical Theory. *Empirical Studies of the Arts, 20*(1), 21–31.

Collier, G. L. (2007).Beyond valence and activity in the emotional connotations of music. *Psychology of Music, 35*(1), 110-131.

Cook, N., & Dibben, N. (2001). Emotion in culture and history: perspective in musicology. En P.N. Juslin y J.A. Sloboda (Eds.). *Music and Emotion, Theory and Research* (pp. 45-70). New York: Oxford University Press.

Cross, I. (2004). Music, meaning, ambiguity and evolution. En D. Miell, R. Mac Donald y D. Hargreaves (Eds). *Musical Communication* (pp. 27-43). Oxford: Oxford University Press.

Davies, S. (2001). *Musical Works and Performances:A Philosophical Exploration.* Oxford: Oxford University Press.

de la Calle, R. (2006). Seis consideraciones a favor de la estética natural [Six Considerations in Favor of Natural Aesthetics]. *Quaderns de filosofia i ciència*, 36, 85-92.

Dietrich, A. (2004). The cognitive neuroscience of creativity. Psychonomic *Bulletin Review, 11*(6), 1011-1026.

Fairley, J. (2006). Dancing back to front: regeton, sexuality, gender and transnationalism in Cuba. En *Popular music*. EUA: Cambridge University Press.

Foucault, M. (2001). *A história da sexualidade: o uso dos prazeres* [History of Sexuality: Use of Pleasures]. Rio de Janeiro: Graal.

Fouce, H., & Pecourt, J. (2008). Emociones en lugar de soluciones. Música popular, intelectuales y cambio político en la España de la transición [Emotions instead of Solutions. Popular Music, Intellectuals and Political Change in the Spain of the Transition]. *Transcultural Music Review*, 19.

Frith, S. (1996). *Performing rites: On the value of popular music.* Oxford, UK: Oxford University Press.

Fubini, E. (1994). *La estética desde la Antgüedad hasta el Siglo XX* [Aesthetics from Antiquity to the XX Century]. Madrid: Alianza Música.

Gértrudix, M. (2000). El icofono musical como portador publicitario [The Musical Tone as Advertising Carrier]. *Icono, 14*(1), 104-121.

Gianetti, C. (2002). *Estética digital. Sintopía del arte, la ciencia y la tecnología* [Digital Aesthetics. Sintopia of Art, Science and Technology]. Barcelona: ACC L'Angelot.

Gregory, A. H. (1997). The Roles of Music in Society. En D.J. Hargreaves and A.C. North (Eds.), *The Social Psychology of Music* (pp. 123-140). Oxford: Oxford University Press.

Hanslick, E. (1957*). The Beautiful in Music.* G. Cohen (Ed). New York: The Liberal Arts Press.

Hargreaves, D. J. (2012). Musical imagination: perception and production, beauty and creativity. *Psychology of Music, 40*(5), 539–557.

Hargreaves, D. J. Hargreaves, J., & North, A. (2012). Imagination and creativity in music listening. En D.J. Hargreaves, D.J.E. Miell, y R.A.R. MacDonald (Eds.), *Musical imaginations* (pp. 156–172). Oxford, UK: Oxford University Press.

Jorquera, M. C. (2010). Las ideas de estudiantes de Maestro sobre educación musical [The Ideas of Maestro Students about Music Education.]. *Investigación en la escuela*, 70, 101-112.

Juslin, P. N. (1997a). Can Results from Studies of Perceived Expression in Musical Performances Be Generalized across Response Formats? *Psychomusicology*, 16, 77-101.

Juslin, P. N. (1997b). Emotional Communication in Music Performance: A Functionalist Perspective and Some Data. *Music Perception, 14*(4), 383-418.

Juslin, P. N. (1997c). Perceived Emotional Expression in Synthesized Performances of a Short Melody: Capturing the Listener's Judgment Policy. *Musicae Scientiae 1*(2), 225-256.

Juslin, P. N., & Västfjall, V. (2008). Emotional Responses to Music: the need to consider underlying mechanisms. *Behavioral and Brain Sciences*, 31, 559-621.

Kivy, P. (1983). The Fine Art of Repetition: Essays in the Philosophy of Music. Cambridge: Cambridge University Press.

Kivy, P. (1990). *Music Alone*. Ithaca, New York: Cornell University Press.

Kivy, P. (2005). La teoría de la estimulación musical: replantearse lo impensable. En *Nuevos Ensayos sobre la comprensión musical* [The Theory of Musical Stimulation: Rethink the Unthinkable. In New Essays on Musical Understanding] (pp. 155-193). Barcelona: Paidós.

Kivy, P. (2007). *Music, Language, and Cognition: And Other Essays in the Aesthetics of Music*. Oxford: Oxford University Press.

Kivy, P. (2008). La moralidad musical [Musical Morality]. *Revue Internationale de Philosophie*, 246, 397-412.

Kivy, P. (2011). *El poseedor y lo poseído. Haëndel, Mozart y Beethoven y el concepto de genio musical* [The possessor and the possessed. Haëndel, Mozart and Beethoven and the Concept of Musical Genius] Madrid: Machado.

Martínez, I. (2014). La escuela y otros ámbitos de la educación musical [The School and other Areas of Music Education]. In C. J. Gómez y A. Escarbajal (Eds.), *Calidad e innovación en educación primaria* [Quality and Innovation in Primary Education] (pp. 109-120). Murcia: Universidad de Murcia, Servicio de Publicaciones.

Martínez, I. (september, 2005). *La audición imaginativa y el pensamiento metafórico en la música* [Imaginative Hearing and Metaphorical Thinking in Music]. Comunication presented in I Jornadas de Educación Auditiva de la Facultad de Bellas Artes de la Universidad Nacional de La Plata. Buenos Aires, Argentina.

Matos, C. (2007). Namoro e briga: as artes do forró [Dating and fighting: the arts of the forró]. In Nelson B. da Costa (Org), O charme dessa nação [The Charm of this Nation]. *Fortaleza: Expressão Gráfica.*

Merriam, A. P. (1964). *The anthropology of music*. Chicago: Northwestern University Press.

Michels, U. (1977). *Atlas de música* [Music Atlas]. Madrid, Alianza Editorial.

Nettl, B. (1956). *Music in primitive cultures.* Cambridge, MA: Harvard University Press.

Obelkevitch, J. (1989). In search of the listener. *Journal of the Royal Musical Association, 114*(1), 102-108

Ockelford, A. (septiembre, 2012). *Memory and creativity: Different sides of the same coin in musical improvisation?* Proposal for paper/practical demonstration at Perspectives on Musical Improvisation Conference en la Universidad de Oxford, UK.

Plutchik, R. (1980). *Emotion: A Psychoevolutionary Synthesis*. New York: Harper y Row

Roberts, J. S., & Wedell, D. H. (1994). Context Effects on Similarity Judgments of Multidimensional Stimuli: Inferring the Structure of the Emotion Space. *Journal of Experimental Social Psychology*, 30, 1–38.

Russell, J. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion Knowledge: Further Exploration of a Prototype Approach. *Journal of Personality and Social Psychology 52*(6), 1061-1086.

Shaw, J. (2001). Turn thrifty lectures into rich jazz recitals. *The Times Higher*, 24.

Sloboda, J. A. (2012). *La mente musical: La psicología cognitiva de la Música* [Musical Mind: Cognitive Psychology of Music]. Madrid: Machado Ediciones.

Stockfelt, O. (1994). Cars, buildings and soundscapes. En Helmi Järviluoma (Ed.), *Soundscapes: Essays on Vroom and Moo.* Tampere: Department of Folk Tradition University of Tampere and Institue of Rhythm Music Seinäjoki. Consultado en http://www.sibetrans.com/trans/trans7/finnegan.htm (16 of September of 2017).

Torrado, J. A., Casas, A., & Pozo, J. I. (2005).Las culturas de la educación musical: aprendiendo a interpretar un instrumento [Cultures of Music Education: learning to play an instrument]. *Estudios de Psicología, 26(*2), 259-269.

Trotta, F. (2009). Música popular, moral e sexualidade: Reflexões sobre o forró contemporáneo [Popular Music, Morality and Sexuality: Reflections on Contemporary Forró]. *Contracampo*, 20, 132-146.

Vilar, G. (2010). ¿Qué clase de experiencia es una experiencia propiamente musical? [What kind of experience is a properly musical experience?] In Alcaraz and Pérez Carreño (Eds.), *Significado, emoción y valor: Ensayos sobre filosofía de la música* [Meaning, Emotion and Value: Essays on Philosophy of Music] (pp. 43-62). Madrid: Antonio Machado.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Role of Enculturation in Music-Induced Emotions: A Study on Psychophysiological Responses during Music Listening

Anastasios Mavrolampados,[1] Geoff Luck[2]

[1,2] *Department of Music, Arts and Culture Studies, University of Jyväskylä*

[1]anmavrol@student.jyu.fi, [2]geoff.luck@jyu.fi

ABSTRACT

Previous cross-cultural studies in music and emotion have mostly focused on emotion recognition and whether basic perceived emotions are recognised across cultures. As a result, the impact of enculturation on music-induced emotions remains largely unexplored. In addition, such studies have relied mainly on subjective self-reports, ignoring other components of emotion such as physiology. Cross-cultural studies have suggested that cultural learning has a differential effect on certain emotional components (subjective feeling, physiology, and facial expression), yet this has not been tested in a music setting. To test this hypothesis, three groups of Finnish, Chinese, and Greek non-musicians listened to 20 excerpts of Western, Chinese, and Greek music that were selected from previous studies in which the emotional character of the music had been rated. Self-reports were used to collect continuous ratings of valence and arousal, along with measures of physiological activity (heart rate, skin conductance, and respiratory rate). Ratings of intensity, familiarity with the excerpt and familiarity with the music style were also collected after each stimulus. Results showed similar levels of familiarity with Western music across nationalities. However, the subjective measurements revealed group differences in the subjective feeling, even when familiarity was controlled for. Arousal was the only subjective rating that did not have a differentiating pattern, in line with previous research that has suggested arousal has a more universal quality. Physiological activity also showed less variation across nationalities, indicating that autonomic nervous system responses to music listening are less mediated by enculturation.

INTRODUCTION

In the context of music, existing cross-cultural studies have focused primarily on emotion recognition, using variations of discrete classification models. Several studies have argued that a few basic emotions (typically joy, sadness, and anger) are universally recognized across cultures (Adachi, Trehub, & Abe, 2004; Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunaga, 2004; Fritz et al., 2009). On the other hand, secondary emotions that share common properties (e.g. sadness and peacefulness that share low valence) are often misidentified and have not produced consistent results (Argstatter, 2015; Balkwill & Thompson, 1999; Laukka, Eerola, Thingujam, & Yamasaki 2013). This might be due to linguistic or taxonomic differences of discrete emotion labels (Mesquita & Frijda, 1992), translation inaccuracies or variations in meaning (Thompson & Balkwill, 2010), or to the existence of emotions that are unique to only certain cultures (Averill, 1982). Using models of emotion with fewer core principles instead of multiple word tags such as the dimensional model (Russell, 1980) could reduce the amount of linguistic errors, although few attempts have been made so far.

Another common challenge in music and emotion research in general is prior exposure to the stimuli (Eerola & Vuoskoski, 2013), which can be especially hard to balance in a cross-cultural setting. Familiarity is known to enhance emotional responses (Schellenberg, Peretz, & Vieillard, 2008) and as it has not been fully controlled in many existing studies, it is hard to conclude whether any observed differences across groups were due to cultural or familiarity factors.

Beyond emotion recognition, induced emotions have largely been ignored, making it unclear how cultural factors affect them. Egermann, Fernando, Chuen and McAdams (2015) investigated felt emotions using the dimensional model and physiological measurements on two groups of Canadian and Congolese Pygmies. Results suggested that physiology and arousal scores have more universal responses to low level acoustic features of music, whereas valence is affected more by cultural learning (Egermann et al., 2015). When it comes to the effect of enculturation in different components of emotion, Soto, Levenson, and Ebling (2005) found that cultures with distinct tendencies towards emotional expression also had distinct subjective emotional responses to aversive acoustic startle stimuli, while their behavioral and physiological responses were less differentiated. However, it is unclear whether these differences between the subjective feeling and physiology can be generalized in a more complex auditory context, such as music listening. It is also remains to be seen whether cultural predispositions towards emotional expression will have an effect on self-reports of music-induced emotions, making some ethnic groups to report stronger emotional responses than others.

To this extend, the current study poses the following research question:

What is the role of enculturation in subjective and physiological responses to music-induced emotions?

To address this question, a cross-cultural study was devised, using three groups of Finnish, Greek, and Chinese participants listening to Western, Greek and Chinese music. The three groups and music styles were selected based on previous experimental and ethnographic research. Subjective self-reports of the emotional experience and physiological measurements were collected. Familiarity was controlled with two sets of ratings, identifying both familiarity with the music excerpt and familiarity with the music style. The hypotheses were as follows:

Hypothesis 1: All groups would have a similar level of exposure to Western music, but a different degree of familiarity for Greek and Chinese music.

Hypothesis 2: The groups will have differentiated subjective ratings, but similar levels of physiological activity.

Hypothesis 3: When familiarity is controlled, some nationalities will consistently report stronger subjective emotional responses than others.

METHODS

PARTICIPANTS

A total of 65 participants (mean age 27.4, *SD* = 6.98, 35 females) were recruited in three groups based on nationality (Finns = 22, Greeks = 22, Chinese = 21). Criteria for selection were that all participants should be nationals, should have grown up and spent most of their life in the target country.

STIMULI

Twenty instrumental recordings (8 Western, 6 Greek, 6 Chinese) were used as stimuli, with durations ranging from 50 to 67 seconds (*M* = 55.3 seconds, *SD* = 4.76). They were edited in Protools (Avid, Version 11.03) to add a two-second fade-out effect at the end of each track and they were equalized in volume (a peak normalization filter was applied). The Chinese and Western stimuli were selected from databases of previous studies that rated the emotional character of the music (Chinese traditional music database: Hu & Lee, 2016; film music database: Vuoskoski & Eerola, 2011) and excerpts were chosen to represent all quadrants of the valence-arousal dimensional model (Russell, 1980). In the case of Greek music, Zacharopoulou & Kyriakidou (2009) provided a database that mainly consisted of songs of Greek traditional music. As there have not been enough existing instrumental excerpts, 12 tracks were pre-selected on premises of similarity with Zacharopoulou & Kyriakidou's database in terms of music style, rhythm and orchestration. These tracks were then tested in a pilot study with Greek listeners (*N* = 15), in order to identify the six final stimuli that better represented the four quadrants of the dimensional model.

MEASURES

The Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) 20-items inventory was used to measure participants' positive and negative affect before the experiment on a 5-point Likert scale ranging from 1 (*Very slightly or not at all*) to 5 (*Extremely*), while a questionnaire with demographic information was completed after the experiment.

Participants were instructed to rate the felt component of emotions ('the emotions awakened to them by the music') through a computer-based interface made in PsychoPy2 1.85.2 (Peirce, 2009). During stimulus presentation, continuous ratings of valence and arousal were collected on bipolar scales. At the onset of each stimulus, a marker appeared in the center of each scale and participants were instructed to adjust the ratings as their emotions evolved. Valence was described as the positive or negative character of the felt emotion (unpleasant – pleasant), and arousal as the energy of the felt emotion (low – high, sleep – awake).

After the end of each stimulus, a further set of intensity and familiarity scores were collected. Intensity was a unipolar scale described as the strength of the experienced emotion, ranging from 1 (*No emotion experienced*) to 7 (*Very strong emotion experienced*). Familiarity consisted of two distinct ratings. The first assessed familiarity with the excerpt on three levels from 1 (*Unfamiliar*) to 3 (*Very familiar*), and the second the familiarity with the music style from 1 (*Very unfamiliar*) to 7 (*Very familiar*).

Physiology was recorded using a MindMedia NeXus-10 MKII biofeedback system and visualized with BioTrace (MindMedia, Version 2017A). Electrocardiography (ECG) was measured with a 3-lead system where the electrodes were placed on the upper and lower chest. A respiration belt was used to record the respiration activity and was placed around the abdominal region. Electrodermal activity (EDA) was measured with two Velcro strap electrodes placed on the distal phalanges of the index and middle finger of the non-dominant hand. The sample rate was 256 Hz for ECG and 30 Hz for EDA and respiration.

PROCEDURE

The experiment was conducted in a sound-proof room, using a Samsung SyncMaster SA450 monitor for stimulus presentation and Audiotechnica ATH-M50x headphones. Participants signed a consent form containing general information about the experiment at the beginning of the session. The cross-cultural aspect was not mentioned, participants were informed that the study was about how music affects listener's emotions. The PANAS questionnaire was completed and the physiology equipment was attached. Then, a two-minute relaxation state recording was completed to obtain the baseline activity and ensure that all the electrodes were properly attached.

After the relaxation period a training session with further information began. The difference between perceived and felt emotions was explained and participants were instructed to rate how the music made them feel, not to rate the emotional character of the music. It was also mentioned that music might not evoke any emotion, which they could report via the intensity rating. Participants were then presented with a training music example to acclimatize them to the experimental setting. During this training trial, the volume was set by the participants to a comfortable level and it remained at the same level for the whole experiment. The order of the stimuli was randomised and a one-minute relaxation period occurred before the onset of each stimulus to eliminate carryover effects across trials. The duration of the music listening was approximately 40 minutes and participants could have a break half-way if they wished.

DATA ANALYSIS

Both the behavioral and physiological data were exported and preprocessed in Matlab (TheMathworks Inc., Version R2016b). The continuous valence and arousal ratings were averaged across excerpt and participant. As felt emotions are known to occur with a certain latency, the time onset for computing the mean valence and arousal score was identified as the first change that the participant inflicted in each score. If the participant had not registered a response in the first 15 seconds, a default time onset was placed in that instant.

The tonic component of EDA was extracted using a continuous decomposition analysis (Benedek & Kaernbach, 2010) in the Matlab-based Ledalab toolbox (V3.4.9). Respiration rate and heart rate scores were also computed from the respiration and ECG scores respectively, using the Biotrace Software. The baseline activity of the one-minute relaxation periods were averaged and subtracted from the raw signal of each succeeding stimulus, in order to correct for inter-subject differences in baseline physiology. Average scores were then computed for each excerpt and they were z-scored for each participants. For both physiological and behavioral data, composite scores were created from the responses of each stimulus, with music style being the grouping variable. This resulted in three scores for each measure, corresponding to the three music styles (Western, Greek, and Chinese).

RESULTS

The composite PANAS scores for positive and negative affect were calculated for each participant. Three participants with negative scores of more than two standard deviations from the grand mean were excluded from the analysis. A further participant was excluded due to equipment failure in the physiological measurements, reducing the total number of participants to 61 (20 Finns, 21 Greeks, and 20 Chinese). The Kolmogorov-Smirnoff test for normality in the behavioral and physiological variables within the groups was not significant, suggesting that assumptions of normality were not violated. Excerpt familiarity scores revealed that in 92%, of the cases participants had not listened to the excerpts before.

Familiarity with the music style was tested with a 3 × 3 (Nationality × Music style) analysis of variance (ANOVA), with music style serving as a repeating measure. Mauchly's test revealed that the assumption of sphericity was violated, therefore degrees of freedom were corrected with Huynh–Feldt estimates of sphericity (ε = .92). The main effect of music styles on the familiarity scores was not significant $F(1.84, 106.46) = 1.12$, $p = .33$, suggesting that the overall familiarity level with each music style was similar across participants. There was a significant interaction effect of music style and nationality $F(3.67, 102.82) = 50.98$, $p < .001$, $\eta^2 = .64$. To further investigate these results, one-way ANOVAs were carried out to test the familiarity of each music style across the nationalities (Figure 1). There was a significant effect on familiarity scores for Greek music, $F(2, 58) = 46.55$, $p < .001$, $\eta^2 = .62$; and for Chinese music, $F(2, 58) = 23.64$, $p < .001$, $\eta^2 = .45$. There was no significant effect on familiarity scores for Western music, $F(2, 58) = 2.53$, $p = .09$.

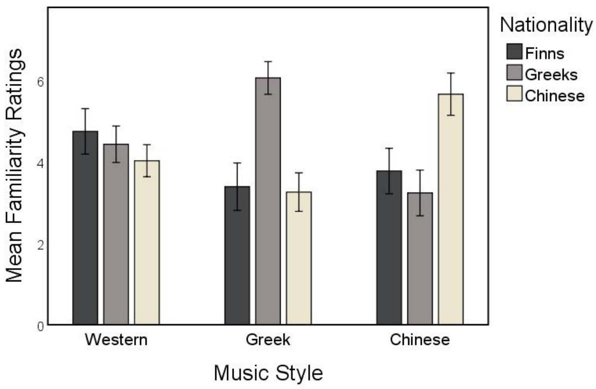


Figure 1. Mean Familiarity ratings per nationality and music style. The error bars denote confidence intervals at *p* = .05.

As only the Western music familiarity scores remained similar across the groups, only responses to Western music were considered in subsequent analyses. In order to investigate the effect of enculturation on the subjective feeling under Hypothesis 2, the valence, arousal and intensity scores were tested in a multivariate analysis of variance (MANOVA). There was a significant main effect of nationality $F(6, 114) = 5.78$, $p < .001$, $\eta^2 = .24$. To elaborate on these findings, separate univariate ANOVAs were carried out. There was a significant effect of nationality on valence and intensity, but not on the arousal scores: Valence, $F(2, 58) = 7.94$, $p = .001$, $\eta^2 = .22$; Intensity, $F(2, 58) = 6.25$, $p = .004$, $\eta^2 = .18$; Arousal, $F(2, 58) = 2.76$, $p = .072$. Pairwise comparisons with Bonferroni corrections were conducted to further investigate the group differences. The Finnish group reported significantly higher valence ratings than the two other groups at $p < .05$, the Chinese participants had the strongest intensity responses, and the Greeks showed significantly lower intensity responses to both other groups ($p < .05$).

Table 1. Means and Standard Deviations for subjective ratings in each group

	Finns		Greeks		Chinese	
Ratings	M	SD	M	SD	M	SD
Valence	6.36	.83	5.57	1.39	5.06	.75
Arousal	5.37	.74	4.91	.97	5.5	.87
Intensity	4.93	.68	4.24	1.04	5.09	.65

A MANOVA was also conducted for the physiology measurements for the Western excerpts. There was not a significant effect of the groups on the physiological responses $F(6, 98) = 1.61$, $p = .15$. One-way ANOVAs were tested for each measure, but only the heart rate scores revealed a significant effect $F(2, 57) = 4.37$, $p = .02$, $\eta^2 = .13$. Pairwise comparisons with Bonferroni corrections revealed that there was a significantly higher heart rate response in the Finnish group when compared to the Greek group ($p = .02$).

DISCUSSION

The aim of this study was to investigate the effect of enculturation on the emotional components of subjecting feeling and physiological responses. In order to control for differences in exposure to the music excerpts and the music style, two sets of familiarity scores were used to assess familiarity with the excerpt and familiarity with the music style. The music excerpt familiarity scores revealed that in 92% of cases participants reported that they either had not listened to the excerpt before or they were not certain. While it is possible that participants had been exposed to some of the excerpts and a recollection bias existed, it can be argued that emotional responses influenced by evoked episodic memories (Juslin & Västfjäll, 2008) remained relatively low. With regards to familiarity with the music styles, there were differences in the Greek and Chinese music, with the respective nationality groups exhibiting considerably higher familiarity scores than the rest (Greeks being more familiar with Greek music and Chinese more familiar with Chinese music). Hypothesis 1 was confirmed, as familiarity with Western music scores was similar and had no significant differences across the groups. This result concurs with Huron's (2008) notion that many non-Western cultures have a level of implicit knowledge of Western music, as a result of globalization. It also supports the idea that Western music can be a suitable candidate for baseline stimuli in cross-cultural research, in order to minimize discrepancies in the prior exposure to the stimuli and music style.

The results of the subjective ratings revealed that only arousal was undifferentiated across the groups. This is in agreement with Egermann et al. (2015), suggesting that arousal shows less variation across different cultures and has more universal tendencies. On the other hand, even when the level of familiarity was relatively similar (Western music), there were significant variations across the groups in valence and intensity scores. As for the autonomic nervous system activity, physiology scores revealed that when the familiarity factor was controlled in Western music, all groups showed similar patterns of activity, with the only difference observed in the heart rate scores between the Finnish and Greek participants. These results seem to concur with previous cross-cultural studies (Levenson, Ekman, Heider, & Freisen, 1992; Soto et al., 2005), suggesting that physiology is less subject to individual differences and societal norms of behavior and expression, whereas subjective feeling is shaped more by cultural learning and can vary greatly across cultures.

Hypothesis 3 held that cultural groups with different tendencies to emotional expression would also consistently report distinct levels of subjective feeling when familiarity factors were controlled. However, the pairwise comparisons revealed that this was not the case, as there were not consistent differences across the groups in all subjective measures. Although the Finnish group had significantly higher responses in terms of valence, this was not the case with intensity and arousal. In a similar vein, while the Chinese group reported lower levels of valence than the other two, it outperformed both in terms of arousal and intensity, with the intensity scores being significant only when compared to the Greek group. Possible explanations for these inconsistencies might be traced to the relatively small sample size and confounding variables that are present in experimental designs of naturalistic listening, having a negative effect on validity. Equally important is that although familiarity levels remained relatively similar across the groups, they were not identical, with the Finnish group reporting the highest level of familiarity with Western music. However, eliminating fully the background exposure discrepancies between participants is unlikely in a music setting, and thus will always introduce some bias in self-report comparisons. Therefore, Hypothesis 3 cannot be confirmed and further research is needed to identify exactly how the subjective feeling of groups of different cultural backgrounds is affected during music listening.

CONCLUSION

The current study investigated how enculturation affects different components of emotional responses during music listening in a cross-cultural setting. Results suggested that Western music can be used as a baseline across certain groups for cross-cultural comparisons. It was shown that the component of subjective feeling was more susceptible to cultural factors and had more variation across the groups compared to physiology. This has implications in future music studies dealing with emotions in cross-cultural settings, as it emphasizes the need to consider mixed methods for accurately measuring emotional activity. In addition, this study highlights the need to control familiarity factors more strictly in future studies and can have methodological implications for the use of dimensional over discrete models, in order to reduce linguistic errors in cross-cultural research.

REFERENCES

Adachi, M., Trehub, S. E., & Abe, J.-I. (2004). Perceiving emotion in children's songs across age and culture. *Japanese Psychological Research, 46*, 322–336. https://doi.org/10.1111/j.1468-5584.2004.00264.x

Argstatter, H. (2015). Perception of basic emotions in music: Culture-specific or multicultural?. *Psychology of Music, 44*(4) 674 –690. https://doi.org/10.1177/0305735615589214

Averill, J. R. (1982). *Anger and aggression: An essay on emotion*. New York, NY: Springer-Verlag.

Balkwill, L. L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*(1), 43–64. https://doi.org/10.2307/40285811

Balkwill, L. L., Thompson, W. F., & Matsunaga, R. I. E. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research, 46*(4), 337-349. https://doi.org/10.1111/j.1468-5584.2004.00265.x

Benedek, M. & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*, 80-91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception, 30*(3), 307–340. https://doi.org/10.1525/MP.2012.30.3.307

Egermann, H., Fernando, N., Chuen, L., & McAdams, S. (2015). Music induces universal emotion-related psychophysiological responses: comparing Canadian listeners to Congolese Pygmies. *Frontiers in Psychology, 5,* 1341. https://doi.org/10.3389/fpsyg.2014 .01341

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current biology*, *19*(7), 573-576. https://doi.org/10.1016/ j.cub.2009.02.058

Hu, X., & Lee, J. H. (2016). Towards global music digital libraries: A cross-cultural comparison on the mood of Chinese music. *Journal of Documentation*, *72*(5), 858-877. https://doi.org/10.1108/JD-01-2016-0005

Huron, D. (2008). Science & music: Lost in music. *Nature*, *453*(7194), 456-457. https://doi.org/10.1038/453456a

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences*, *31*(5), 559-575. https://doi.org/10.1017/S0140525X08005293

Laukka, P., Eerola, T., Thingujam, N. S., Yamasaki, T., & Beller, G. (2013). Universal and culture-specific factors in the recognition and performance of musical affect expressions. *Emotion, 13*(3), 434–449. https://doi.org/10.1037/a0031388

Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra. *Journal of personality and social psychology*, *62*(6), 972.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112*, 179–204.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in neuroinformatics*, *2*, 10. https://doi.org/10.3389/neuro.11.010.2008

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *59*, 1161-1178.

Schellenberg, E. G., Peretz, I., & Vieillard, S. (2008). Liking for happy-and sad-sounding music: Effects of exposure. *Cognition & Emotion*, *22*(2), 218-237. https://doi.org/10.1080/02699930701350753

Soto, J. A., Levenson, R. W., & Ebling, R. (2005). Cultures of moderation and expression: emotional experience, behavior, and physiology in Chinese Americans and Mexican Americans.

Emotion, *5*(2),154-165. https://doi.org/10.1037/1528-3542.5.2.154

Thompson, W. F., & Balkwill, L-L. (2010). Cross-cultural similarities and differences. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion. Theory, research, and applications* (pp. 755–788). New York, NY: Oxford University Press.

Vuoskoski, J. K., & Eerola, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, *15*(2), 159-173. https://doi.org/10.1177/1029864911403367

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54*(6), 1063-1070.

Zacharopoulou, K., & Kyriakidou, A. (2009). A cross-cultural comparative study of the role of musical structural features in the perception of emotion in Greek traditional music. *Journal of Interdisciplinary Music Studies*, *3*(1), 1-15.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The effects of tone duration in discriminating major/minor modes

Solena Mednicoff,[1] Stephanie Mejia,[1] Jordan Ali Rashid,[1] and Charles Chubb[1]

[1] *Department of Cognitive Sciences, University of California, Irvine, USA*

[1]smednico@uci.edu, [1]smejiane@uci.edu, [1]jrashid@uci.edu, [1]cfchubb@uci.edu

Abstract

The major and minor scales play vital roles in western music; however, many listeners have difficulty discriminating between major vs. minor melodies (Leaver & Halpern, 2004). Tone-scrambles are stimuli designed to isolate musical qualities produced by variations in scale from other aspects of musical structure. Most listeners (≈70%) perform near chance in classifying major vs. minor tone-scrambles (rapid, random sequences comprising 8 G5's, 8 G6's, 8 D6's, and either 8 B6's (major) or 8 Bb6's (minor)), while the other 30% perform near perfect (Chubb et al., 2013). Moreover, the sensitivity required for this task generalizes to other tone-scramble tasks requiring judgments unrelated to differences between the major vs. minor scales (Dean & Chubb, 2017). In previous tone-scramble studies, all stimuli were very rapid. This raises the possibility that high-performers differ from low-performers solely in being able to extract scale-generated qualities from these rapid, musically degenerate stimuli. If so, when the stimuli are presented more slowly, the gap in sensitivity separating high- and low-performers should disappear. The current study tested this prediction. Seventy-three participants with variable musical training were tested in 4 tasks. In the *k*-task (k=1,2,4,8) each (randomly sequenced) tone-scramble contained *k* copies of each of the notes G_5, G_6, D_6, and either *k* B_6's or *k* Bb_6's. The duration of each tone was 520/*k* ms; thus all stimuli lasted 2.08 sec. Listeners strove (with feedback) to classify stimuli as major vs. minor. Performance (a) was strongly correlated across tasks, (b) conformed to the same bimodal distribution as observed previously, and (c) was equally good in the 2-, 4- and 8-tasks but significantly worse in the 1-task. Thus, high-performers in tone-scramble tasks do not differ from low-performers solely in their ability to extract scale-properties from very rapid stimuli. The difference in sensitivity persists across all four conditions.

Introduction

The major and minor musical modes are central to Western music. The differences in these modes are thought to contribute to the emotional expressiveness of music. For example, music in the major mode has been said to sound "happy," whereas music in the minor mode sounds "sad" (Blechner, 1977; Crowder, 1984; 1985a,b; Gagnon and Peretz, 2003; Gerardi and Gerken, 1995; Heinlein, 1928; Hevner, 1935; Kastner and Crowder, 1990; Temperley and Tan, 2013). Despite the important role played by the major and minor modes in western music, however, there is evidence that many listeners cannot discriminate major vs. minor melodies (Halpern, 1984; Halpern et al., 1998; Leaver and Halpern, 2004).

Chubb et al. (2013) investigated this issue further using stimuli called "tone-scrambles" created in order to isolate effects of mode from other aspects of musical structure such as rhythm, timbre, loudness, etc. The tone-scrambles used by Chubb et al. (2013) were randomized sequences of tones drawn from either the G major (G_5, B_5, D_6, and G_6) or G minor (G_5, Bb_5, D_6, and G_6) scale. Each tone was 65 ms in duration and occurred eight times in the stimulus sequence. On a given trial, the listener attempted to classify the stimulus as major vs. minor and was given immediate feedback. The tone-scramble classification task partitioned listeners cleanly into high-performing and low-performing groups: 30% of listeners were near perfect whereas 70% were near chance at the task.

The results from Chubb et al. (2013) suggest that there exists some cognitive resource that high-performing listeners possess in greater measure than low-performing listeners. Dean and Chubb (2017) used new types of tone-scrambles to probe the nature of this cognitive resource. Again, each tone-scramble stimulus contained 32 randomly sequenced 65 ms tones, with 8 copies each of G_5, D_6, and G_6 to establish G as the tonic center throughout the experiment. The differences across the five tasks were in the target tones. In the "2" task, the target tone was either the diminished second or major second, Ab_5 or A_5; the "3" task had either the minor or major third, Bb_5 or B_5; the "4" task had either the perfect fourth or tritone, C_6 or Db_6; the "6" task had either the minor or major sixth, Eb_6 or E_6; and the "7" task had either the minor or major seventh, F_6 or Gb_6.

Dean and Chubb (2017) tested 139 listeners, *k,* in each of these five tone-scramble tasks, *t,* and were able to describe their results using a model in which performance in all of the 2, 3, 4, 6 and 7 tasks is determined by a single cognitive resource, *R*. This model is defined below in equation 1. In this equation, $d'_{k,t}$ represents performance as reflected by the level of d-prime achieved by listener *k* in task *t*, R_k represents the amount of *R* possessed by listener *k,* and F_t represents the strength which with which *R* facilitates performance in task *t*.

$$d'_{k,t} = R_k F_t + measurement\ noise \qquad (1)$$

The model accounted for 79% of the variance in $d'_{k,t}$ across all listeners in all five tasks. Again just as in Chubb et al. (2013), most listeners (around 70%) had R_k values around zero, which produces near-chance performance in all five tasks, whereas the other 30% of listeners had much higher levels of R_k. Some tasks, however, were facilitated more strongly by *R* than others. In particular, $F_2 \approx F_3 \approx F_6 > F_4 > F_6$. Dean and Chubb (2017) concluded that the resource *R* confers sensitivity not just to the difference between the major vs. minor musical modes but more generally to the spectrum of qualities that music can create by establishing a tonic and varying the ensemble of notes that compose a scale relative to the tonic. Accordingly, they called the resource *R* "scale-sensitivity."

The current study

The interpretation of *R* as "scale-sensitivity" may be unwarranted, however. In both Chubb et al. (2013) and Dean & Chubb (2017), the tone-scramble stimuli were composed of rapid 65 ms tones. It is possible that the main difference between high- vs. low-performing listeners is that high-performing listeners can extract scale-defined properties from these rapid sequences whereas low-performing listeners cannot. If this conjecture is true, then when stimuli are presented more slowly, the gap in sensitivity that separates low- and high-performing listeners should disappear. The present study addresses this question.

Methods

Participants

Seventy-three listeners participated in this study. All participants were undergraduate students recruited through the UCI SONA-system, were compensated with course credit, and had self-reported normal hearing. All methods were approved by the UCI Institutional Review Board.

Stimuli

The stimuli in this experiment were major and minor tone-scrambles that contained scale degree three. Each tone was a pure tone that was windowed by a raised cosine function that had a 22.5 ms rise time. The five notes used were the same as in Chubb et al. (2013) and drawn from the standard equal-tempered chromatic scale: G_5 (783.99), $B\flat_5$ (932.33 Hz), B_5 (987.77 Hz), D_6 (1174.66 Hz), and G_6 (1567.98 Hz).

The differences in the tone-scramble stimuli used in each condition can be seen in table 1. The four conditions are designated by n = 1, 2, 4, and 8, where the n-task contains n each of the tones G_5, D_6, and G_6, as well as n copies of the target tone. For the major tone-scramble stimuli, the target tone is the major third, B_5, and for the minor tone-scrambles, the target tone is the minor third, $B\flat_5$. The individual tones in each tone-scramble stimulus in the n-task lasted for 520/n ms. Thus, in each task the entire tone-scramble had a total duration of 2.08s. For instance in the 2-task, a major tone-scramble is composed of an eight note randomized sequence, with each tone lasting 260ms. Of the eight notes in this stimulus, 2 are G_5, 2 are B_5, 2 are D_6 and the remaining 2 are G_6.

Design and procedure

Each listener was tested in separate blocks in the 1-task, 2-task, 4-task, and 8-task. In each task, listeners first heard four alternating demos each of the n-each major and the n-each minor tone-scrambles. Next, listeners participated in two blocks of 50 trials each, with 25 "major" tone-scrambles and 25 "minor" tone-scrambles, and a brief break between the two blocks. The task order was counterbalanced across listeners using a Latin square design.

On a given trial in the n-task, the listener would be presented with either an n-each-major or an n-each-minor tone-scramble. The listener was then visually prompted to press "1" on the keyboard for a "HAPPY (major)" type, or "2" for a "SAD (minor)" type. Visual correctness feedback was given after each trial, and proportion correct was presented at the end of each block. For the analysis, the first block of each n-task was treated as practice, and d' was calculated from performance on the second block as the dependent measure. At the start of the experiment, the listener filled out a brief questionnaire; however, the only information from the questionnaire used in our analysis is the number of years of musical training. All participants wore noise-cancelling headphones with volume adjusted to comfortable listening levels during testing.

Results

For each listener in each task, d' was estimated from the 50 trials in the second block. If a listener responded correctly on all 25 "major" (or "minor") stimuli in a block, then the probability of a correct "major" ("minor") response was adjusted to $\frac{24.5}{25} = 0.98$ (as suggested by Macmillan and Kaplan (1985)); this yielded d' values of 4.1075 for listeners who performed perfectly across all trials in a given task.

Figure 1 shows scatterplots of d' for all listeners in each task vs. all other tasks. In each of the scatterplots, there exists a large group of listeners with d' values near 0 in both tasks and a smaller group of listeners with d' values near 4.1075 (or near perfect) in both tasks. In addition, each plot shows some intermediate listeners between these two extreme groups, and the relative difficulty of the two tasks being compared in a given scatterplot can be seen in the distribution of this group.

Looking at the left column of Figure 1, this intermediate group of listeners tends to have higher d' in the 2-, 4-, and 8-task vs. the 1-task. Specifically in each of these three plots, a majority of the points corresponding to these intermediate listeners fall above the main diagonal. The difference in difficulty between the 1-task and each of the three other tasks h is highly significant as confirmed by a paired-samples t-test of the null hypothesis that the mean value of $d'_{k,h} - d'_{k,1}$ is equal to 0. In contrast, the tests comparing performance between the 2-, 4-, and 8-tasks did not reach significance. This evidence suggests that the 1-task is more difficult than the other three tasks, which are roughly equal in difficulty.

Table 1. Description of tone-scrambles used in the 1-, 2-, 4- and 8-tasks. Table adapted from Mednicoff, Mejia, Rashid, and Chubb (under review).

Task	Stimulus type	Tone duration	# G_5's	# $B\flat_5$'s	# B_5's	# D_6	# G_6's
1-task	1-each-major	520 ms	1	0	1	1	1
	1-each-minor	520 ms	1	1	0	1	1
2-task	2-each-major	260 ms	2	0	2	2	2
	2-each-minor	260 ms	2	2	0	2	2
4-task	4-each-major	130 ms	4	0	4	4	4
	4-each-minor	130 ms	4	4	0	4	4
8-task	8-each-major	65 ms	8	0	8	8	8
	8-each-minor	65 ms	8	8	0	8	8

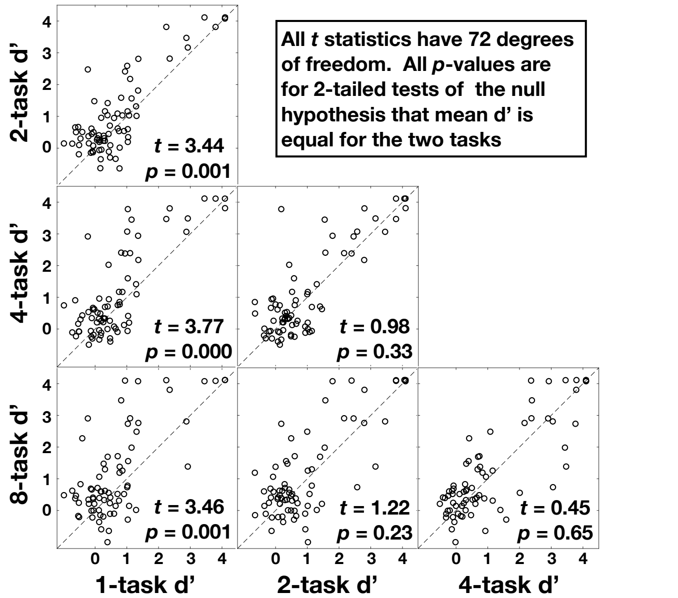


Figure 1. Scatterplots comparing *d*' performance of each listener across two tasks for all tasks in the experiment.

The bilinear model

We follow Dean and Chubb (2017) in using the model of Eq. (1) to describe the current results. This "bilinear" model proposes that $d'_{k,t}$, the value of d' achieved by listener k in task t, is proportional to (1) the level R_k of some congnitive resource R possessed by listener k and (2) the strength F_t with which R facilitates performance in task t. The reader will note that the model of Eq. (1) is under-constrained: for any real number A, predictions identical to those of Eq. (1) are given by a model in which AR_k is substituted for R_k and F_t/A is substituted for F_t. We therefore impose the additional constraint that $F_{1-task} + F_{2-task} + F_{4-task} + F_{8-task} = 4$.

This constraint is convenient for two reasons. First, if all four tasks are equally facilitated by R, then F_t will be 1 for all tasks t; therefore, deviations from 1 directly indicate deviations of relative facilitation strength. Second, this constraint makes R_k (the level of R possessed by listener k) equal to the average value of $d'_{k,t}$ achieved by listener k across the four tasks.

Estimated values for F_t are shown in figure 2. It should be noted that F_{1-task} is lower than F_{2-task}, F_{4-task}, and F_{8-task}, which is consistent with listeners finding the 1-task more difficult than the other three tasks.

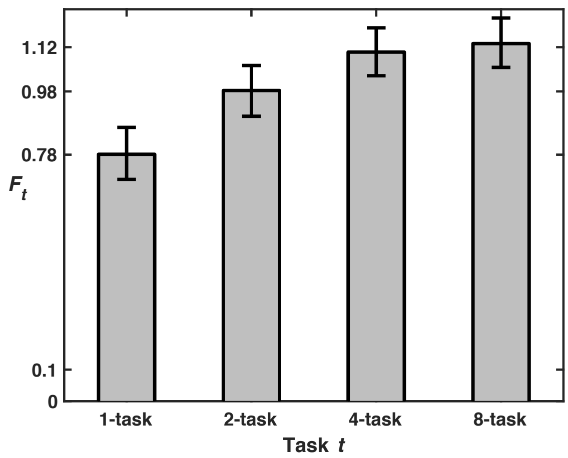


Figure 2. Estimated F_t for all listeners for each task based on the bilinear model.

Figure 3 shows the histogram of R values for all 73 listeners on the left, while the right shows the distribution of predicted proportion correct of these listeners in the 8-task. The histogram on the left is very similar to the histogram of observed R values seen in Dean and Chubb (2017): it is positively skewed with a large number of listeners with R values near 0, yet it is not bimodal. In contrast, when we plot the histogram of predicted proportion correct (assuming all listeners used optimal criteria) based on the model, this results in a bimodal distribution very similar to the bimodal distribution in performance observed in Chubb et al. (2013).

But how well does the bilinear model capture the current data? Figure 4 plots the estimates of $d'_{k,t}$ derived from the data for each listener k in each task t against the values predicted by the bilinear model (Eq. (1)). As suggested by the strong linear relationship evident in this figure, the model does a good job of describing the data. Specifically, it accounts for 84% of the variance in the $d'_{k,t}$ values for the 73 listeners across all four tasks.

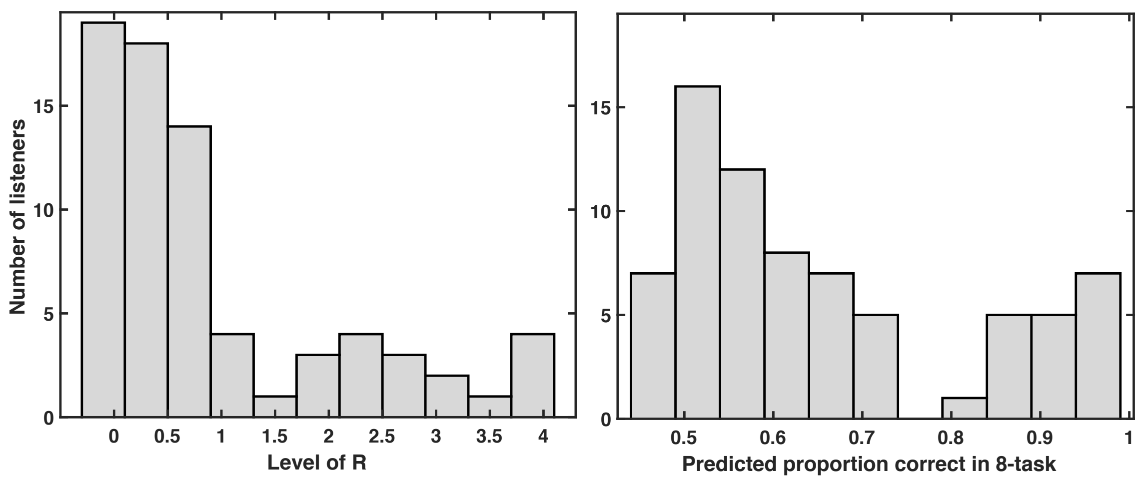


Figure 3. *Left:* Distribution of level of *R* possessed by all listeners. *Right*: Distribution of predicted proportion correct of all listeners in the 8-task based off of the model R_k values.

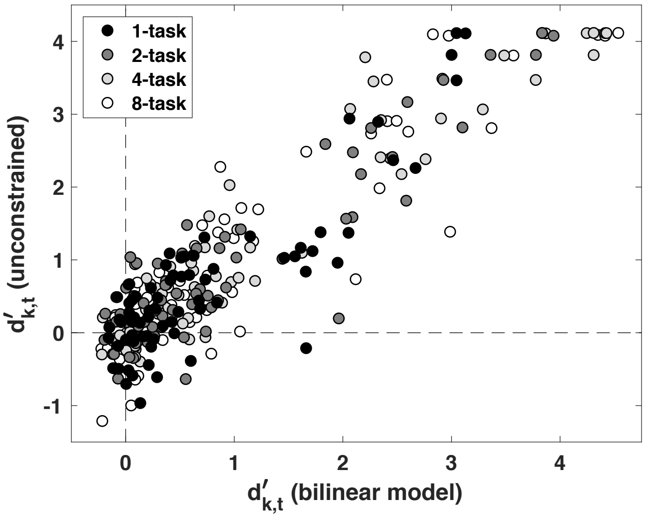


Figure 4. Scatterplot comparing the predicted $d'_{k,t}$ based on the bilinear model against the unconstrained $d'_{k,t}$ values from the data for all four tasks.

One might wonder whether the level of R possessed by a listener can be predicted by years of musical training. Figure 5 aims to answer this question by plotting the levels of R possessed by listeners as a function of their self-reported years of musical training. Although the correlation of 0.50 seems high, it is due mainly to the large number of listeners with no musical training possessing very low levels of R. It should be noted that there are many listeners with five or more years of musical training who possess $R_k < 1$ – which corresponds to an average of < 70% correct across all four of the tasks. Thus, musical training is not sufficient to produce a high level of R. On the other hand, none of our listeners with less than 3 years of musical training had levels of R greater than around 2 (approximately half the maximum level of R observed in our sample). The current results are thus consistent with the idea that musical training may be necessary to achieve the highest levels of R.

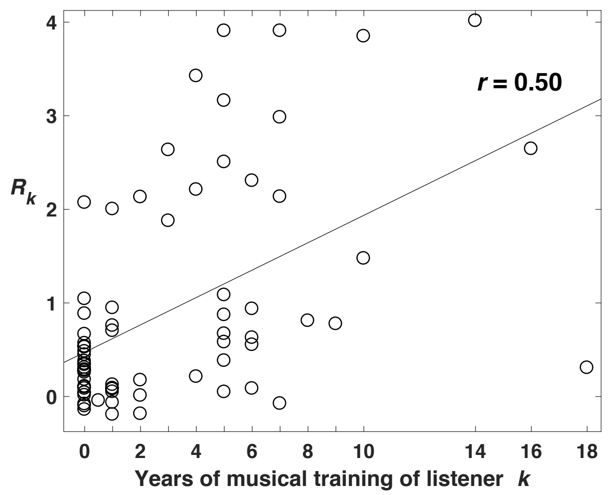


Figure 5. Correlation between the cognitive resource, *R*, and the years of musical training for listener *k*.

Conclusion

All the stimuli presented in Chubb et al. (2013) and Dean and Chubb (2017) were at a rate of 32 tones per 2.08s. This raises the possibility that high-performing listeners in these previous studies may differ from low-performing listeners solely in being able to extract scale-produced qualities from rapid, random stimuli. If so, then the difference between high- vs. low-performing listeners should disappear if stimuli are presented more slowly. Specifically, listeners who perform poorly on the 8-task should perform better on the 4-, 2-, and especially the 1-task. As our results reveal, this was not the case; in fact, the 1-task, in which each tone was presented for more than half a second, proved to be the most difficult on average.

In ruling out the possibility that high-performers differ from low-performers merely in being able to extract scale-produced qualities from rapid, random stimuli, the current findings support the idea proposed by Dean and Chubb (2017) that the cognitive resource underlying performance in tone-scramble tasks confers sensitivity to the range of qualities that music can achieve by establishing a tonic and selecting an ensemble of notes to serve as a scale relative to that tonic. We conclude that the name "scale-sensitivity" proposed by Dean and Chubb (2017) for this cognitive resource seems to be appropriate.

Acknowledgements. This work is supported by the Center for Hearing Research at the University of California, Irvine and NIH Training Grant T32-DC010775.

References

Blechner, M. J. (1977). *Musical skill and the categorical perception of harmonic mode* (Unpublished doctoral dissertation, Yale University.).

Chubb, C., Dickson, C. A., Dean, T., Fagan, C., Mann, D. S., Wright, C. E., ... & Kowalsky, E. (2013). Bimodal distribution of performance in discriminating major/minor modes. *The Journal of the Acoustical Society of America*, *134*(4), 3067-3078.

Crowder, R. G. (1984). Perception of the major/minor distinction: I. Historical and theoretical foundations. *Psychomusicology: A Journal of Research in Music Cognition*, *4*(1-2), 3.

Crowder, R. G. (1985a). Perception of the major/minor distinction: II. Experimental investigations. *Psychomusicology: A Journal of Research in Music Cognition*, *5*(1-2), 3.

Crowder, R. G. (1985b). Perception of the major/minor distinction: III. Hedonic, musical, and affective discriminations. *Bulletin of the Psychonomic Society*, *23*(4), 314-316.

Dean, T., & Chubb, C. (2017). Scale-sensitivity: A cognitive resource basic to music perception. *The Journal of the Acoustical Society of America*, *142*(3), 1432-1440.

Gagnon, L., & Peretz, I. (2003). Mode and tempo relative contributions to "happy-sad" judgements in equitone melodies. *Cognition & Emotion*, *17*(1), 25-40.

Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception: An Interdisciplinary Journal*, *12*(3), 279-290.

Halpern, A. R. (1984). Perception of structure in novel music. *Memory & cognition*, *12*(2), 163-170.

Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1998). Perception of mode, rhythm, and contour in unfamiliar melodies: Effects of age and experience. *Music Perception: An Interdisciplinary Journal*, *15*(4), 335-355.

Heinlein, C. P. (1928). The affective characters of the major and minor modes in music. *Journal of comparative Psychology*, *8*(2), 101.

Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, *47*(1), 103-118.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception: An Interdisciplinary Journal*, *8*(2), 189-201.

Leaver, A. M., & Halpern, A. R. (2004). Effects of training and melodic features on mode perception. *Music Perception: An Interdisciplinary Journal, 22*(1), 117-143.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin*, *98*(1), 185.

Temperley, D., & Tan, D. (2013). Emotional connotations of diatonic modes. *Music Perception: An Interdisciplinary Journal*, *30*(3), 237-257.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Genre Indication as a Tool for Musical Signification

Pablo Mendoza-Halliday

Doctorado en Música, UNAM, Mexico

pmholly@gmail.com

Abstract

Musical genre is a form of categorization that groups musical entities, such as musical works, which share affinity criteria. However, when we listen to music or we think of it, we can categorize it in many ways depending on the context and the background of our experiences. One way to stabilize the categorization of music is through a genre indication: a textual information that explicitly states a category that refers to a musical entity. Genre indications can work as a form of hermeneutic guide to music: it guides toward attitudes, conjectures and expectations around the cultural artefact or the musical fact as social phenomena. This study analyzes the function of genre indication as a sign in music and its possibilities as a tool for musical signification. To achieve this, it is necessary to understand the structure of genres as musical categories; to identify the possible representations of the genre by means of paratexts, and to recognize the function of this sign of music in the pragmatic context of music listening. The model presented in this paper gives an account of the way genre indications influence the perception of music by creating expectations and meanings. The analytical tools provide arguments that composers, musicologists, producers, or critics can use to think about the way they want the listeners to perceive the music they share.

I. Introduction

Categorization is cognitively inevitable for human beings. Music is turned into categories when we refer to what we hear. In this way, to understand music is not only to understand the real acoustic and social phenomena but also the way categories put them in our minds.

Musical genre is the form of categorization we use to refer to kinds of musical entities—such as musical works. Genres help us to apprehend music explaining its characteristics so that the most relevant are highlighted and then we can relate the musical entity with others that share the same characteristics and differentiate it from the rest of the universal corpus.

Music's meanings change according to its context and to the personal relationship, and these meanings affect the form music is categorized. But also, the categories conventionalized in the musical cultures affect the way we understand music and how we interact with it. Therefore, there is a reciprocal relationship between music signification and genre signification.

The representations of music can divert depending in how they operate in each mind. A way to find agreements between individuals is to share the same representations using genre indications. These are texts that explicitly state a category that refers to a musical entity. They can be manifested in a written or oral discourse about music—a musical paratext. By being shared information about music that can affect the way it is comprehended, genre indications work as hermeneutic guides to music. In this paper, I analyze the way genres works as a form of categorization and as signs of music to explain its hermeneutical potential.

II. Categorization and Classification of Music

The categories in which we think are conceived as *cognitive categories.* For several decades, cognitive sciences have explored the properties of this kind of categories. Actual theories say that the structure of these categories is based on prototypes—idealizations of what would be a hypothetical perfect example of the category. The inclusion of a member is not based in necessary and sufficient conditions, but instead the mind works with family resemblances. This means that something—such as a piece of music—is included in a category—such as a genre—if it resembles to the prototype or to the other members in some way, not necessary by sharing all the same attributes.

However, every mind works differently according to its own capacities, experiences, context and cultural values. Therefore, no piece of music belongs tacitly to some genre. But, on another hand, musical taxonomies work with genre definitions, and definitions propose a model of necessary and sufficient conditions of inclusion to the category. Does this mean that one model is right and the other wrong? Not necessarily. Cognitive categories and taxonomic classes are just stages of a more complex dynamics that reflects in greater depth the nature of categories.

Unlike cognitive categories, taxonomic classes do not depend on cognitive capacities, experience, and context, since they are based on stable conceptual systems. They are not the categories that exist on our mind but rather they exist as logical sets. By the nature of music, these sets can be better explained with fuzzy logic.

Musical genres can be cognitive categories and at the same time taxonomic classes because these two operate at different instances. However, taxonomies exist since someone previously has "thought" in those categories. If cognitive categories precede taxonomic classes, they need to stabilize their conceptual system, and this is achieved by means of social conventions. These conventions, as cultural units, are another form of categorization. Cultural units are not the categories in our individual minds nor the abstract logical sets; they are intersubjective categories established by semiotic codes within a musical community.

With these three stages of categorization we can think of three types of categories. The first type are categories that are exclusively cognitive. We call them *ad hoc* categories for they arise by specific cognitive objectives. They do not need to be (or cannot be) socialized or remain in the memory. Usually these categories do not have labels and it is difficult to name them. An example could be "the kind of music my neighbor plays that is not familiar to me".

A second type of category reaches the state of social convention, but without some power that institutionalizes a definition, the category is volatile and is not established as a taxonomic class. We call them *de facto* categories for they exist in social environments but are context-dependent. An example could be "the music played at weddings". The third type of category are the ones that, through definitions, become taxonomic classes. We call them official categories because the definitions are a byproduct of the institutionalization of knowledge. "Official" genres, such as "reggae", could be examples of this.

These types of categories are better explained as a dynamic process than as immanent conceptual objects. Cognitive categories, social conventions, and taxonomic classes impinge on each other, so they are constantly changing through time. Figure 1 explains these dynamics.

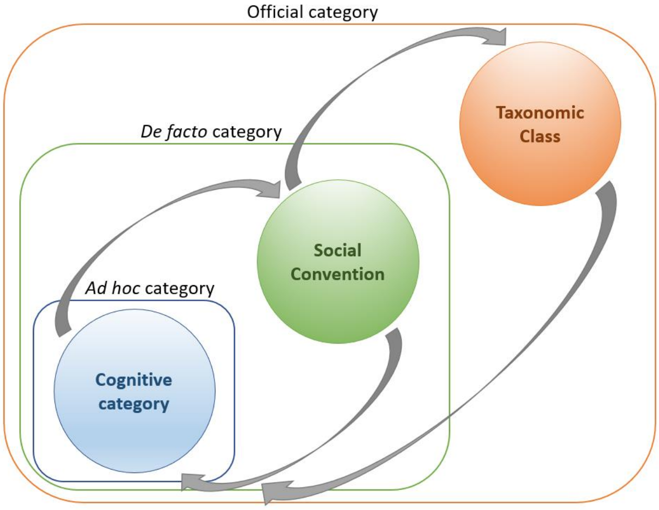


Figure 1. Types of categories

Traditionally, the musical genre is understood as an official category—one that is described in a music taxonomy. However, musicologists interested in theorizing about musical genre, such as Fabbri (2012), have also considered *de facto* categories as music genres. I claim that, as categories of musical entities, musical genres work at all stages of categorization; therefore, a genre may adopt any type of category, even *ad hoc*. There may be a difference in the social function between *ad hoc* categories and the others, but cognitively they have the same purpose of organizing the information and facilitating its comprehension.

III. The Signs of Music and Genre

We can relate music and genre as signs using the semiotic theory of C.S. Peirce. For Peirce, every sign has a triadic relation that involves a sign vehicle, the object referred, and an interpretant—the understanding reached in the relation between the previous, which is itself another sing. In musical semiotics, a musical entity can be the sign vehicle that refers to a sensitive experience as its object. One of the possible interpretants of the sign is the categorization of such entity: the musical genre. As another sign that also refers to the same sensitive experience, the genre produces its own interpretants, for example, the salient attributes of that kind of music.

The sign also works in the other direction. The genre can be interpreted in terms of musical experience. The figure below expresses the triadic form of the sign and the chain of interpretants.

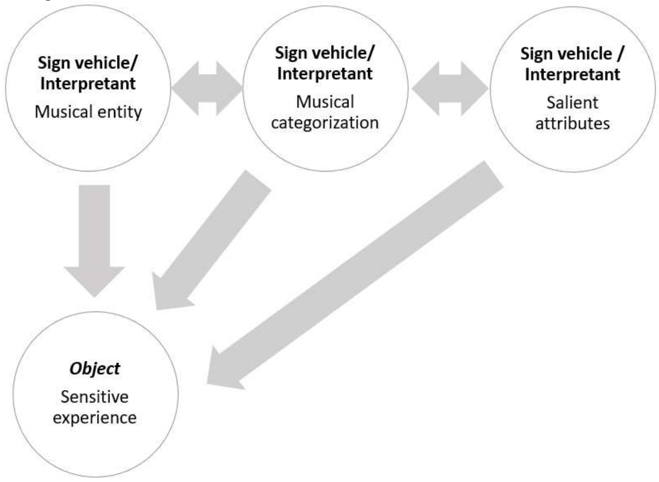


Figure 2. The signs of music and genre

If we analyze these signs within the pragmatic dimension of music listening, we find that the genre indication plays a decisive role in the direction of the semiotic process. When listeners have access to an information that tells them what kind of music they are going to listen, the indicated genre is a sign that represents music. Conversely, when listeners do not have access to that information, they categorize the music according to their listening competencies. In this case, it is music the sign that represents the genre. Being created in the mind of each listener, this interpretant can adopt any type of category: it can be an *ad hoc* genre, a *de facto* genre, or an official genre.

When the genre is indicated, the interpretation of the sign depends on the listeners' actual knowledge of it, or their familiarity with that kind of music. If the listeners are not familiarized with that kind of music, the interpretants of the genre sign will be the meanings of the linguistic category. These meanings will seek a correspondence with the musical attributes.

This can be an advantage if the interpreted meanings have a real relationship with the music. For example, if someone is not familiarized with non-western music and is told that he will hear a Javanese gamelan, probably the category "gamelan" will not mean something concrete or will be interpreted as something exotic, but if he knows what Java is, there will be a "correct" interpretation as music that comes from that island. However, if the interpreted meanings do not have a real relationship with the music, probably the genre indication will mislead the understanding of the music. For example, if the music to be heard is referred as "jungle music", and this label is understood as something wild, or coming from a rainforest, then it will divert the comprehension of its origin in the Jamaican junglists. In those cases, the expectations created by the genre indication will just be truncated during the audition or they will seek to be satisfied with any possible associations with the musical attributes, even if they are not actual facts.

On the other hand, if the listeners are familiarized with the genre, the interpretants of the sign will be the attributes outlined in the genre's concept, according to the current state of the convention that operates in their culture. This will create expectations of correspondence with the attributes of the referred music. Figure 3 shows how the signs of genre and music are related in this scenario.

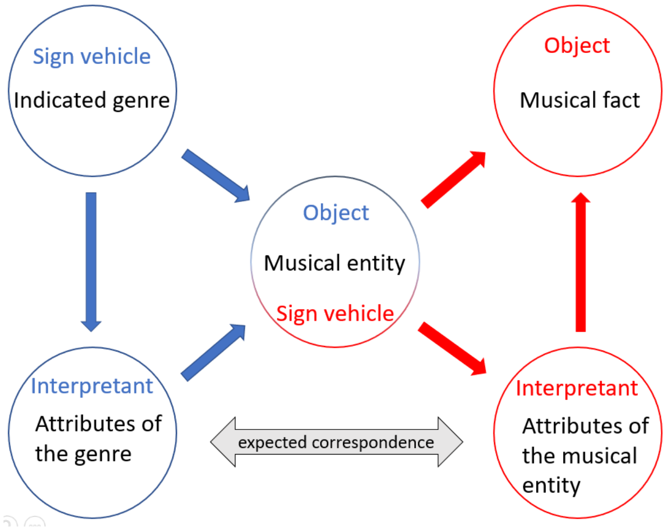


Figure 3. Expectations created by the genre indication

Although the main objective of the genre indication is to lead to the satisfaction of those expectations, it is not the only one. Genre indications can also be used to create surprise, leading to unexpected outcomes that can end in a positive response.

IV. Conclusion

When we give genre indications we generate expectations that have an impact on the way music is listened, the way we behave towards music, and the way music is judged. Furthermore, as socialized signs of music, genre indications have the possibility to create interpretation habits. In this way, the sign has an impact on music listening not only when it is present but also when it stays on memory or when it works tacitly in the culture.

Those interpretation habits also transform the categories. An official genre is what it is because, first, someone thought the category, and then, it was conventionalized within some musical community by means of genre indications. This means that if *de facto* genres—and even *ad hoc*—are used in musical paratexts, the conceptual model can be conventionalized so to be transformed into official genres. Therefore, the use of genre indications is not just something circumstantial in the musical fact. If we understand the nature of genres—and categories in general—we can take advantage of its possibilities to impact music listening.

An official genre should not be understood as the "true" genre, but as the one conventionalized in a certain musical community, in a certain historical moment. Moreover, since musical entities do not have a unique classification, they can be indicated with different types of genres. It can be other official genres that highlight other attributes of the musical entity. It can also be *de facto* genres, according to what members of a musical community think that it represents, regardless if this is musicologicaly or "officially" valid. It can even be *ad hoc* genres, considering that this would be a particular perspective that does not represent a social agreement of what the musical entity is, but this does not mean that it is wrong or invalid.

If we identify the type of category in a genre indication and recognize its semiotic properties, we can evaluate its possibilities as a hermeneutical guide for music. With more complex semiotic systems we can foresee the way they can influence on music's meanings and, therefore, on the way we use music. Composers, musicologists, producers, or critics can use these arguments to think about the way they want listeners to perceive the music they share. The use or the avoidance of genre indications is essential to that goal.

References

Barsalou, L. (1983). Ad hoc categories. *Memory & Cognition,* 11 (3), 211-223.

Brackett, D. (2002). (In search of) musical meaning: genres, categories and crossover. In D. Hesmondhalgh & K. Negus (Eds.). *Popular Music Studies* (pp. 65–83). London: Arnold.

Eco, U. (1999). *Kant and the Platypus. Essays on Language and Cognition.* New York: Harcourt Brace and Company.

Eco, U. (1976). *A Theory of Semiotics*. Bloomington: Indiana University Press.

Fabbri, F. (1981). A theory of musical genres: two applications. In D. Horn and P. Tagg (Eds.). *Popular Music Perspectives.* Göteborg & Exeter: IASPM.

Fabbri, F. (1999). Browsing Music Spaces: Categories And The Musical Mind. Paper presented at the *3rd Triennial British Musicological Societies' Conference*, University of Surrey, Guildford, UK.

Fabbri, F. (2012). How Genres Are Born, Change, Die: Conventions, Communities and Diachronic Processes. In: S. Hawkins (Ed.). *Critical Musicological Reflections. Essays in honour of Derek B. Scott* (pp. 179–192). Oxon, U.K.: Routledge.

Fabbri, F. (2014, December). Music Taxonomies: an Overview. Paper presented as a keynote at the conference "Musique Savante/Musiques Actuelles: Articulations. JAM 2014: Journées d'analyse musicale 2014 de la Sfam (Societé française d'Analyse Musicale)". Salle Igor Stravinsky, Ircam, Paris.

Hatten, R. (1994). *Musical Meaning in Beethoven. Markedness, Correlation, and Intepretation*. Bloomington: Indiana University Press.

Huron, D. (2006). *Sweet Anticipation. Music and the Psychology of Expectation.* Cambridge, MA: MIT Press.

Jakob, E.K. (2004). Classification and Categorization: A Difference that Makes a Difference. *Library Trends,* 52 (3), 515–540.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things, What Categories Reveal about the Mind*. Chicago: The University of Chicago Press.

Miller, C.R. (1984). Genre as Social Action. *Quarterly Journal of Speech,* 70, 151–167.

Monelle, R. (1992). *Linguistics and semiotics in music, Contemporary music studies*. Philadelphia: Harwood Academic.

Peirce, C.S. (1931). *Collected Papers*. Cambridge, MA: Harvard University Press.

Tagg, P. (2012). *Music's Meanings. A modern musicology for non-musos*. New York & Huddersfield: Mass Media Music Scholars' Press.

Varela, F., E. Thompson & E. Rosch. (1992). *De cuerpo presente. Las ciencias cognitivas y la experiencia humana.* Barcelona: Gedisa.

Zbikowski, L. (2002). *Conceptualizing music: cognitive structure, theory, and analysis. AMS studies in music*. Oxford: Oxford University Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Stereotypes of Listeners and Producers of Different Music Genres

Kenneth Michniewicz and Laura Edelman
Psychology Department, Muhlenberg College, USA
kennethmichniewicz@muhlenberg.edu, lauraedelman@muhlenberg.edu

Abstract

We explored the stereotypes concerning social class and motivation for both listeners and producers of different music genres. We used the 14 genres that Rentfrow and Gosling (2007) suggest as the best level of analysis for understanding how music preferences relate to stereotypes of music listeners. One hundred and twenty participants were asked to rate the listeners of each genre on the degree to which they were part of a privileged group, were wealthy, listened in order to influence other peoples' impressions of them, and intrinsically enjoyed the music. Participants were asked to rate the producers of each genre on similar factors. Our results for privilege and wealth of listeners are the same as previous findings for social class. Classical, rap, and religious music were rated high on both impression management and also intrinsic enjoyment. The relationship between these two rating may be based on how tightly associated these three genres are to specific social classes. There was no overlap of genres for low ratings on these two measures. For music producers the patterns were very different. The producers coming from privilege were expected to produce classical, pop, and soundtracks. Pop and soundtracks are associated with greed, pop with low intrinsic enjoyment, but classical music is associated with high intrinsic enjoyment. Low ratings of privilege were associated with blues, rap, and soul. Blues and soul have low rating for greed but rap has high rating for greed. For producers, greed and intrinsic enjoyment seem to be opposing motivations.

Introduction

Previous research indicates fairly clear stereotypes about the personality and other characteristics of people who listen to certain genres of music (North & Hargreaves, 1999; Rentfrow & Gosling, 2003). Hargreaves and North (1999; 2007) suggest that music preferences reflect one' social identity and can be used to indicate social group membership such that specific music preferences are associated with "high culture" and other preferences with "low culture." Rentfrow, McDonald, and Oldmeadow (2009) showed that ethnicity and social class were associated with specific music preferences; specifically, Rentfrow and Gosling (2007) found that prototypical fans of classical music tended to be White and from upper middle class socioeconomic backgrounds, whereas prototypical rap music fans were Black and from low-income and working-class backgrounds. The current study was designed to expand on our understanding of music stereotypes by comparing music listeners with music producers. We also explored motivations for listening to or producing specific genres of music. We compared the listeners' and producers' intrinsic and extrinsic motivations. Intrinsic motivation for both listeners and writers was defined as their own enjoyment of the genre of music. Based on Rentfrow and Gosling's (2003) findings that people prefer certain music styles to not only confirm a person's self-identity but also to determine how others perceive them, we defined perceived extrinsic motivation for listeners as the extent to which their listening to the genre served to influence other people's impressions of them (impression management). Extrinsic motivation of producers was defined as their desire to make money (greed). We finally assessed the degree to which those listening to or producing each genre were expected to have relatively high or low privilege in society (privilege).

We explored the stereotypes concerning social class and motivation for both listeners and producers of different music genres. We hypothesized that there would be differences in both social class and motivations for the listeners and producers of different genres of music.

Method

Participants

One hundred and three participants total (age *Md* = 25) participated in the study with 71 receiving $0.50 through Amazon's Mechanical Turk (www.mturk.com), and 52 receiving course credit in an undergraduate Psychology course at a small liberal arts institution, in exchange for participation. Participants described themselves as White (80.3%), Black (5.5%), Asian (6.3%), Hispanic or Latinx (8.7%), or another non-identified race or ethnicity (3.1%). No data on participant gender was recorded.

Design

A 2 x 4 x 14 within-subjects design was used to assess stereotypes. The independent variables included evaluated target (listener or producer), evaluated characteristic (privilege, wealth, intrinsic motivation, and extrinsic motivation), and music genre. The dependent variable was the rating on each measure.

Materials and Procedure

An online survey of 112 stereotype questions was developed. Each participant evaluated listeners and producers on items pertaining to privilege, wealth, intrinsic motivation, and extrinsic motivation separately for each of 14 genres of music. We used the 14 genres that Rentfrow and Gosling (2007) suggest as the best level of analysis for understanding how music preferences relate to stereotypes of music listeners. Participants rated their agreement with each item using a scale that ranged from 1 (*Not at All*) to 9 (*Extremely*). The order of each music genre was counterbalanced. After evaluating listeners and producers of all 14 genres, participants completed several demographic measures before receiving a debriefing statement.

Results

Privilege and wealth were correlated for all genres and so only privilege was used in further analyses. A 2 x 14 within-subjects analysis of variance was completed for each question. For ratings of privilege, producers were rated overall as more privileged (M = 5.08, SD = 1.31), F(1,120) = 33.16, p < .01, η^2 = .22. There was a statistically significant interaction between evaluated target and music genre, F(13,1560) = 8.38, p < .01, η^2 = .06 (see Figure 1). Fisher's LSD post-hoc tests show that producers were rated significantly more likely to be privileged than listeners for country, pop, and soundtracks (all ps < .01).

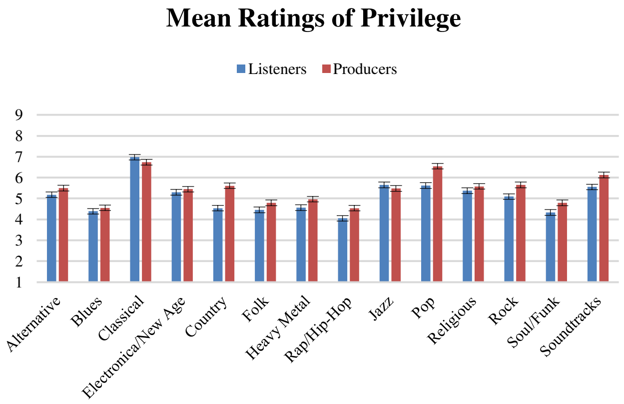


Figure 1. Mean ratings of privilege by evaluated target and music genre. Error bars span two standard errors.

Producers were also rated higher on extrinsic motivation (M = 4.84, SD = 2.23) than listeners, (M = 4.42, SD = 2.14), F(1,115) = 13.77, p < .01, η^2 = .11. There was a statistically significant interaction between evaluated target and music genre, F(13,1495) = 22.93, p < .01, η^2 = .17, see Figure 2. Fisher's LSD post-hoc tests show that producers were rated significantly more likely to have an extrinsic motivation than listeners for alternative, electronica/new age, country, pop, rap/hip-hop, rock, soul/funk, and soundtracks (all ps < .01). However, listeners were rated as having higher extrinsic motivation for classical and religious music (both ps < .01).

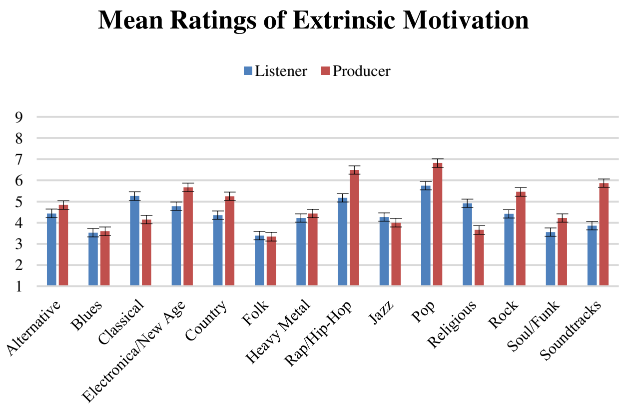


Figure 2. Mean ratings of external motivation for listeners (impression management) and producers (greed) as a function of music genre. Error bars span two standard errors.

Producers were also rated higher on intrinsic motivation (M = 6.69, SD = 1.85), than listeners, (M = 5.24, SD = 2.37), F(1,118) = 66.64, p < .01, η^2 = .36. There was a statistically significant interaction between evaluated target and music genre, F(13,1534) = 10.16, p < .01, η^2 = .08, see Figure 3. Fisher's LSD post-hoc tests show that producers were rated significantly more likely to have an internal motivation than listeners for all genres except pop (all ps < .01). For pop music, the listeners and producers were not rated differently on intrinsic motivation.

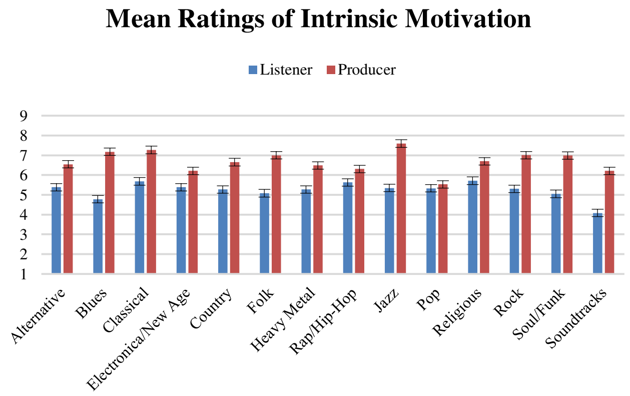


Figure 3. Mean ratings of intrinsic motivation by evaluated target and music genre. Error bars span two standard errors.

Discussion

While previous research has demonstrated that music genres are associated with stereotypes such as socioeconomic status, race, or ethnicity (e.g., Rentfrow & Gosling, 2007), the current study sought to broaden our understanding of these relationships. Specifically, we explored whether or not those who listened to, compared to those who produced, different genres of music were perceived differently on the privilege held in society as well as their internal and extrinsic motivation for either choosing to listen to or write each genre.

With regard to privilege, our findings parallel previous research (North & Hargreaves, 1999, 2007; Rentfrow & Gosling, 2007), in that classical and pop music were associated with relatively more privilege, whereas blues, country, folk, and rap/hip-hop genres were associated with relatively less privilege. Given the likely association of privilege with relevant characteristics (e.g., socioeconomic status), these findings seem consistent with and may in fact replicate prior research.

In addition, producers of music almost always received higher evaluations on all three characteristics than listeners. While additional research is needed to address the mechanism for this, we propose that the time commitment involved in writing or producing any music, compared to simply listening to this music, may underlie a relationship between these three measures. Because successful song composition presumably requires skill and training, people may expect those who pursue a career producing a certain music genre to have made a substantial investment in doing so. This investment, in turn, would require a justification in the form of either genuine enjoyment (intrinsic motivation) or expected affluence (greed and, consequently, privilege). Future research should examine the role of expected time commitment, both prior to becoming a musician (e.g., education and training), and during the composition process (i.e., song writing), in this relationship for music producers.

Within evaluation measurements, producers specifically were also rated higher on intrinsic motivation than extrinsic motivation on all genres except for pop (for which there was on difference). While there exist genre-specific exceptions, future research should examine the financial prospects or likelihood of success for an aspiring musician compared to other possible career trajectories. Accordingly, people may simply expect music to be less promising or lucrative and thus substitute the intrinsic motivation in place of the extrinsic one. Additionally, we did not inform participants about the context in which someone decides to produce music: we suspect that evaluations of motives differ when someone composes music as their career relative to when they do so as a hobby. Still, future research should examine if these factors yield different evaluations of music producers.

By comparison, listeners were rated as having similar levels of internal and external motivation for all genres. These findings support previous research showing that people have clear stereotypes about people based on their musical preferences (Renfrow & Gosling, 2007). These stereotypes include the motivations that drive individuals to listen to or write specific genres of music.

In light of these findings, we offer two broad, potentially promising areas of future research. First, in an effort to assess "pure" stereotypes regarding these music genres and those who compose or listen to them, we offered virtually no information about the context in which these evaluations were provided. Similarly, our observations of the stereotypes provided were fairly limited. For example, while previous research highlights the relationship between certain genres and stereotypes regarding race, ethnicity, or socioeconomic status (Rentfrow & Gosling, 2003), we remain uncertain if these associations would strengthen or weaken as a function of whether or not the target evaluated is composing or listening to the music. Thus, future research should examine the stereotypes that participants automatically conjure in more depth.

Additionally, future research should distinguish between the nature of the stereotypes and the subsequent consequences. For example, other researchers distinguish between stereotypes that merely describe what a group is like (i.e., descriptive stereotypes), and stereotypes that describe what groups *should* be like (i.e., prescriptive stereotypes; Rudman & Glick, 2001). This research accordingly proposes that these prescriptive stereotypes, when violated, promote negative or hostile reactions towards the targets violating the stereotype (e.g., Rudman & Fairchild, 2004). Applying this model to music genres, we may expect that composers or listeners to various genres may receive different evaluations of their motives, and likely varying interpersonal evaluations, as a function of whether or not the stereotypes described here are merely descriptive or are prescriptive.

Despite the limited information that participants received, we still find that people hold stereotypes about music genres and those associated with them. Importantly, these stereotypes strongly influence impressions of those who either listen to particular music genres or compose them.

References

Hargreaves, D. J., & North, A. C. (1999). The functions of music in everyday life: Redefining the social in music psychology. *Psychology of Music, 27*(1), 71-83. doi:10.1177/0305735699271007

North, A. C., & Hargreaves, D. J. (2007). Lifestyle correlates of musical preference: 1. Relationships, living arrangements, beliefs, and crime. *Psychology of Music, 35*(1), 58-87. doi:10.1177/0305735607068888

North, A. C., & Hargreaves, D. J. (2007). Lifestyle correlates of musical preference: 2. Media, leisure time and music. *Psychology of Music, 35*(2), 179-200. doi:10.1177/0305735607070302

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology, 84*(6), 1236-1256. doi:10.1037/0022-3514.84.6.1236

Rentfrow, P. J., & Gosling, S. D. (2007). The content and validity of music-genre stereotypes among college students. *Psychology of Music, 35*(2), 306-326. doi:10.1177/0305735607070382

Rentfrow, P. J., McDonald, J. A., & Oldmeadow, J. A. (2009). You are what you listen to: Young people's stereotypes about music fans. *Group Processes & Intergroup Relations, 12*(3), 329-344. doi:10.1177/1368430209102845

Rudman, L. A., & Fairchild, K. (2004). Reactions to counterstereotypic behavior: the role of backlash in cultural stereotype maintenance. *Journal of Personality and Social Psychology, 87*(2), 157. doi: 10.1037/0022-3514.87.2.157

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*(4), 743-762. doi: 10.1111/0022-4537.00239

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Entropy and acceptability: information dynamics and music acceptance

Lorena Mihelač,[1] Geraint A. Wiggins,[2,3] Nada Lavrač,[4] Janez Povh,[5,6]

[1]*ŠC Novo mesto & International Postgraduate School Jožef Stefan, Ljubljana, Slovenia*
[2]*AI Lab, Vrije Universiteit Brussel, Belgium*
[3]*School of Electronic Engineering Computer Science, Queen Mary University of London, UK*
[4]*Jožef Stefan Institute Ljubljana, Slovenia*
[5]*Faculty of Mechanical engineering Ljubljana, Slovenia*
[6]*Institute of mathematics, physics and mechanics Ljubljana, Slovenia*
[1]lorena.mihelac@sc-nm.si, [2]geraint.wiggins@vub.be, [4]Nada.lavrac@ijs.si, [5]janez.povh@fs.uni-lj.si

Abstract

Shannon's Information Theory, developed to determine the optimal encoding of messages for radio transmission, has also been adopted as an analytical tool in music cognition. Early research in this area was mostly focused to measuring the entropy of musical styles rather than that of individual musical elements (melody, rhythm, harmony). However, detailed modelling of these elements shows that complex combinations of them influence the entropy of music. As harmony plays a fundamental role in all musical styles, and is one of the foundations of Western tonal music, it is important to explore how the entropy of harmony affects the listener's acceptance of music. Studies from Mihelač (2017) and Mihelač and Povh (2017) have shown that the prediction of the acceptance of a musical piece can be obtained by analysing the entropy of the harmony, as the entropy of the harmonic flow correlates with the acceptance (pleasantness) of the music piece. In this paper we have extended the empirical study from Mihelač and Mihelač & Povh by analysing the harmonic flow in depth. We used the same data and evaluated it twice by 20 new evaluators from secondary vocational school (10 with and 10 without musical knowledge). Our new results are aligned with findings of the previous research: the effect of musical style on perception of listening difficulty is significant. Listening difficulty significantly decreases on the second hearing of a piece, while pleasantness, recognition and repeatability significantly increase, with one exception: participants without musical knowledge showed no significant difference in mean repeatability. Furthermore, our results suggest that subliminal irregularities in the harmonic flow affect the listeners' acceptance of music.

Introduction

How can we understand the laws of attraction in order to predict the human preference for a particular music and to predict the success of a song (or any artefact)? According to Blume (2011), there is no prediction of a hit-song as everything is just "a matter of technique". According to the famous and highly successful Hollywood screenwriter William Goldman (Case, 2016), "nobody knows anything".

The real message of these two statements is that it is very difficult to predict the acceptance of a piece of music. We have to deal (in the process of the prediction) with a complex combination of micro-emotions that are related to our personal history, to the special features of the song and to some elements, which are still not quite understandable (Pachet, 2011).

In spite of this (seemingly) unsolvable task, there have been some efforts and steps forward in predicting the acceptance and success of a song automatically in the field of Hit Song Science. Unfortunately, the exact mechanisms behind these predictions are not clearly disclosed, and very little reproducible data is provided to check the accuracy of these predictions (Li, 2011; Herremans et al., 2000).

Proceeding from the idea, that each musical piece is information (Madsen & Widmer, 2006), the empirical studies from Mihelač (Mihelač, 2017) and Mihelač & Povh (2017) have used the measurement of entropy in explaining the complexity of musical dimensions and the listener's acceptability of music. Results from these studies have shown that the prediction of the acceptance of a musical piece can be obtained by analyzing the entropy (measure of uncertainty) of the harmony, as the entropy of the harmonic flow correlates with the acceptance (pleasantness) of a musical piece.

The results of these studies have shown, furthermore, that the musical structure organized in a particular way in some musical pieces is affecting the listener's understanding. By evaluating the same data twice in our new research with twenty new evaluators, we examine these pieces with an in-depth analysis of their harmonic flow, in order to define the irregularities and their impact on the listener's acceptance/rejection of a musical piece. As the irregularities seem to be perceived by the evaluator without being aware of it, as "subliminal" (Ramsøy & Overgaard, 2004), affecting the pleasantness and the feeling of complexity, we refer to these irregularities hereafter as "subliminal irregularities".

Method

Evaluators

Twenty evaluators (N=20), ten with musical knowledge (more than five years of formal musical training) and 10 without musical knowledge, participated in the evaluation of the data in both evaluations; all of the evaluators were 20 females, between 15-16 years old ($M = 15.4$, $SD = 0.49$).

Music stimuli

Data from Mihelač (2017) and Mihelač & Povh (2017) has been used in both evaluations. The data consists of 160 musical pieces from different musical styles, from baroque until the 20th century and is between 14 *s* and 18 *s* in duration, adjusted to an equal loudness level.

In the current research, twenty evaluators evaluated the same 160 musical pieces twice, each by the same four main criteria on a Likert scale from one to five. The criteria were: a) the difficulty of listening to the musical piece, b) the pleasantness perceived while listening to the musical piece, c) the recognition of the musical piece and d) the repeatability of the musical piece (the readiness of the evaluator to listen to the whole music piece, not only to a part of it).

Procedure

The open source software Moodle was used for evaluating the musical pieces in the first and second listening session. In the second evaluation, performed 10 days after the first evaluation, the order of musical pieces was randomized to avoid response bias. For each evaluator, two questionnaires were prepared (for first and second evaluation) with a compulsory login procedure on behalf of his unique ID. Each evaluation was time limited; however, each evaluator had two days available to complete the task due to the demanding listening to the complete data. In analysing the data, we have used One-way ANOVA, paired t-test and two sample t-test.

Results

Paired-samples t-tests were conducted to compare the first and second evaluation of the same 160 musical pieces by 20 evaluators (10 with musical knowledge and 10 without musical knowledge). Four main variables: difficultness, pleasantness, repeatability and recognizability were evaluated. Firstly, we compared the average values for the difficultness, pleasantness, repeatability and recognizability of each song obtained on the first and on the second evaluation. In Table 1 we show results of paired-samples t-test for the difficulty, pleasantness, repeatability and recognizability of listening to the musical piece. We can see that the average of the difficulty is on the second evaluation slightly (but still statistically significantly) lower compared to the first evaluation. For the other three variables, we observe significant increase of the mean values.

Table 1. Results of paired t-test (all evaluators)

variable	Difficulty	pleasantness
1st ev.	M = 1.75, SD = 0.482	M = 3.58, SD = 0.608
2nd ev.	M = 1.61, SD = 0.316	M = 3.77, SD = 0.316
	t(159) = 7.06 p = 4.76e-11	t(159) = -8.22 p = 6.889e-14
	Repeatability	**recognizability**
1st ev.	M = 3.44, SD = 0.652	M = 2.67, SD = 1.13
2nd ev.	M = 3.52, SD = 0.579	M = 3.09, SD = 0.941
	t(159) = -3.01 p = 0.003	t(159) = -13.58 p = 2.2e-16

Additionally, we repeated these tests for the average values of difficulty, pleasantness, repeatability and recognizability, obtained only by evaluators with and without music knowledge. Results are in Tables 2-3.

Table 2. Results of paired t-test (evaluators with musical knowledge)

variable	difficulty	pleasantness
1st ev.	M = 1.69, SD = 0.386	M = 3.71, SD = 0.616
2nd ev.	M = 1.51, SD = 0.269	M = 3.95, SD = 0.494
	t(159) = 7.39 p = 7.784e-12	t(159) = -10.36 p = 2.2e-16
	repeatability	**recognizability**
1st ev.	M = 3.71, SD = 0.615	M = 2.90, SD = 1.13
2nd ev.	M = 3.82, SD = 0.615	M = 3.33, SD = 0.970
	t(159) = -4.67 p = 6.331e-06	t(159) = -13.28 p = 2.2e-16

Table 3. Results of paired t-test (evaluators without musical knowledge)

variable	difficulty	pleasantness
1st ev.	M = 1.92, SD = 0.452	M = 3.37, SD = 0.567
2nd ev.	M = 1.70, SD = 0.467	M = 3.60, SD = 0.566
	t(159) = 7.59 p =2.557e-12	t(159) = -10.36 p = 2.2e-16
	repeatability	**recognizability**
1st ev.	M = 3.14, SD = 0.664	M = 2.50, SD = 0.978
2nd ev.	M = 3.21, SD = 0.664	M = 2.85, SD = 0.994
	t(159) = -1.82 p = 0.070	t(159) = -9.89 p = 2.2e-16

We can see that perception of the difficulty significantly decreases on the second hearing of a piece, while pleasantness, recognition and repeatability significantly increase, with one exception: participants without musical knowledge showed no significant difference in average repeatability.

Our results support Madison and Schiölde (2017): repeated listening can increase participants' liking of music regardless of complexity (entropy).

We have also analysed how the evaluators with and without music knowledge perceived difficulty, pleasantness, repeatability and recognizability. For each musical piece, we compared the average values of both groups of evaluators (10 evaluators with musical knowledge and 10 without musical knowledge) and used paired samples t-test. Tables 4-5 show results for both evaluations.

Table 4. Results of paired t-test (evaluators with and without musical knowledge, 1st evaluation)

variable	difficulty	Pleasantness
with	M = 1.69, SD = 0.386	M = 3.71, SD = 0.616
without	M = 1.92, SD = 0.452	M = 3.37, SD = 0.567
	t(159) = -6.55 p = 7.664-10	t(159) = 11.001 p = 2.2e-16
	repeatability	**recognizability**
with	M = 3.71, SD = 0.672	M = 2.90, SD = 1.13
without	M = 3.142,SD =0.610	M = 2.50, SD = 0.978
	t(159) = 14.5 p = 2.2e-16	t(159) = 9.52 p = 2.2e-16

Table 5. Results of paired t-test (evaluators with and without musical knowledge, 2nd evaluation)

variable	Difficulty	Pleasantness
with	M = 1.51, SD = 0.269	M = 3.95, SD = 0.494
without	M = 1.70, SD = 0.467	M = 3.60, SD = 0.566
	t(159) = -5.64 p = 7.431e-08	t(159) = 10.01 p = 2.2e-16
	repeatability	**recognizability**
with	M = 3.82, SD = 0.615	M = 3.33, SD = 0.970
without	M = 3.21, SD = 0.664	M = 2.85, SD = 0.994
	t(159) = 14.26 p = 2.2e-16	t(159) = 10.76 p = 2.2e-16

We can see that there exist statistically significant differences in perceiving the evaluated musical examples between both groups of evaluators, for both evaluations.

Evaluators with musical knowledge compared to evaluators without musical knowledge are (on average) perceiving the musical pieces as less difficult and more pleasant and are more successful in recognizing the musical pieces. The evaluators with musical knowledge compared to the evaluators without musical knowledge are also more prepared to listen to the entire musical piece.

Although musical knowledge is obtained in different indoor and outdoor activities during the adolescence (North, Hargreaves, & O'Neill, 2000), these findings support the impact of formal musical knowledge on preferences and in understanding the structure of musical elements (Dobrota & Ercegovac, 2014; Moore & Johnson, 2001).

A one-way ANOVA was conducted in both evaluations to compare the impact of musical styles (baroque, beat, classical, disco, epic music, folk, hard rock, jazz, pop, rhythm and blues, rock and romantic) on the perception of the difficulty.

Significance of musical styles on difficultness at the p<.05 level (Table 6) was found in both evaluations.

Table 6. Results of ANOVA (difficulty and musical style)

eval.	1st evaluation	2nd evaluation
All	[F(11,148) = 2.49, p = 0.007]	[F(11,148) = 2.56, p = 0.005]
With	[F(11,148) = 2.08, p = 0.025]	[F(11,148) = 2.39, p = 0.009]
Without	[F(11,148) = 2.61, p = 0.005]	[F(11,148) = 2.24, p = 0.015]

Compared to the results obtained in the previous empirical study (Mihelač, 2017; Mihelač & Povh, 2017), no significance was found in any of the evaluations of musical styles on pleasantness (Table 7).

Table 7. Results of ANOVA (pleasantness and musical style)

eval.	1st evaluation	2nd evaluation
All	[F(11,148) = 1.09, p = 0.37]	[F(11,148) = 1.30, p = 0.232]
With	[F(11,148) = 1.34, p = 0.21]	[F(11,148) = 1.26, p = 0.254]
Without	[F(11,148) = 1.07, p = 0.386]	[F(11,148) = 1.57, p = 0.112]

These results are surprisingly as we are dealing with evaluators, adolescents, aged between 15-16 years old, which prefer to listen to music that reflects their personal characteristics and/or issues they face in their maturing process (Schwartz & Fouts, 2003).

We have extended the previous studies with the analysis of the impact of musical style on repeatability and recognizability.

The results of the one-way ANOVA conducted in both evaluations to compare the impact of musical style on repeatability have shown no significance (Table 8).

Table 8. Results of ANOVA (repeatability and musical style)

eval.	1st evaluation	2nd evaluation
All	[F(11,148) = 1.67, p = 0.087]	[F(11,148) = 1.430, p = 0.163]
With	[F(11,148) = 1.69, p = 0.08]	[F(11,148) = 1.74, p = 0.069]
Without	[F(11,148) = 1.49, p = 0.141]	[F(11,148) = 1.45, p = 0.156]

The results of the one-way ANOVA conducted in both evaluations to compare the impact of musical style on recognizability have shown significance in all groups (Table 9).

Table 9. Results of ANOVA (recognizability and musical style)

eval.	1st evaluation	2nd evaluation
All	[F(11,148) = 2.03, p = 0.029]	[F(11,148) = 1.85, p = 0.05]
With	[F(11,148) = 2.02, p = 0.03]	[F(11,148) = 1.88, p = 0.0467]
Without	[F(11,148) = 2.00, p = 0.0326]	[F(11,148) = 1.88, p = 0.0466]

It seems that some styles are affecting more or less the recognizability of a musical piece due to the music material, which is more/less familiar for the evaluator or more/less complicated to memorize (Jónasson & Lisboa, 2016).

The previous two studies from Mihelač (2017) and Mihelač & Povh (2017) has shown that some musical pieces are perceived as "difficult" by evaluators.

As both evaluations of our recent studies have shown the same results in same musical pieces, the harmonic flow of these pieces was analysed in-depth.

The findings from 53 musical pieces (from 160) suggests that these pieces are not meeting the listener's expectancies. Listeners tend to use a set of basic perceptual principles, which are applied to different musical styles and are depending on the kind of music the listeners are exposed to (Krumhansl, 2004).

If the content of a musical piece does not meet these principles it might be, that the listener is confused and cannot recognise the given information (Edmonds, 1995).

In this case, the complexity, sometimes posited as a mid-point between order and disorder (Grassberger, 1989) seems to be perceived more in the sense of a disorder or "subliminal irregularity" as it is affecting the pleasantness and the feeling of complexity below the threshold of consciousness.

To examine the impact of the subliminal irregularity on the evaluator's acceptance of a musical piece, 160 musical pieces were categorized in two main categories: "regular" (107 musical pieces) and 53 musical pieces defined as "irregular" consisting of "subliminal irregularities".

Two sample t-tests were conducted to examine the impact of subliminal irregularity on the evaluation of the same 160 musical pieces by 20 evaluators (10 with musical knowledge and by 10 without musical knowledge).

The same main four variables (difficultness, pleasantness, repeatability and recognizability) were evaluated. The findings are presented in Table 10-15.

Significance of the impact of subliminal irregularity on all the main four variables (difficulty, pleasantness, repeatability and recognizability) at the p<.05 level was found in both evaluations.

Table 10. Results of t-test (1st evaluation/all evaluators)

variable	difficulty	pleasantness
irregular	M = 2.05, SD = 0.35	M = 3.17, SD = 0.45
regular	M = 1.68, SD = 0.29	M = 3.72, SD = 0.52
	t(158) = 7.01 p =6.459e-11	t(158) = -6.56 p = 7.356e-10
	repeatability	**recognizability**
irregular	M = 3.08, SD = 0.52	M = 2.50, SD = 0.82
regular	M = 3.60, SD = 0.55	M = 2.85, SD = 1.05
	t(158) = -5.69 p = 5.916e-08	t(158) = -3.87 p = 0.0002

Table 11. Results of t-test (2nd evaluation/all evaluators)

variable	difficulty	pleasantness
irregular	M = 1.79, SD = 0.34	M = 3.47, SD = 0.42
regular	M = 1.52, SD = 0.26	M = 3.93, SD = 0.44
	t(158) = 5.49 p = 1.529e-07	t(158) = -6.39 p = 1.735e-09
	repeatability	**recognizability**
irregular	M = 3.16, SD = 0.51	M = 2.67, SD = 0.84
regular	M = 3.69, SD = 0.53	M = 3.3, SD = 0.93
	t(158) = -6.06 p = 9.315e-09	t(158) = -4.14 p = 5.666e-05

Table 12. Results of t-test (1st evaluation/ evaluators "with")

variable	difficulty	pleasantness
irregular	M = 1.92, SD = 0.39	M = 3.3, SD = 0.49
regular	M = 1.57, SD = 0.33	M = 3.92, SD = 0.57
	t(158) = 6.01 p = 1.184e-08	t(158) = -6.70 p = 3.389e-10
	repeatability	**recognizability**
irregular	M = 3.32, SD = 0.6	M = 2.43, SD = 0.96
regular	M = 3.9, SD = 0.62	M = 3.14, SD = 1.14
	t(158) = -5.70 p = 5.861e-08	t(158) = -3.87 p = 0.0002

Table 13. Results of t-test (2nd evaluation/evaluators "with")

variable	difficulty	pleasantness
irregular	M = 1.67, SD = 0.29	M = 3.64, SD = 0.44
regular	M = 1.43, SD = 0.22	M = 4.11, SD = 0.44
	t(158) = 5.83 p = 3.104e-08	t(158) = -6.37 p = 2.016e-09
	repeatability	**recognizability**
irregular	M = 3.48, SD = 0.6	M = 2.96, SD = 0.89
regular	M = 4, SD = 0.55	M = 3.51, SD = 0.96
	t(158) = -5.47 p = 1.703e-07	t(158) = -3.52 p = 0.0006

Table 14. Results of t-test (1st evaluation/evaluators "without")

variable	difficulty	pleasantness
irregular	M = 2.18, SD = 0.47	M = 3.05, SD = 0.48
regular	M = 1.79, SD = 0.39	M = 3.53, SD = 0.54
	t(158) = 5.55 p = 1.169e-07	t(158) = -5.49 p = 1.583e-07
	repeatability	**recognizability**
irregular	M = 2.84, SD = 0.56	M = 2.12, SD = 0.75
regular	M = 3.29, SD = 0.58	M = 2.68, SD = 1.02
	t(158) = -4.68 p = 6.131e-06	t(158) = -3.56 p = 0.0005

Table 15. Results of t-test (2nd evaluation/evaluators "without")

variable	difficulty	pleasantness
irregular	M = 1.53, SD = 0.51	M = 3.29, SD = 0.5
regular	M = 1.61, SD = 0.42	M = 3.75, SD = 0.54
	t(158) = 3.94 p = 0.0001	t(158) = -5.12 p = 8.723e-07
	repeatability	**recognizability**
irregular	M = 2.84, SD = 0.57	M = 2.39, SD = 0.87
regular	M = 3.39, SD = 0.64	M = 3.08, SD = 0.97
	t(158) = -5.29 p = 4.081e-07	t(158) = -4.39 p =2.092e-05

We can see from the results that subliminal irregularity in harmonic flow is a significant factor in explaining the perception of musical pieces in both evaluations and for all groups of evaluators.

However, no significance was found neither in the repeated exposure to subliminal stimuli on the increasement on "liking"/pleasantness as it was found in the studies of Bornstein (1989) and Kunst-Wilson & Zajonc (1980) nor in a more favorable evaluation when the stimuli was recognized by evaluators as it has been reported in the study of Lee (2001).

Conclusion

Two evaluations of the same data used for the empirical research from Mihelač and Mihelač & Povh (2017) have shown that repeated listening to music significantly decreases the difficulty, while pleasantness, recognition and repeatability significantly increase, except in evaluators without musical knowledge where no significant difference in average repeatability was found. Our results support Madison and Schiölde (2017): repeated listening can increase participants' liking of music regardless of complexity (entropy).

Formal musical knowledge is a significant factor in explaining all the main four variables (difficulty, pleasantness, repeatability and recognizability). Evaluators with formal musical knowledge compared to evaluators without it, are experiencing music as more pleasant, recognizable and less difficult, and are more prepared to listen to the entire composition.

Perceived difficultness and recognizability of musical pieces are significantly different for different music styles. This is not the case for pleasantness and repeatability. Our interpretation is that the evaluators are experiencing the entropy differently in different musical styles.

Furthermore, our findings provide additional information about the irregularities in the harmonic flow. The analysis of the harmonic flow of 160 musical pieces casts a new light on the understanding of harmony and suggest consideration of subliminal irregularities in harmonic flow when measuring acceptability of music.

Future research will examine these subliminal irregularities in-depth as it might extend the explanations of the acceptability of music in listeners.

References

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968-1987. *Psychological Bulletin, 106,* 265-289.

Case, C. (2018, June 4), *Nobody Knows Anything (except William Goldman).* Retrieved from https://www.kickstarter.com/projects/1939276694/nobody-knows-anything-except-william-goldman

Conklin, D. & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research, 24,* 51-73.

Dobrota, S., & Reić Ercegovac, I. (2014). The Role of Music Education, Characteristics of Music and Personality Traits. *Croatian Journal of Education, 16* (2), 363-384.

Edmonds, B. (2018, May 2). *What is complexity? The philosophy of complexity* per se *with application to some examples in evolution.* Retrieved from http://cogprints.org/357/4/evolcomp.pdf

Febbres, G. & Jaffe, K. (2018, May 27). *Music viewed by its Entropy content: A novel window for comparative analysis.* Retrieved from https://pdfs.semanticscholar.org/fe36/6a735cd8d905fdcae0b895c896d0e66e708e.pdf

Grassberger, P. (1989). Problems in quantifying self-organized complexity. *Helvetica Physica Acta, 62,* 498-511.

Herremans, D., Martens, D., & Sörensen, K. (2000). Musical preferences during and after relaxation and exercise. *Journal of New Music Research, 43* (3), 291-302.

Jónasson, P., & Lisboa, T. (2016). Shifting the paradigm: Contemporary Music, Curriculum Changes and the Role of Professional Musicians as Researchers. *Proceedings of the 21st International Seminar of the ISME Commission on the Education of the Professional Musician*. (pp. 78-91). UK: University of St. Andrews.

Krumhansl, C. L. (2004). The cognition of tonality – as we know it today. *Journal of New Music Research, 33* (3), 253-268.

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science, 207,* 557-558.

Lee, A. Y. (2001). The mere exposure effect: An uncertainty reduction explanation revisited. *Personality and Social Psychology Bulletin, 27,* 1255-1266.

Li T., Ogihara, M., & Tzanetakis, G. (2011). *Music data mining*. Boca Raton, FLA: CRC Press.

Madison, G. & Schiölde, G. (2017, May 29). *Repeated listening increases the liking for music regardless of its complexity: Implications for the appreciation and aesthetics of music.* Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5374342/

Madsen, S. T., & Widmer, G. (2006). Music complexity measures predicting the listening experience. *Proceedings of the Third International Conference on Music Perception and Cognition* (pp. 57-60). Belgium: Université de Liège.

Margulis, E. H. & Beatty, A. (2008). Musical style, psychoaesthetics, and prospects for entropy as an analytic tool. *Computer Music Journal, 32 (4),* 64-78.

Mihelač, L. (2017). *Napovedovanje slušne sprejemljivosti glasbe na osnovi entropije harmonije [Predicting the acceptability of music with entropy of harmony](Master thesis).* Faculty of information studies, Novo mesto, Slovenia.

Mihelač, L., & Povh, J. (2017). Predicting the acceptability of music with entropy of harmony. In V. Rupnik & L. Bogataj (Ed.), *Proceedings of the 14th International Symposium on Operational Research,* (pp. 371-376). Slovenia: Slovenian Society Informatika, Section for Operational Research.

Moore, R., & Johnson, D. (2001). Effect of Musical Experience on Perception of and Preference for Humor in Western Art Music. *Bulletin of the Council for Research in Music Educcation, 149,* 31-37.

North, A. C., Hargreaves, D. J., & O'Neill, S. A. (2000). The importance of music to adolescents. *The British Journal of educational Psychology, 70* (2), 255-272.

Pachet, F. (2011). Hit Song Science. Pitch, harmony, and neural nets: A psychological perspective. In T. Li, M. Ogihara & G. Tzanetakis (Eds.), *Music and Data Mining.* (pp. 306-314).London: Taylor & Francis Group.

Pearce, M. (2018, May 18). *The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition*. (PhD thesis). Department of Computing, City University, London, London, UK. Retrieved on 26 May 2018 http://openaccess.city.ac.uk/8459/

Ramsøy, T. Z., & Overgaard, M. (2004). Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences, 3* (1), 1-23.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Systems Technical Journal, 27,* 379-423, 623-656.

Schwartz, K. D., & Fouts, G. T. (2003). Music preferences, personality style and developmental issues of adolescents. *Journal of Youth and Adolescence, 32* (3), 205-2013.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Relationships Between Genre Preference, Aural Skills, and Tonal Working Memory

Elizabeth Monzingo,[1] Emily M. Elliott,[2] Daniel Shanahan,[3] David John Baker,[4] Juan Ventura,[5] and Katherine Vukovics[6]

[1,4,5,6] *School of Music, Louisiana State University, USA*
[2] *Department of Psychology, Louisiana State University, USA*
[3] *School of Music, Ohio State University, USA*

[1]emonzi1@lsu.edu, [2]eelliott@lsu.edu, [3]shanahan.37@osu.edu, [4]dbake29@lsu.edu, [5]kvukov1@lsu.edu

Abstract

Musical training and cognitive abilities appear to be related to one another, but why? Recent research has used measures of musical sophistication to predict measures of working memory capacity, especially tonal working memory capacity, and vice versa, but definitive mediators of the relationship have yet to be identified. Musicians may have an advantage over non-musicians due to their likely enrollment in aural skills courses, in which they learn strategies for melodic dictation, a complex tonal working memory task. They may also have an advantage due to an accruement of aural skills implicitly learned through more meaningful engagement with music, or due to engagement with specific types of music. The aim of this paper is to investigate the role of explicitly and implicitly learned aural skills as potential mediators of the relationship between musicianship and working memory capacity. Results suggest that musicians are more likely to engage strategies that encourage deeper levels of processing for tonal working memory tasks than non-musicians and that musicianship and aural skills achievement help predict tonal working memory capacity. Exploratory analysis of genre preference suggests listening to classical music, jazz, or heavy metal correlates to higher tonal WMC, which encourages further research into genre preference. Considering these findings, we suggest that the "musician advantage" in working memory tasks may be found in the acquisition of valuable strategies for decreasing working memory load gained through the pursuit of musical mastery.

Introduction

Students pursuing a degree in music will almost certainly be asked to learn to notate melodies by ear, a task known as melodic dictation. Though the specific strategies and assessment methods incorporated by the instructor will vary, this task generally asks the students to hear a short melody and write it down accurately, given a limited number of play-throughs and a limited amount of technical information (such as the time signature, the clef, etc.). A student's ability to hold the melody accurately in their working memory, both in between and after play-throughs, is beneficial to the successful completion of the task. When viewed through the lens of cognitive science, the melodic dictation task begins to look like a complex assessment of tonal working memory capacity.

The relationship between musical sophistication and cognitive abilities has been the object of research for some time, and recent literature has focused on the particular cognitive abilities related to working memory tasks (see Müllensiefen et al., 2014 & Ollen, 2006). Meinz and Hambrick (2010) found that variance in sight-reading ability could be predicted by measures of working memory capacity (WMC) beyond sight-reading experience or musical training, and Colley et al. (2018) similarly found that an individual's WMC helped account for differences in the ability to tap along to expressive timing in music. However, other researchers have found musical training to contribute significantly to performance on working memory tasks. Slevc et al. (2016) found musical ability to predict better performance on both auditory and visual updating tasks, or tasks that involve the ability to both monitor information continuously and quickly add and remove information from working memory. Swaminathan et al. (2017) similarly found evidence that supports musical aptitude as a contributing factor in predicting individual differences in general fluid intelligence between musicians and non-musicians. Talamini et al. (2017) conducted a meta-analysis to clarify whether musicians perform better than non-musicians in memory tasks, and their findings suggest that musicians do seem to have a large advantage with tonal stimuli in particular.

Whether musical ability provides an advantage in WMC or a highly functioning WMC provides an advantage in musical ability, research supports a relationship between the two mechanisms. The objective of this work is to explore potential mediators of this relationship, regardless of which presupposes which. One potential advantage that musicians have over non-musicians is that they are likely to explicitly learn and develop the skills to accomplish melodic dictation, or in other words, they explicitly practice strategies to hold tonal information in their working memory while they simultaneously apply themselves to the task of writing it down in a specific nomenclature. Another potential advantage is found in what musicians implicitly learn through the music they engage with on a regular basis. It is common for people to listen to music on a daily basis, but musicians also actively play, read, and create music. Perhaps the type of music we engage with, and the way in which we engage with the music, influences our ability to work with tonal information in our working memory.

In this paper, we first investigate the contribution of explicit aural skills learning on tonal WMC. We apply a stepwise hierarchical multiple linear regression to identify whether musical sophistication and, more specifically, aural skills achievement are significant factors in participant success in the tonal working memory task. We also examine the strategies that musicians and non-musicians used to complete the tonal working memory task, and we specifically hypothesize that musicians will employ more complex, and more explicitly musical, strategies than non-musicians.

As Wolf and Kopiez found, the prognostic validity of tests administered to incoming students to gauge their aural skills and music theory mastery are quite low (2014); perhaps there are other measures that can predict academic success better than explicit theory and skills knowledge. We investigate the contribution of implicit aural skills learning through an exploration of potential relationships between aural skills achievement, tonal WMC, and genre preference. We chose to explore genre preference due to its potential to serve as an indicator of what type of music the participants regularly enjoyed, and thus the type of music to which participants were regularly exposed.

Methods

Participants

Two hundred and fifty-four students enrolled at Louisiana State University completed the study. Students were recruited from the Department of Psychology and the School of Music and received course credit or $20. Participants were excluded in the analysis if they reported hearing loss or taking medication that would alter cognitive performance, or if their performance on any task was greater than 3 standard deviations from the mean score of that task. Thus, 15 participants were excluded (hearing loss: 8, age: 1, univariate outliers on one or more WMC tasks: 6). The remaining 239 eligible participants were between the ages of 17 and 43 (M = 19.72, SD = 2.74; 148 females).

Procedure

Participants completed a battery of tests and surveys measuring cognitive ability, musical sophistication, aural skills experience, and musical genre preferences. The tasks included the Goldsmiths Musical Sophistication Index (Gold-MSI) self-report inventory (Müllensiefen et al., 2014), the Short Test of Musical Preferences (STOMP; Rentfrow & Gosling 2003), a demographic questionnaire, two tests of general WMC (Symmetry Span and Operation Span, Unsworth et al., 2005), a novel test of tonal WMC (ToneSpan), perceptual tests from the Gold-MSI (Melodic Memory, Beat Perception, Sound Similarity), and two tests of general fluid intelligence (Number Series; Thurstone, 1938, Raven's Advanced Progressive Matrices; Raven et al., 1998). Researchers later collected final grades for aural skills and music theory courses completed at Louisiana State University. Only the measures used for analysis are included below.

Goldsmiths Musical Sophistication Index Self-Report (Gold-MSI). Participants completed a 38-item self-report survey which included free response and Likert scale questions (the complete survey can be found at goo.gl/dqtSaB, Müllensiefen et al., 2014). We also added the Short Test of Musical Preferences (STOMP) to this survey, which asked participants to indicate their preference for 14 genres of music on a 7-point Likert scale.

Operation Span (OSPAN). Participants were tasked with completing a two-step math operation and then recalling a letter (F, H, J, K, L, N, P, Q, R, S, T, or Y) in an alternating sequence (Unsworth et al., 2005). The letter was presented visually for 1000ms after each math operation. During letter recall, participants were presented with a 4x3 matrix of all possible letters, each with its own check box. Participants checked the boxes for each letter in the serial order they recalled them being presented.

Tone Span. Participants were tasked with completing a two-step math operation and then recalling a tone (high, middle, low) in an alternating sequence (based on Unsworth et al., 2005). The three tones were modelled after Li, Cowan, & Saults (2005), using frequencies outside of the equal tempered system (200Hz, 375Hz, 702Hz). The tone was presented aurally for 1000ms after each math operation. During tone recall, participants were presented with the three possible tones: H, M, and L (High, Medium, and Low), each with its own check box. Participants checked the boxes for each letter in the serial order they recalled them being played.

Aural Skills Achievement. Aural skills achievement was operationalized as a composite score of final grades in aural skills courses taken at Louisiana State University. Students who passed into higher-level aural skills courses upon their arrival were given the equivalent of an A for each skipped lower-level aural skills course.

Results

Regression Analysis

We conducted a stepwise hierarchical multiple linear regression to investigate whether musical sophistication and aural skills achievement predict tonal working memory capacity. We operationalized musical sophistication as the General score from the Gold-MSI, and we operationalized tonal working memory capacity as the Tone Span score.

The results of our models can be seen in Table 1. Model 1 predicted Tone Span based on the self-report General score from the Gold-MSI. In the second step of the regression analysis, we added the composite aural skills achievement score to the model. There was a significant difference between model 1 and model 2, $F(2, 211) = 5.99$, $p < .02$ and the adjusted R^2 value increased from .18 to .20.

Table 1. Regression results using Tone Span score as the criterion.

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit	Difference
(Intercept)	29.51**	[22.38, 36.64]							
GENERAL	0.29**	[0.20, 0.37]	0.43	[0.30, 0.55]	.18	[NA, NA]	.43**		
								R^2 = .181** 95% CI[NA,NA]	
(Intercept)	32.05**	[24.71, 39.38]							
GENERAL	0.24**	[0.15, 0.33]	0.36	[0.22, 0.49]	.10	[.03, .18]	.43**		
Composite Aural	0.60*	[0.12, 1.09]	0.17	[0.03, 0.30]	.02	[-.01, .06]	.31**		
								R^2 = .203** 95% CI[NA,NA]	ΔR^2 = .023* 95% CI[-.01, .06]

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

Tone Span Strategies

Due to apparatus error, strategy responses could not be retrieved for 113 participants, leaving a total of N=126 participants in analyses involving Tone Span strategy data. To compare the strategies employed by musicians to those employed by non-musicians, we divided the participants by the way in which they were recruited; participants recruited through the School of Music were considered musicians, and those recruited through the Department of Psychology were considered non-musicians. We chose to operationalize in this manner for this comparison because we were particularly interested in the potential influence of aural skills courses on strategy use.

We asked participants to explain what strategies they employed in order to complete the Tone Span task in a free answer format. Five coders independently coded the responses as indicative of one or more of six different strategies we were interested in; if two or more coders agreed on a strategy type, the participant response was coded accordingly. As illustrated in Figure 1, a similar percentage of musicians and non-musicians indicated that they employed rehearsal and ordinal strategies. No musicians responded in a manner that indicated no strategy was used, while some non-musicians responded in a manner that indicated they did not use a strategy. A larger percentage of musicians than non-musicians indicated that they employed each of the remaining strategies. Table 2 shows the percentages of musician and non-musician responses indicating each type of strategy.

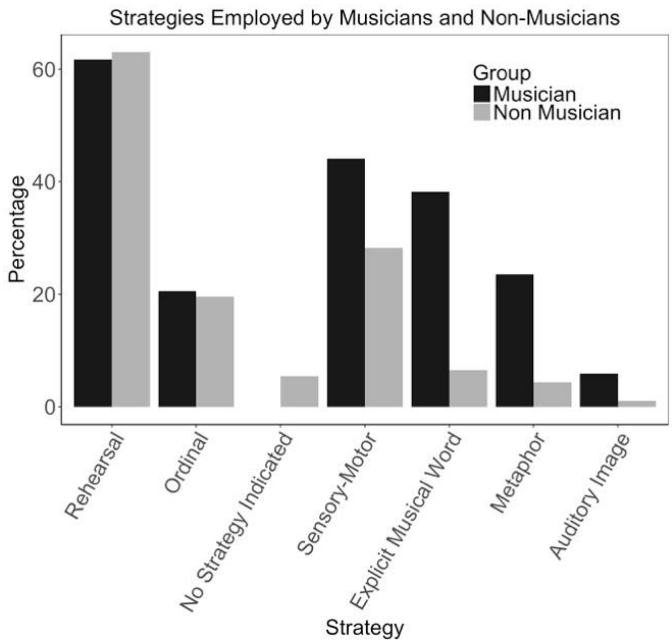


Figure 1. Percentage of musicians and non-musicians who indicated employing each type of strategy in the Tone Span task.

Table 2. Percentage of musicians and non-musicians who indicated employing each type of strategy in the Tone Span task.

	Percentage	
Strategy Employed	**Musician**	**Non-Musician**
Rehearsal	61	63
Ordinal	21	20
No Strategy Indicated	0	5
Sensory-Motor	44	28
Explicit Musical Word	38	7
Metaphor	23	4
Auditory Image	6	1

Exploratory Relationships

We found a significant and positive correlation between the composite aural skills score and the Tone Span score, as can be seen in Figure 2. However, when we focused on the musicians alone, the correlation disappears (see Figure 3). This finding may reflect the "musician advantage" in tonal working memory capacity, but may fail to illustrate an explicit link between tonal working memory capacity and aural skills achievement.

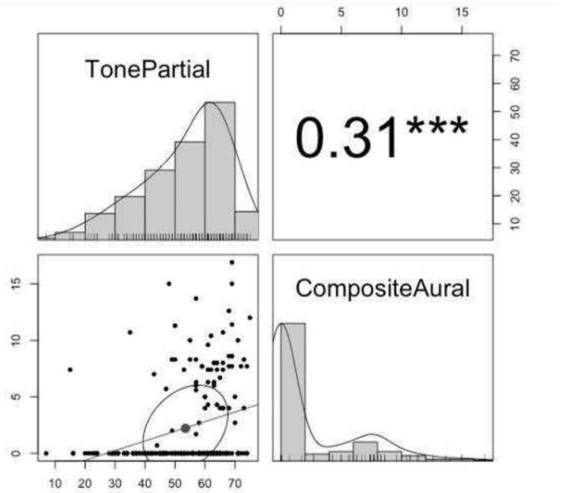


Figure 2. Correlation between Tone Span score and composite aural skills achievement score, including entire subject pool (both musicians and non-musicians). * indicates $p < .001$.**

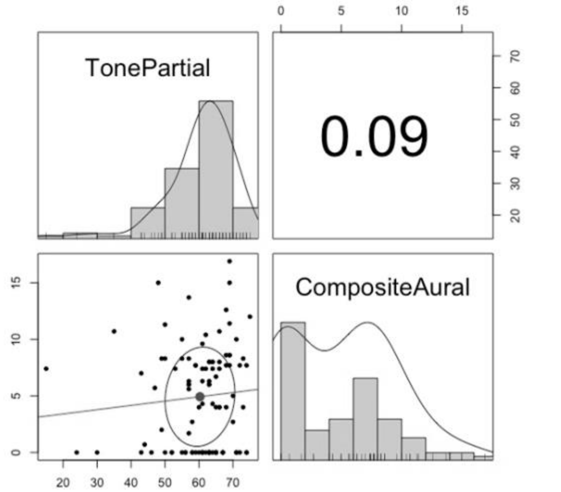


Figure 3. Correlation between Tone Span score and composite aural skills achievement score, using only musicians from the subject pool.

We next explored the correlations between genre preference and Tone Span score. We found indicated preference for three of the fourteen genres to correlate significantly and positively with Tone Span score after applying a Bonferroni correction for Type I error, as can be seen in Table 3.

Table 3. Correlations between indicated preference for genre and Tone Span score. * indicates $p < .05$, ** indicates $p < .01$, and * indicates $p < .001$. With a Bonferroni correction, only correlations marked with *** remain significant.**

Genre	Correlation with Tone Span Score
Classical	0.29***
Blues	0.20**
Country	-0.11
Dance/Electronica	0.01
Folk	0.20**
Rap/Hip-Hop	-0.13
Soul/Funk	0.12
Religious	0.01
Alternative	0.11
Jazz	0.31***
Rock	0.18*
Pop	-0.02
Heavy Metal	0.26***
Soundtracks/Theme Songs	0.10

Discussion

Musicians seem to have an advantage over non-musicians in tonal working memory tasks, which is what motivated our exploration of potential reasons for this advantage. One possibility is that musicians explicitly practice strategies for maintaining tonal information in their working memory, strategies they are likely to learn in an aural skills classroom. To investigate this possibility, we examined the ability of musicianship and aural skills achievement to predict our measure of tonal working memory capacity (WMC). We found that musicianship was able to account for approximately 18% of the variance in tonal WMC, and when we added aural skills achievement to the regression model, we were able to account for an extra 2% of variance. This finding serves to support the claim that musicians have an advantage in tonal working memory tasks, and it also suggests that success in aural skills courses may contribute to that advantage.

We then took a closer look at the strategies participants used to maintain tonal information in their working memory in our tonal working memory task, hypothesizing that musicians would be more likely to use strategies that encouraged deeper levels of processing, as well as strategies that were explicitly musical in nature. We found that a similar percentage of musicians and non-musicians indicated they employed rehearsal and ordinal strategies, while a higher percentage of musicians indicated they employed all the other strategies we were interested in. These findings support the theory that musicians are more likely to learn, and more likely to be aware enough to report using, strategies that encourage deeper levels of processing so as to potentially lessen working memory load.

Finally, we explored relationships between aural skills achievement, tonal working memory capacity, and genre performance. We first examined the entire participant pool and found a significant a positive correlation between aural skills achievement and tonal WMC. However, as roughly half of our sample had very low aural skills achievement scores (due to not having taken aural skills courses), we decided to further investigate this correlation within the pool of musician participants alone. This second correlation was non-existent. These findings likely serve to add support to the theory that musicianship is highly related to tonal working memory capacity, as musicians were more likely to have higher aural skills achievement scores than non-musicians. However, as the correlation disappears when focusing on musicians alone, it calls into question whether the skills musicians learn in aural skills courses are contributing in a meaningful way to tonal WMC. It may be that our population lacked in sufficient diversity to see the relationship clearly, or that, as Wolf and Kopiez state, grade inflation has served to diminish the predictive ability of grades (2014). Given the findings of the regression model, the types of strategies employed, and the significant correlation found between aural skills achievement and tonal WMC, we propose that aural skills courses are still worthy of future investigation into the question of the "musician advantage" in tonal working memory tasks.

We next considered the relationship between tonal WMC and genre preference, and we found that participants who indicated a higher preference for classical music, jazz, or heavy metal were also more likely to achieve a higher tonal WMC score. These correlations may serve to indicate that musicians are more likely to enjoy these genres than non-musicians, which makes intuitive sense given that undergraduate music students are likely to be studying either classical or jazz music, though the strong correlation between tonal WMC and preference for heavy metal music is somewhat surprising. The correlations may also serve to indicate that the music typical to these genres in some way facilitates the learning of tonal working memory strategies that allow for higher tonal WMC. Future research in a controlled setting would help illuminate the reasons behind the relationships we have explored in this study.

Conclusion

In this paper we examined the relationships between aural skills achievement, tonal working memory capacity, and genre preference in an effort to further explore the "musician advantage" in tonal working memory tasks. Our findings suggest that musicianship and aural skills achievement may contribute to higher tonal WMC, and that the strategies musicians develop to manipulate tonal information in their working memory may be a significant mediator of the relationship between musicianship and tonal WMC. Future research will continue exploring implicit learning of these strategies through engagement with particular genres.

Acknowledgements. The authors would like to thank Connor Davis, Hailey Holt, Hannah Keller, Brian Ritter, and Christopher Young for their help running participants for this study.

References

Colley, I. D., Keller, P. E., & Halpern, A. R. (2018). Working Memory and Auditory Imagery Predict Sensorimotor Synchronization with Expressively Timed Music. *Quarterly Journal of Experimental Psychology*, doi:10.1080/17470218.2017.1366531.

Harrison, Carole S., Edward P. Asmus, and Richard T. Serpe. (1994). "Effects of Musical Aptitude, Academic Ability, Music Experience, and Motivation on Aural Skills." *Journal of Research in Music Education 42* (2). SAGE Publications Inc: 131–44.

Honing, Henkjan, and Olivia Ladinig. (2009). "Exposure Influences Expressive Timing Judgments in Music." *Journal of Experimental Psychology: Human Perception and Performance 35* (1): 281–88.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science*, *21*(7), 914-919.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one, 9*(2), e89642.

Ollen, J. E. (2006). A criterion-related validity test of selected indicators of musical sophistication using expert ratings [electronic resource]. Available at http://www.ohiolink.edu/etd/view.cgi?osu1161705351.

Raven, J., Raven, J. C., & Court, J. H. (1998). Manual for Raven's Progressive Matrices and Vocabulary Scales. Oxford, England: Oxford Psychologists Press.

Rentfrow, Peter J., and Samuel D. Gosling. (2003). "The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences." *Journal of Personality and Social Psychology 84* (6): 1236–56.

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition, 152*: 199-211. doi:10.1016/j.cognition.2016.03.017.

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? *Intelligence, 62*: 119-124. doi:10.1016/j.intell.2017.03.005.

Talamini, Francesca, Gianmarco Altoè, Barbara Carretti, and Massimo Grassi. (2017). "Musicians Have Better Memory than Nonmusicians: A Meta-Analysis." *PloS One 12* (10): e0186773.

Thurstone, L. L. (1938). Primary mental abilities. Chicago, IL: University of Chicago Press.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37* (3): 498-505. doi:10.3758/BF03192720.

Wolf, Anna, and Reinhard Kopiez. (2014). "Do Grades Reflect the Development of Excellence in Music Students? The Prognostic Validity of Entrance Exams at Universities of Music." *Musicae Scientiae: The Journal of the European Society for the Cognitive Sciences of Music 18* (2): 232–48.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Effects of Melodic Contagions in the Oral Transmission of Melodies

Elizabeth Monzingo,[1] Daniel Shanahan,[2] Anthony Marasco,[3] Emily M. Elliott,[4] David John Baker,[5] Connor Davis,[6] and Joshua Albrecht[7]

[1,2,3,5,6] *School of Music, Louisiana State University, USA*
[2] *School of Music, Ohio State University, USA*
[4] *Department of Psychology, Louisiana State University, USA*
[7] *Music Department, University of Mary Hardin-Baylor, USA*

[1]emonzi1@lsu.edu, [2]shanahan.37@osu.edu, [3]amarasco@lsu.edu, [4]eelliott@lsu.edu, [5]dbake29@lsu.edu, [6]cdav224@lsu.edu, [7]jalbrecht@umhb.edu

Abstract

Previous research into the transmission and altering of musical signals has primarily examined recognition tasks, emphasizing the role of memory (e.g. Halpern & Bartlett, 2010). Recent studies suggest that the physical act of melodic production might also play a role. Shanahan and Albrecht (forthcoming) found that ascending stepwise motion at cadences tended to be replaced by descending stepwise motion over the course of oral transmission of melodies in a linear transmission train. The current study hypothesizes that this effect will be amplified over the course of oral transmission in a diffusion-based model, in which we examine the influence of a "contagion" cadence. We also examine the ability of measures of working memory capacity and musical sophistication to predict participant choice of cadence. Findings show that the contagion cadence serves to influence the type of cadence participants sing, and the descending contagion cadence exerts more influence than the ascending contagion cadence. Further, measures of working memory capacity and musical sophistication have no significant effect on the cadence chosen. We argue that chosen cadences align with physiological affordances, such that descending cadences are more likely to be sung than ascending cadences.

Introduction

There are a number of reasons why one melody might be more likely to be transmitted than another. Perhaps certain melodies contain features that are reminiscent of specific events; perhaps they are sung by more influential people, and connote status; perhaps melodies that are more prototypical are more likely to be transmitted, or perhaps it's about simplicity. Recent work by Shanahan and Albrecht (forthcoming) has examined how melodies that are less physically affordant (i.e. those with cadences that ascend, rather than descend), are more likely to be transformed than a descending melody. Put more succinctly, "ti-do" melodies are far more likely to transform into "re-do" melodies than vice versa.

The field of Social Learning focuses on why certain ideas might be more easily lost in transmission than others, and is defined as "learning that is influenced by observation of, or interaction with, another animal…" (Hoppitt & Laland, 2013). The study of the transmission of ideas encompasses a broad scope, including evolutionary biology, animal behavior, and decision theory, among other fields. An example from the animal world includes a Payne and Payne study (1985), that found that male humpback whales in a certain population all sang a song that would gradually change through the season, as various changes were gradually introduced. Similarly, Garland et al. (2011) found that migration of humpback whales created a change in song types that followed the migratory patterns.

One of the earliest examples of a study focusing on information lost during transmission was Bartlett's (1932) early work on memory and transmission, which involved participants being presented with either text or a picture, recalling it, and having the result of that recall passed to the next participant, who in turn produced their own version of their recollection to another participant. Bartlett concluded that there are always some features —"dominant features"— that are more easily maintained than others throughout transmission. This type of methodology is known as a "transmission chain." The simplest form of transmission chain is a linear transmission chain, in which one person directly demonstrates to one observer (see Figure 1).

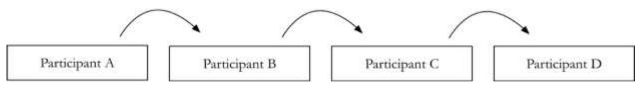


Figure 1. A linear transmission chain, demonstrating a melody from a single demonstrator to a single observer.

The transmission chain methodology seems particularly appropriate for examining how a musical signal like a melody changes over time. One benefit of this approach is the ability to conduct an experiment in a controlled laboratory environment. On the other hand, it is unlikely that an idea is presented to an observer by only one demonstrator. For example, it is far more likely that a melody will have been sung to someone by multiple people before that person teaches it to someone else.

A more ecologically valid approach would be to enlist a diffusion-based model, in which each observer is presented with multiple versions of a melody (see Figure 2). In this paradigm, the observer not only tries to replicate a previously-encountered melody, but they are forced to make a choice regarding which melody they should replicate. In a more complex paradigm, this observer would then become one of four more demonstrators, and this would continue for a number of rounds.

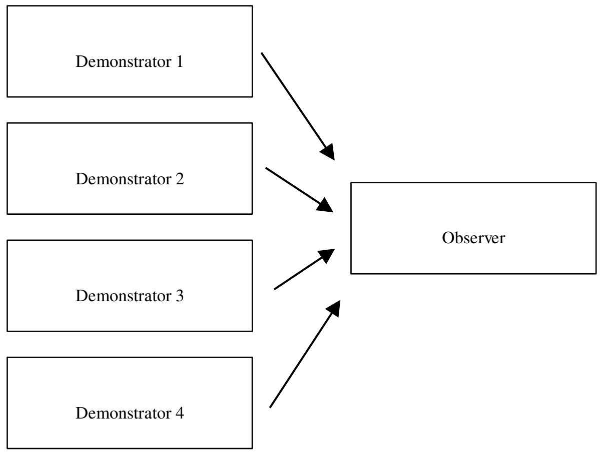


Figure 2. A diffusion-based model, in which one observer is presented with options from multiple demonstrators.

For this study, we were interested in whether the melodies changed in the predicted fashion given multiple demonstrators and a "contagion" cadence. Specifically, we were interested in whether "re-do" (^2-^1) melodies would be more likely to be transmitted than "ti-do" (^7-^8) melodies when presented by multiple individuals, and we wondered whether this transformation would be amplified by presenting a single "re-do" cadence along with the "ti-do" cadences, or by presenting a single "ti-do" cadence along with the "re-do" cadences. We were also interested in how measures of working memory capacity and musical sophistication might be predictive of the choices observers made.

Methods

Participants

Seventy-two students enrolled at Louisiana State University participated in this study. Forty-two participants were recruited from the School of Music, and thirty participants were recruited from the Department of Psychology. Two participants were excluded due to attrition and four participants were excluded due to working memory capacity scores more than 3.5 SD from the mean (final N=66). Eligible participants were between the ages of 18 and 26 (M = 19.55, SD = 1.63; 31 females). Participants received course credit.

Melodies

Eight melodies were chosen as the tonal stimuli for this study. These melodies were chosen in a previous study (Shanahan and Albrecht, forthcoming), because they were meant to mirror the process of oral transmission as much as possible, but were not prevalent enough that participants would be aware of them. The previous study therefore employed an orally transmitted repertoire that would be unfamiliar to our participants. In order to mitigate the effects of lyrics, the original study by Shanahan and Albrecht set out to find stimuli in which the songs were sung on a neutral syllable. A suitable corpus that fulfills all of these criteria can be found in Weiss, et al. (2012). This study involved "unfamiliar folk melodies from the United Kingdom and Ireland [that]...conformed to Western tonality" sung by "an amateur female (alto) singer without lyrics (i.e. "la" for each note) in an everyday (non-operatic) manner" (p. 1075). Four of the chosen melodies included a re-do cadence, and four of the melodies included a ti-do cadence. Four vocalists (two men and two women) at Louisiana State University recorded two versions of each melody; one version included the original cadence, while the second version included the altered cadence.

Goldsmiths Musical Sophistication Index (Gold-MSI)

Participants completed the Gold-MSI Self Report, Beat Perception, Melodic Memory, and Sound Similarity Tasks. We used the General score from the Self Report, in which participants completed a 38-item self-report survey that included free response and Likert scale questions (the complete survey can be found at goo.gl/dqtSaB, Müllensiefen et al., 2014).

Measures of Working Memory Capacity

Participants completed one block each of three measures of working memory capacity (Foster et al., 2014). Each included a practice trial before the test trial.

Operation Span (OSPAN). Participants were tasked with completing a two-step math operation and then recalling a letter (F, H, J, K, L, N, P, Q, R, S, T, or Y) in an alternating sequence. The letter was presented visually for 1000ms after each math operation. During letter recall, participants were presented with a 4x3 matrix of all possible letters, each with its own check box. Participants clicked the check boxes for each letter in the serial order they recalled them being presented.

Symmetry Span (SSPAN). Participants were tasked with completing a two-step symmetry judgement and then recalling a visually presented red square on a 4x4 matrix in an alternating sequence. The square was presented in one of sixteen locations on a 4x4 matrix for 650ms after each symmetry judgement. In the symmetry judgement, participants were shown an 8x8 matrix with random squares filled in black, and they decided if the black squares were symmetrical about the matrix's central vertical axis. During square recall, participants were presented with the 4x4 matrix and clicked the locations in the serial order they recalled the squares being presented.

Rotation Span (RSPAN). Participants were tasked with completing a two-step rotation match judgement and then recalling a visually presented arrow in an alternating sequence. The arrow was either of short or long length and pointed in one of eight different directions. In the rotation match judgement, participants were shown a rotated letter, and they decided whether the letter was presented correctly or as a mirrored image of the letter. During arrow recall, participants were presented with the sixteen possible arrows and clicked them in the serial order they recalled the arrows being presented.

Procedure

Participants in this study completed eight tasks in about 75 minutes. In the first task, participants were asked to listen to, and then record themselves singing, eight unique melodies.

Upon arrival, participants were assigned to one of two conditions (Condition A or Condition B). In each condition, each melody was demonstrated four times, as if by four different demonstrators. As Figure 3 illustrates, there were two possible presentation types: Descending Dominant Presentation or Ascending Dominant Presentation, labeled in accordance with the cadence that dominated the demonstrations in each presentation. In each set of four demonstrations, a single melody demonstration included a contagion cadence that opposed the other three demonstration cadences. More specifically, in Condition A, Melodies 1-4 were each presented in sets of 4 that included 3 re-do cadences and 1 ti-do cadence, while Melodies 5-8 were each presented in sets of 4 that included 3 ti-do cadences and 1 re-do cadence. In Condition B, Melodies 1-4 were each presented in sets of 4 that included 3 ti-do cadences and 1 re-do cadence, while Melodies 5-8 were each presented in sets of 4 that included 3 re-do cadences and 1 ti-do cadence. Table 1 illustrates how these melody and condition combinations ultimately sorted into the two presentation types.

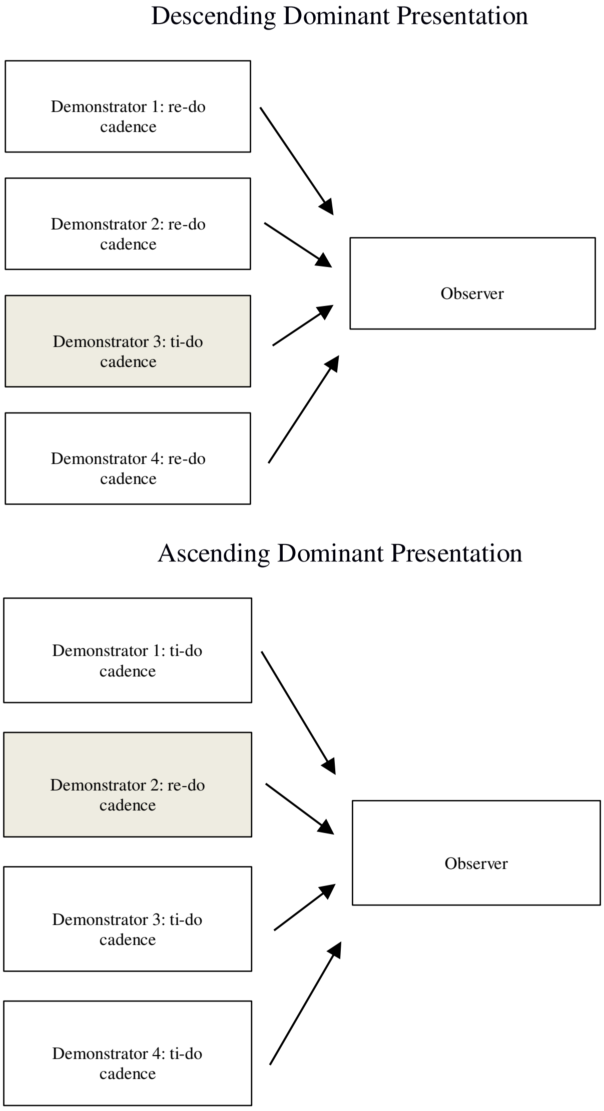


Figure 3. The diffusion model includes a single "contagion" demonstration in each presentation type.

To carry out the experiment, we designed software in Max/MSP. The interface played the four versions of each unique melody, as recorded by four vocalists (two men and two women), in a random order with two seconds of silence between them. Then, following a ten-second break after the final version of the melody played, the interface recorded the participant's own sung version of the melody. Participants were instructed to match the melody presented as best they were able. When they were finished recording, participants were given the option to either erase their current recording and re-record their rendition, or to continue on to the next melody set. During their session, the software logged data to an embedded dictionary, keeping track of their subject number, the playback order of each melody set's audio files, and the number of times they opted to re-record their rendition of the melody before moving on to the next melody. At the end of each participant session, the research assistant is exported and saved the embedded dictionary log as a JSON file. Participants were able to learn this process during a practice trial before continuing on to the experimental melodies.

Table 1. Each melody was demonstrated in one of two presentation types.

Presentation Type	Cadences Demonstrated	Melody + Condition
Descending Dominant Presentation	Demonstrators 1–3: "re-do" Demonstrator 4: "ti-do"	Melody 1 + Condition A Melody 2 + Condition A Melody 3 + Condition A Melody 4 + Condition A Melody 5 + Condition B Melody 6 + Condition B Melody 7 + Condition B Melody 8 + Condition B
Ascending Dominant Presentation	Demonstrators 1–3: "ti-do" Demonstrator 4: "re-do"	Melody 1 + Condition B Melody 2 + Condition B Melody 3 + Condition B Melody 4 + Condition B Melody 5 + Condition A Melody 6 + Condition A Melody 7 + Condition A Melody 8 + Condition A

Participants then completed the Gold-MSI Self-Report, Beat Perception, Melodic Memory, and Sound Similarity Tests. Following that, they completed one block each of OSPAN, SSPAN, and RSPAN. Each task was administered on a desktop computer. Sounds were presented at a comfortable listening level for the tasks that required headphones. All participants provided informed consent.

Results

Our hypothesis stated that ti-do cadences would transform more often into re-do cadences than vice-versa through the act of oral transmission, despite the presentation of the opposite cadence alongside the original. Melodies recorded by participants sourced from the Department of Psychology were often difficult to label with solfege; as a result, we operationalized the re-do cadence as a descending cadence and the ti-do cadence as an ascending cadence. To test our hypothesis, 5 coders independently coded a subset of the melodies in the following way: assigned solfege to the final three pitches (if possible), discounting repeated pitches; provided a contour vector for the last three pitches; and labeled the movement from the penultimate pitch to the final pitch as "ascending" or "descending."

A chi-square test was performed for each presentation type. Under the null hypothesis, participants would sing the dominantly presented cadence 75% of the time, and the weakly presented cadence 25% of the time. We expected to see the re-do contagion influence participants to sing the descending cadence more often than expected, and more often than the ti-do contagion influenced participants to sing the ascending cadence. Following the Ascending Dominant Presentation of a melody, participants were more likely to sing a descending cadence than expected, X2 (1, N=258) = 32.25, p = .0001. Following the Descending Dominant Presentation of a melody, participants were more likely to sing an ascending cadence than expected, X2 (1, N=260) = 5.25, p = .02. This suggests that the contagion demonstration asserted a significant influence on which cadence the participant sang, regardless of whether the contagion was a descending or ascending cadence.

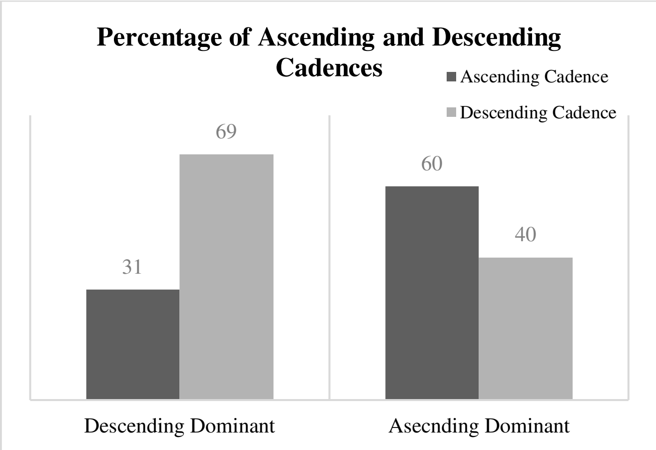


Figure 4. The percentage of ascending and descending cadences observed in each presentation type.

In order to see if there was an effect of musicianship or working memory capacity, we examined measures of musical sophistication and working memory capacity as predictors in whether the penultimate note was changed. We operationalized musical sophistication as the General score from the Gold-MSI. We converted the working memory task recall scores to z-scores and made a composite measure of working memory capacity by averaging across all three z scores. A mixed-effects logistic regression was conducted, with working memory capacity and musical sophistication scores (log-transformed) as random effects, and the melody sung as a fixed effect. No effect was seen on penultimate note change (p = .65, n=432). This would suggest that the task was possibly less about memory or musicianship than about physiological affordances (see Figure 5).

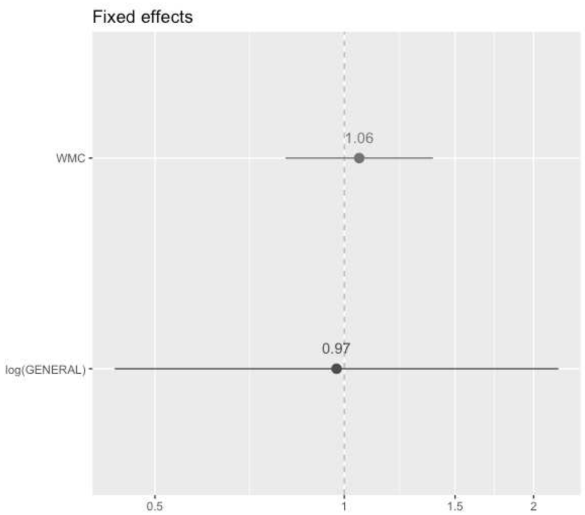


Figure 5: A log-odds ratio of both working memory capacity and general musical sophistication, showing no significant effect.

Discussion

In both cases, the contagion was influential on the participant's choice of cadence. However, as Figure 4 and a comparison of the two chi-square values would indicate, the descending cadence contagion exerted a more powerful influence than the ascending cadence. This serves to support the theory that physiological affordances may be playing a significant role. As observed in Shanahan & Albrecht (forthcoming), physiological affordances can serve to minimize elements that deviate from the norm. In the same way that pitch tends to decline over the course of an utterance (see Collier, 1975; Vassiere, 1984; Hart, Collier, & Cohen, 2006), it may be the case that melodies also decline in pitch toward the end of a phrase. The near-ubiquitous effect found in speech is a convincing example of the "principle of least effort" (Bloomfield, 1933); perhaps the same principle also applies to melodies.

Physiological affordances may be a contributing factor in the transformation of melodies as they disseminate through oral transmission. In this study, we presented multiple demonstrations to a single observer, an improvement of ecological validity from the linear transmission chain previously used, in which each observer heard the melody from a single demonstrator. However, our study only included a single step - one transmission from multiple demonstrators to a single observer - instead of a transmission chain. We plan to expand upon this study in the future, such that each observer becomes one of multiple demonstrators to another observer. This future work will examine the transformation of cadences through the oral transmission of melodies over multiple transmission points to investigate the physical affordances theory further.

Conclusion

In this paper we examined the transformation of cadences through oral transmission. A "contagion" cadence demonstration significantly influenced whether the participant sang a descending or ascending cadence, and the descending cadence contagion was more influential than the ascending cadence contagion. We propose that this is due to physiological affordances, especially given that neither working memory capacity nor musical sophistication were predictive of cadence transformation. Future studies will extend the oral transmission chain to involve multiple points of dissemination.

Acknowledgements. The authors would like to thank Madison Casey, Hannah Keller, Andrew Owen, Brian Ritter, Adam Rosado, Katie Vukovics, and Christopher Young for their assistance in this study.

References

Bartlett, F. C. (1932). Remembering: An Experimental and Social Study. *Cambridge: Cambridge University*.

Bloomfield, L. (1933). *Language history: from Language (1933 ed.)*. Holt, Rinehart and Winston.

Collier, R. (1975). Physiological, correlates of intonation patterns. *The Journal of the Acoustical Society of America*, *58*(1), 249–256.

Foster, Jeffrey L., Zach Shipstead, Tyler L. Harrison, Kenny L. Hicks, Thomas S. Redick, and Randall W. Engle. (2015). "Shortened

Complex Span Tasks Can Reliably Measure Working Memory Capacity." Memory & Cognition, 43 (2), 226–36.

Garland, E. C., Goldizen, A. W., Rekdahl, M.s L., Constantine, R., Garrigue, C., Hauser, N. D., Noad, M. J. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Current Biology: CB*, *21*(8), 687–691.

Halpern, A. R., & Bartlett, J. C. (2010). Memory for Melodies. In Music Perception (pp. 233–258). Springer, New York, NY.

Hart, J. T., Collier, R., & Cohen, A. (2006). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge University Press.Hoppitt, W., & Laland, K. N. (2013). Social Learning: An Introduction to Mechanisms, Methods, and Models. Princeton University Press.

Müllensiefen, D., & Halpern, A. R. (2014). The role of features and context in recognition of novel melodies. Music Perception: An Interdisciplinary Journal, 31 (5), 418–435.

Payne, K., & Payne, R. (1985). Large Scale Changes over 19 Years in Songs of Humpback Whales in Bermuda. *Zeitschrift Für Tierpsychologie*, *68*(2), 89–114.

Shanahan, D. and Albrecht, J. (forthcoming). Examining the Effect of Oral Transmission on Folksongs. Music Perception.

Smith, K., Brighton, H., & Kirby, S. (2003). Complex Systems in Language Evolution: The Cultural Emergence of Compositional Structure. Advances in Complex Systems, 06(04), 537–558.

Vaissière, J. (1984). PROSEIDON, Automatic Detection of Prosodic Cues in Continuous Speech. XIIIes JEP, 53–66.

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: enhanced memory for vocal melodies. Psychological Science, 23(10), 1074–1078.

Intervallic Awareness: the definition of a musical construct

Neto, Pedro A. S. O

Institute of Arts, University of Campinas, Brazil

pasoneto@gmail.com

Abstract

This work undertook the scientific definition and operationalization of a new psychological construct. The Intervallic Awareness Construct (IAC) has been defined as the ability to recognize and to manipulate intervals within different melodic contexts. It has been established that the IAC could be observed throughout the tasks of intervallic comparison, addition, inversion and substitution, which served as basis for the development of the Intervallic Awareness Test (IAT). The IAT has been administered to 21 college level music students, and psychometric concepts, such as criterion validity, internal consistency and inter-rater reliability, have been used to assess the instrument's quality. Inter-rater reliability has been estimated at 0,983 (P<0,001), Cronbach's alpha indicated an internal consistency of 0,864 and Spearman's rho showed a moderate to strong correlation between the scores that subjects reached at the IAT and at a semester of college level Ear Training (0,744, P<0,01). The IAC, as well as the IAT, might be used to fulfill some research needs on music learning and performance. The test hereby proposed could be used to predict and to explain the development of many melodic perception skills. For instance, researchers could look for correlations and causalities that might exist between proficiency on the IAT and on tasks such as sight-singing, memorizing and transcribing of melodies.

Introduction

A well-defined construct may serve as framework for researchers to investigate theories, to test hypothesis and to predict the behaviour of certain variables. Aiming to develop new tools of investigation in the field of music education and cognition, this work has been dedicated to the definition of a new psychological construct which has been called Intervallic Awareness Construct (IAC). The theoretical foundations for the development of this construct were gathered from Dowling's (1970, 1978, 1981, 1986) and Goldemberg's (2011, 2015) works, which have similarly indicated that melodic intervals and scales have different cognitive implications for the performance on certain musical tasks.

Dowling (1986) has demonstrated, both theoretically and experimentally, that the perception of intervals and scales is not necessarily inseparable. The author came to such conclusion by noticing that intervallic perception might be different for individuals with various levels of musical training. Non-musicians, for instance, tended to be sensitive to changes in absolute interval sizes, while musicians with intermediate musical training were not (Dowling, 1986). The author suggested that intermediate levels of musical training could make individuals sensitive to contour and scale-step information, while insensitive to absolute interval sizes.

Dowling's findings about highly experienced subjects, however, is what provided the most insightful thoughts to the present investigation. The author has found that those individuals were able to combine scale-step representations with interval representations, and therefore they presented the most refines skills of melodic perception and recognition.

In a similar manner, Goldemberg (2015) suggests that the perception of melodic patterns and specific interval sizes might be interpreted, respectively, as bottom-up and top-down processing of music. The author suggests, then, that the capacity to combine those two melodic structures might be a prerequisite for fluent sight-singing, just like the combination of bottom-up and top-down processing of verbal language is an indispensable skill for the experienced reader.

The concept of integration between scale-step representations and interval representations constitute, in this work, the theoretical foundations which served as base for the definition of the IAC. Dowling suggests a parallel between years of experience and the capacity to conjugate scale-step representations with interval representations, while Goldemberg speculates about the cognitive advantages that might emerge from a similar strategy. It becomes possible, therefore, to delineate a musical concept which that takes into account the ability to work with those two features of melodic structure.

Besides Dowling's and Goldemberg's works, the definition of the IAC has been based upon some psycholinguistic studies. The next section of this paper reports, therefore, the logic and the methodologies which were employed in order to define a new musical construct.

Development of the Intervallic Awareness Construct

IAC as a linguistic correlate.

Phonological Awareness is a psycholinguistic construct which has been defined as the ability to perceive and to manipulate the constituent sounds of speech, regardless of their syntactic and semantic implications (e.g. Cassady et. Al, 2005; Ribeiro, 2011).

The investigation of phonological awareness is only possible, in the field of linguistics, due to the fact that verbal language is organized in terms of discrete elements (Patel et. al, 1998). In other words, distinctive features are combined to form phonemes, phonemes are combined to form syllables and syllables are combined to form words.

Music, in a similar manner, emerges sequentially, in a way that frequencies are combined to form complex notes, notes are combined to form intervals, and intervals are combined to form phrases. This point of convergence between music and language enables, in this work, the adoption of a highly interdisciplinary strategy of investigation.

Linguists describe and measure the Phonological Awareness construct accordingly to a set of verbal and perception tasks and skills. Comparison tasks, for instance, requires the ability to recognize, between three given words (e.g. frost, frog and dip), the ones that begin with the same sounds (frost and frog).

Table 1 presents some of the abilities which are usually attributed to the construct of Phonological Awareness. It becomes relatively easy to recognize, throughout the examples given at **Table 1**, that many of those tasks are very well suited to the field of music.

Table 1. Table of abilities and examples attributed to phonological awareness.

Abilities	Examples
Word to word matching	Among the words frog, frost and drip, individuals are able to indicate which ones begin with the same sounds. In this case, frog and frost (Ramos, 2005; Yopp, 1988).
Addition/Agglutination	Individuals are able to add, in the beginning of the word ''ice'' the phoneme /m/, thus forming the word ''mice'' (Ramos, 2005).
Inversion	Given the word ''bus", individuals are able to invert the phonemes and say the word ''sub'' (Ramos, 2005; Yopp, 1988)
Substitution	Given the word ''fight'', individuals are able to substitute /f/ by /s/ and form the word ''sight'' (Ramos, 2005; Yopp, 1988).

It has been noticed that, while linguists talk about phonemic inversion tasks, it would be possible to talk about intervallic inversion tasks in regard to music. It is not to say, however, that those tasks, when adapted from linguistics, will preserve the same, or even similar cognitive effects which they had in regard to processing and learning of verbal languages.

Some studies (e.g. Lamb and Gregory, 1993; and Anvari et. al, 2002) have actually indicated that Phonological Awareness might be significantly correlated with certain music perception skills. However, establishing such parallel was not the aim of the present investigation. Such construct has simply been used as a way to define new musical tasks, just like *priming* paradigms have been successfully adapted from linguistics to music.

Operationalization

The operationalization of the IAC has been based upon the tasks described in **Table 1.** Basically, instead of manipulation and perception of certain phonemes within a verbal context, the IAC tasks were focused on the manipulation of intervals within melodic contexts.

Comparison task. Ability to compare melodic excerpts and tell if they present or not the same absolute pattern of interval sizes. Individuals should be able to tell that melodies 1 and 2, from **Figure 1** are different, considering that they maintained scale step and contour dispositions.

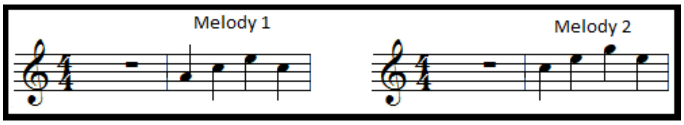


Figure 1. Example of the comparison task.

Addition task. Ability to sing a given interval within a melodic/harmonic context. Individual would be stimulated twice. First with a tonal/modal excerpt, and then they would hear the stimulus presented in **Figure 2**. After hearing the stimuli, individuals would be asked to sing, from the last note of the melody (e.g. E), a given interval (e.g. Major Second). The response should be given according to the response section in **Figure 2**.

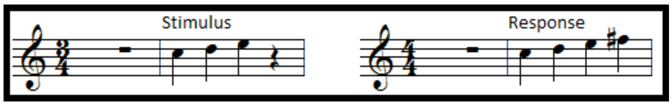


Figure 2. Example of the addition task.

Inversion task. Ability to invert the order of a given sequence of intervals, maintaining the initial note. In **Figure 3**, individuals would hear the stimulus and invert the order of the presented intervals, thus forming the melody represented under "responses".

Figure 3. Example of the inversion task.

Substitution task. Ability to substitute a given interval within a melodic/harmonic context, without altering any other interval from the excerpt. This task is fairly similar to the inversion task. The only difference is that individuals would have to maintain the second interval presented within the excerpts of **Figure 4.**

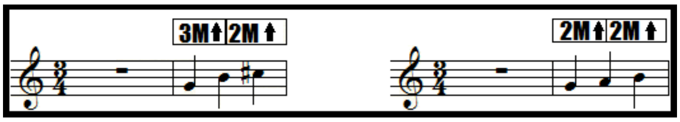


Figure 4. Example of the substitution task.

In general, every task attributed to the IAC aimed to assess intervallic perception within a melodic/scalar context. In order to comprise Dowling's scale step and interval representations, the tasks created require that individuals sing a specific interval, which is inserted in a melodic context.

Addition tasks, for instance, would establish scale step representations by presenting a tonal stimulus. Afterwards, individuals would have to access their interval representations in order to sing the specific interval required by the task. In **Figure 2**, therefore, individuals would have to ignore their scale step representations in order to successfully sing an interval which falls off of the diatonic context.

Definition of the IAC

Definition for the IAC has also been supported by some definitions of the Phonological Awareness construct. As well as linguists have described it as the ability to manipulate and to perceive phonemes within words, the IAC has been defined as ''the ability to perceive and to manipulate melodic intervals within different melodic contexts''. (Neto and Orio, 2016, Neto, 2016, 2017a and 2017b).

Those definitions, as well as the musical tasks developed, were used in order to develop a psychometric instrument. The Intervallic Awareness Test (IAT) was created to evaluate individual's abilities to perceive and to manipulate intervallic features within melodic contexts. The final form of the instrument comprised all of the tasks that were mentioned above.

Methodology

Procedures. Music undergraduates from the University of Campinas (N=21) volunteered to take the Intervallic Awareness Test. The instrument was administered with approval from the university's ethics committee (process number 58417116.2.0000.5404), and the administration occurred inside the university's campus.

All items were explained to the subjects prior to the beginning of each section. Examples were also given to each section of the test, and if the tasks remained unclear, subjects were encouraged to stop the test and ask for further explanation.

The auditory stimuli were given in MIDI format and all the answers were recorded (mono) by the software Voice Record Pro 3.1.9, downloaded on an Ipad 4. The test took an average of 32,84 minutes to be completed.

A preliminary analysis of the audios indicated a pattern of mistakes that were being committed by the subjects. Inconsistencies were frequently noted in regard interval accuracy, maintenance of pulse and number of trials taken to respond each item. The scoring system was, therefore, based upon those criteria. A 4 point Likert scale was adopted to describe subject's performance and subjects should be rated as follows:

- 4 points – Correct interval; steady pulse; first trial
- 3 points – Correct interval; steady pulse OR first trial
- 2 points – Correct interval; unstable pulse AND more than one trial
- 1 point – Incorrect interval

In order to assess the validity of the scoring system adopted, Hermilson Garcia, Doctor of music and Professor at the University of Campinas, scored the performance of 5 subjects according to his own criteria.

Data analysis and results. Cronbach's alpha was used as the internal consistency measure, and its value was estimated at 0,864. Insofar as the Intervallic Awareness Test was designed to measure contents that are similar to those usually addressed in Ear Training courses, a high correlation would be expected between the scores subjects reached in the test and in the first semester of Ear Training. Spearman's rho was estimated at 0,744 ($P<0{,}01$) between the scores of 16 individuals. Pearson's r was estimated at 0,983 ($P<0{,}001$) between the scores given by the professor and by the author of this study, thus indicating a fairly high correlation between the professor's criteria and the criteria previously adopted.

Subject 17 was the only individual who reported having Absolute Pitch (AP). Analysis showed that this subject was an outlier, and therefore he had a significant effect of decrease on internal consistency levels of the IAT. The analysis reported above were already made without subject 17. It is interesting to consider, still, that AP individuals might rely on dissimilar strategies when taking the IAT. Such outcome is still to be further explored.

It has been found, also, that the scores obtained by subjects tended to increase throughout the administration of the instrument. It means that individuals improved their performance at each step of the IAT. **Figure 5** presents a line of tendency for such effect.

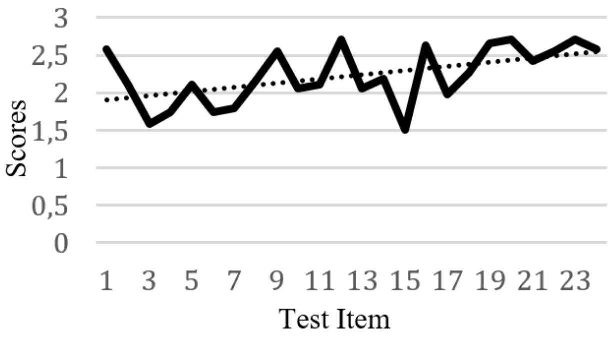


Table 5. Graphic representation of scores obtained throughout the test.

Discussion

Even though the instrument showed satisfactory levels of psychometric validity, it is important consider the need for larger sample sizes. Besides, it is possible to speculate that some tasks vary in regard to their loadings on the specific factor which the test is designed to measure.

Every analytical effort hereby undertaken evaluated the test as a whole, and no specific information was extracted from each of the variables present within the test.

The next investigative steps are being taken towards deeper evaluations of construct validity, and a set of factor analysis are being conducted in order to assess the instrument's homogeneity.

In regard to the line of tendency represented in **Figure 1**, future administrations of the test should consider a familiarization drill. If individuals are previously familiarized with the musical tasks that will be proposed, one could hypothesize that the slope of improvement could be attenuated.

Conclusion

The definition of scientific constructs may reach significance once it enables the formulation and the evaluation of new hypothesis and theories. As well as linguists concern themselves with the correlations that might exist between Phonological Awareness and reading proficiency, for instance, it would be fair to investigate the correlations that might exist between Intervallic Awareness and sight-singing proficiency, or between Intervallic Awareness and working memory for melodies.

The scientific capabilities of the IAT may still be enhanced if it could also be administered to individuals without musical training. The next investigative steps are also being taken in order to make this instrument administrable in a broader sense.

It is possible to say, until then, that the IAC, as well as the IAT, might serve as a good indicatives of a real set of musical abilities. Good levels of internal consistency, as well as criterion validity might suggest that it is interesting to further explore and improve the construct hereby presented.

References

Anvari, S. H., Trainor, L. J., Woodside, J., et Levy, B.A. (2002). Relation among musical skills, phonological processing and early reading ability in preschool children. *Journal of experimental psychology*, 83, 111-130.

Cassady, J. C.; Smith, L. L.; Huber, L. K. (2005) Enhancing Validity in Phonological Awareness Assessment through Computer-Supported Testing, *Practical Assessment, Research & Evaluation*. 10(18), 1-13

Dowling, W. J. (1978) Scale and Contour: Two Components of a Theory of Memory for Melodies. *Psychological Review*, 85(4), 341-354

Dowling, W. J. (1986) Context Effects on Melody Recognition: Scale-Step versus Interval Representations. *Music Perception*, 3(3), 281-296

Dowling, W. J. and Bartlett, J. C. (1981) The Importance of Interval Information in Long-Term Memory for Melodies. *Psychomusicology. A Journal of Research in Music Cognition*, 1(1) 30 – 49.

Dowling, W. J. and Fujitani, D. S. (1970) Contour, Interval, and Pitch Recognition in Memory for Melodies. *The Journal of the Acoustical Society of America*, 6, 12, 524-531.

Goldemberg, Ricardo. (2011) Modus Novus e a abordagem intervalar da leitura cantada à primeira vista. *Opus*, 17(2), 107-120

Goldemberg, Ricardo. (2015). Uma avaliação da abordagem ascendente para a leitura cantada à primeira vista. Revista da *Abem*, 23(34), 84-94

Lamb, S. J., et Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educational psychology*, 13(1), 19-27.

Neto, P. A. S. O. (2017a) Consciência Intervalar: desenvolvimento e operacionalização de um teste psicométrico. [Intervallic Awareness: development and operationalization of a psychometrical. *ABEM*, 25(39)

Neto, P. A. S. O. (May, 2016) Criação de instrumentos de pesquisa: Consciência intervalar e proficiência em solfejo [Development of research instruments: Intervallic Awareness and Solfegio Proficiency] Conference Paper presented at the XII International Symposium for Music Arts and Cognition, Porto Alegre, Brazil.

Neto, P. A. S. O. (October, 2017b) Consciência Intervalar: Desenvolvimento e operacionalização de um construto musical [Intervallic Awareness: development and operationalization of a musical construct]. Paper presented at the XXV Conference for scientific initiation. Campinas, São Paulo – Brazil.

Neto, P. A. S. O; Orio, A. A. A. D. (November, 2016) Convergências processuais entre as linguagens musical e verbal: novas portas se abrem para o estudo interdisciplinar [Processual convergences between language and music: new possibilities for interdisciplinary investigation] Conference Paper presented at the II International Symposium about Language and Cognition, São Paulo, Brazil.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998) Processing syntactic relations in language and music: An event-related potential study. Journal of Cognitive Neuroscience, 10, 717-733

Ramos, N. S. C. (2005) Consciência Fonológica do português do Brasil: descrição e análise de cinco testes. [Phonological Awareness of Brazilian Portuguese: description and analysis of five tests]. (Master's Dissertation). Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil.

Ribeiro, V. S. (2011) Consciência fonológica e aprendizagem da leitura e da escrita: uma análise dessa relação em crianças em fase inicial de alfabetização. [Phonological Awareness and learning to write and read: analysis directed to children in early stages of development]. *Entrepalavras*, 1(1), 100-116

Yopp. H. K. (1988). The Validity and Reliability of Phonemic Awareness Tests. *Reading Research Quarterly*, 23(2), 159-177

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Meaning Beyond Content: Extramusical Associations are Plural but not Arbitrary

Jason Noble[1] and Stephen McAdams[2]

[1,2] *Music Perception and Cognition Lab, Schulich School of Music, McGill University, Canada*

[1]jason.noble@mail.mcgill.ca

Abstract

Extramusical meaning is often conceptualized via a dubious metaphor —MEANING IS CONTENT—likening music to a container that delivers extramusical meaning to listeners and implying that extramusical meaning is fixed within the music. Skeptics observe the plurality of listeners' extramusical associations and conclude that extramusical content is arbitrary or nonexistent. We argue that the solution is not to dismiss extramusical associations, but to conceptualize them differently. We propose replacing the "content" metaphor with a dynamical model in which meaning arises through homologous or topical relations between musical and extramusical domains. Music presents a complex of attributes, and listening involves selectively attending to subsets of those attributes. The attributes attended during a given listening (e.g., noisy timbre, busy texture) may share properties with attributes of an extramusical domain (e.g., busy crowd), and this homology may provide a plausible basis for a cross-domain mapping. But those same attributes may also provide plausible mappings with other extramusical domains (e.g., swarm of bees, agitated state of mind, etc.), and furthermore, attending to a different subset of attributes may yield a different set of potential mappings. Still other associations could arise from non-homologous relations such as musical "topics." Thus, a plurality of plausible, non-arbitrary extramusical associations with a given musical example may coexist, drawing on different attributes and/or relations. We support this hypothesis with an empirical study in which listeners rated excerpts of 20th- and 21st-century music along a battery of semantic scales. Results show significant consistency between listeners in their extramusical associations, which in many cases could be explained in terms of homologous attributes between musical and extramusical domains and in others could be explained in terms of topical significance. We conclude that plurality of extramusical associations does not imply arbitrariness, and that this pervasive and complicated aspect of musical experience merits further study.

Introduction

Although listeners routinely experience extramusical associations in response to music, such associations are excluded from formalist study and the concept of "absolute" music, and are relatively little-studied in music perception and cognition research. Why such reluctance to engage this aspect of musical experience? For many critics, such as the composer Witold Lutosławski (1913-1994), the problem seems to be the absence of *fixed* meaning, or the *plurality* of possible meanings, in music. Lutosławski felt that since each person can interpret music differently, extramusical reference is inherently unstable and it is therefore best to compose music that makes no attempt to express extramusical content (Jakelski, 2014). He preferred to bracket the question: "We must definitely assert that music is not an unequivocal art if one can attach to it so many different meanings. Because of this unequivocalness, it is better to consider music as an asemantic art." (Lutosławski, published 2007).

Lutoslawski's reflections reveal an important issue in music perception and cognition: even granting that some music is composed with the *intent* to refer outside itself and other music is not, this may have little bearing on how the music is *interpreted*. Listeners may fail to make extramusical associations intended by the composer, and they may draw other extramusical associations that have nothing to do with compositional intent. The difficulties posed by such plurality and inconsistency of interpretations have led many scholars to avoid the subject, preferring to discuss features of music that can be asserted more positivistically. But conceiving of music as isolated from other domains of experience and meaning is, we believe, predicated on a solipsistic fantasy: as Nicholas Cook says, "music never *is* 'alone,'…it is always received in a discursive context…it is through the interaction of music and interpreter, text and context, that meaning is constructed." (Cook, 2001). We sought a theoretical account of extramusical associations that allows for plurality but avoids arbitrariness. We believe that music can mean many things, but its meanings are nevertheless constrained by its attributes: music is polysemic, it is not omnisemic or asemic. We sought to ground our account of extramusical meaning in metaphor theory and cognitive semiotics, and to demonstrate it empirically.

Metaphor Theory and Extramusical Meaning

Metaphor theory, as articulated by Lakoff and Johnson, asserts that metaphor is not merely a rhetorical and poetical device but is "pervasive in everyday life, not just in language but in thought and action" (Lakoff & Johnson, 1980). Metaphors are mitigated by cultural assumptions, values, and attitudes, such that all experience is "cultural through and through." Many metaphors are so deeply ingrained that we often do not even recognize them as metaphors. A familiar example from musical experience is PITCH IS HEIGHT; as Lawrence Zbikowski remarks, "[p]erhaps more remarkable than the long tradition of construing pitch relations in terms of 'up' and 'down' are the ready reminders of how arbitrary a construal it is" (Zbikowski, 2002). He mentions three other cultures that use different metaphors for pitch: the Kaluli of Papua New Guinea, who describe melodic relations in terms of characteristics of waterfalls; Balinese and Javanese cultures, in which pitches are conceived as 'small' and 'large' because smaller sounding bodies tend to vibrate more rapidly than larger ones; and the Suyá of the Amazon, who describe pitches as 'young' and 'old' because voices tend to deepen with age. Such metaphors, while they may be naturalized within their respective cultures to the extent of evading notice, are not without consequence for understanding: Zbikowski notes that correlating pitch with vertical position "leads quite naturally to an imaginary world in which pitches become things that move through space." As it happens, metaphors of 'up' and 'down' are prolific in our culture. Lakoff and Johnson list several other cultural domains that map onto vertical position in physical space: HAPPY IS UP; SAD IS DOWN; CONSCIOUS IS UP, UNCONSCIOUS IS DOWN; MORE IS UP, LESS IS DOWN; GOOD IS UP, BAD IS DOWN, and so forth. The shared image schema of vertical position between such domains and musical pitch creates the potential for extramusical associations rooted in cross-domain mapping.

"MEANING IS CONTENT"

Another, even more fundamental metaphor is at work in the concept of extramusical meaning (and of meaning in general): MEANING IS CONTENT. This is a generalization of a metaphor identified by Lakoff & Johnson, LINGUISTIC EXPRESSIONS ARE CONTAINERS FOR MEANINGS, which, as the authors say, entails "that words and sentences have meanings in themselves, independent of any context or speaker." The implication of MEANING IS CONTENT is that if music has extramusical meaning, then music must be like a container that delivers extramusical meaning to the listener: the meaning must be somehow inside the music, there to be unpacked. Meaning on this account is a property of the music, something that can be revealed through exegesis, something fixed, something about which claims may be true or false. This metaphor is seen to be at work, either positively or negatively, in many descriptions of extramusical meaning. For example:

> I view any discourse about the so-called content of a composition with some scepticism; to my mind this content is absent. (Lutosławski, quoted in Kaczynski, 1984)

> Music is capable of modeling semantic *content* as well by motivating the construction of scenarios in musical space that model conceptual content itself... (Nussbaum, 2007)

> [Musical beauty] is self-contained and in no need of content from outside itself... (Hanslick, 1986; originally published 1854)

We would like to suggest that the metaphor MEANING IS CONTENT may be a source of much of the controversy and confusion about extramusical meaning. Perhaps there are other, less problematic ways to conceptualize extramusical meaning: not something that the music delivers, but something that arises in the act of interpretation; not something fixed, but something dynamic and contextual; not an objective attribute, but a kind of interaction between subject and object. The crucial point, we believe, is that when we speak of music's extramusical content (or semiotic content, semantic content, narrative content, emotional content, etc), we fashion meaning as something *contained* within the music, which is difficult to square with the plural (and especially with contradictory) interpretations of listeners. One solution is to adopt a formalist position and abandon the question of extramusical meaning. But another option might be to abandon the metaphor MEANING IS CONTENT, and conceive of extramusical meaning differently.

Extramusical Meaning, Attribute Selection, And Homology

The model of extramusical meaning we advocate is closely related to Nicholas Cook's theory of attribute selection. As Cook describes it, musical works (or "traces") "can be thought of as bundles comprised of an indefinite number of attributes from which different selections will be made within different cultural traditions, or on different occasions of interpretation." From the vast (but not infinite) bundle of attributes presented in a given musical entity (for example: bright timbre, smooth contour, etc.), the listener (or culture) selects a finite number of attributes that may be understood in terms of models provided by source domains with homologous attributes (e.g., bright light, smooth surface, etc.). But as homologous attributes may be shared by multiple extramusical domains, multiple mappings may be equally plausible for any given musical attribute. The potential for plurality is multiplied greatly when one considers that *groups* of attributes may be attended in any given act of interpretation, and even more when one considers that the subset of attributes attended may differ between interpretations. Nevertheless, the interpretations are grounded in meaningful relations between musical and extramusical attributes. Neither bereft of meaning nor bound to fixed meanings, Cook argues that "musical works are unstable aggregates of potential signification." Zbikowski makes much the same point, saying that musical meaning "is not, in the final analysis, simple or direct but multivalent and contingent," reflecting a rich set of activated correspondences. Nussbaum characterizes musical performance as "a non-propositional symbolic utterance that motivates the construction of mental models." We believe that homology of the kind referenced in these accounts can account for a great deal of extramusical meaning, but also that meaningful relations with extramusical domains can arise from non-homologous relations such as musical "topics." Topical associations, which we have discussed in detail elsewhere (Noble, 2018), arise from the context in which music is experienced rather than from inherent properties of the music itself, and this kind of signification has been demonstrated to be an important aspect of extramusical meaning in a variety of musical styles (Huovinen & Kaila, 2015).

Metaphor and the Discourse of Contemporary Composers

Extramusical domains play an important role in the metaphorical way many contemporary composers talk about their music. Many of these domains appear to share homologous image schemata with the music. For example:

> The polyphonic structure...remains hidden in a microscopic, under-water world...composition is like letting a crystal form in a supersaturated solution. (Ligeti, 1983)

> I proposed a world of sound-masses, vast groups of sound-events, clouds, and galaxies...thus stochastic music was born. (Xenakis, 1971)

To investigate the assumption that extramusical associations are arbitrary, we designed a study to assess (1) to what extent listeners may be consistent in their extramusical associations with selected excerpts of contemporary music, (2) to what extent the extramusical domains invoked by the composers of those excerpts may be relevant to listeners' associations, and (3) to what extent listeners' associations can be explained in terms of homologous relations with extramusical domains.

Method

Thirty-eight participants (24 female) between 18 and 50 years of age (M = 24.9, SD = 6.0) completed the experiment. Participants heard a series of 40 excerpts selected from commercially available recordings of 20th- and 21st-century music featuring sound mass [1] and related fusion-based aesthetics, and rated them along three batteries of semantic scales presented in three blocks. Stimuli had an average duration of 15s. Because we were interested in studying participants' responses to ecologically valid stimuli, the excerpts were not matched for loudness. Each excerpt was heard and rated in each block, such that each of the 40 excerpts was encountered three times (for a total of 120). On each encounter, participants heard the excerpt at least once, and had the option to hear it a second time.

The semantic scales were based on metaphorical descriptions by composers and theorists. For example, following from the quotes by Ligeti and Xenakis above, participants would use a continuous slider to indicate the extent to which each musical excerpt reminded them of crystals, clouds, and galaxies. The first block featured terms involved in the definition of sound mass (Density, Complexity, Homogeneity). The second block featured adjectival metaphors used to describe sound mass (Volatile, Atmospheric, Busy, Static, Formless, Impenetrable, Voluminous, Kaleidoscopic). The third block featured nominal metaphors used to describe sound mass (Gas, Liquid, Solid, Clouds, Wind, Water, Webs, Galaxies, Crystals, Machinery, Herds/Crowds/ Swarms). Additionally, in Blocks 2 and 3, there was an optional "Other (please specify)" slider that participants could use to add a scale if they wished to indicate an association not listed.

In Block 2 (adjectival metaphors), the instruction was worded "Please rate the degree to which you perceive the example to be:" with a range defined for each individual category (for example, from "Not volatile at all" to "Very volatile"; from "Not formless at all" to "Completely formless"; etc.). In Block 3 (nominal metaphors), the question was worded "Please rate the degree to which the example reminds you of:" with a range for each individual category from "Very much" to "Not at all."

We expected to observe correlations between sonic/musical attributes in the excerpts and the ratings participants provided for the categories. We further expected these correlations to be explicable in terms of homologous attributes between the excerpts and the categories, sometimes based in sonic similarity (for example, the excerpt from Trevor Wishart's *Vox 5*, which samples the sound of buzzing bees, would probably be rated highly for "herds/crowds/swarms") and sometimes in abstract affinities (for example, the excerpt from Krzysztof Penderecki's *Threnody to the Victims of Hiroshima*, which consists of a loud quarter-tone cluster saturating a wide range of the audible spectrum, would probably be rated highly for "impenetrable"). Finally, we hypothesized that excerpts which achieve sound mass integration in the same way (for example, by exploiting the low register, or through dense kinetic activity) would have some similar musical attributes and would therefore invite some similar mappings.

Results

If listeners' associations had been arbitrary, then we would expect a random or unpredictable distribution of ratings. This was not the case: ratings showed significant variability between excerpts and consistency between subjects. In many cases, especially the Block 2 (adjectival) categories, the ratings correlated with clearly identifiable musical attributes with intuitive, homologous mappings between the musical and extramusical domains. For example:

- 'Volatile' was associated with absence of stable pitch, emphasis on noisy or unstable timbres (i.e., electroacoustic sounds or instrumental extended techniques), kinetic and dynamic textures, and continuous and/or unpredictable change. There was a wide range of responses, indicating that participants perceived considerable variance between the excerpts along this category.

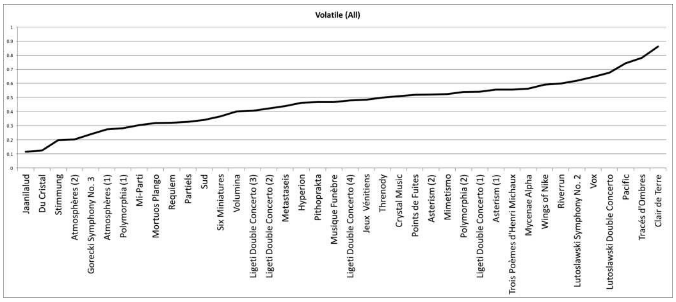


Figure 1. Graph of averaged participant ratings for 'Volatile' (excerpts in ascending order for ratings in that category).

- 'Kaleidoscopic' tended to be associated with timbral heterogeneity and internal dynamism or process, as well as with mid-high registers. Several excerpts featuring "Shepard tone"-like patterns were rated highly, perhaps suggesting an affinity between two cross-modal types of circular or cyclical motion. Excerpts rated lowly for this category tended to be rhymthically static, timbrally homogeneous, and situated in a low register.

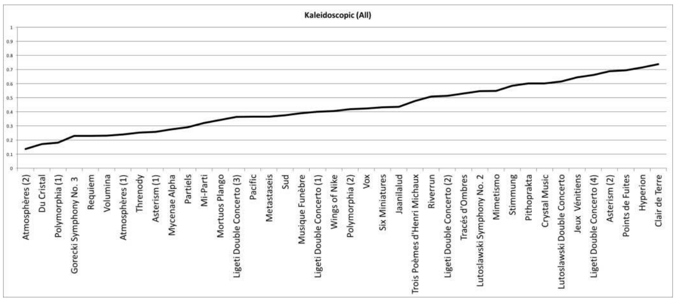


Figure 2. Graph of averaged participant ratings for 'Kaleidoscopic' (excerpts in ascending order for ratings in that category).

Note that the highest-rated example for both categories is an excerpt from Robert Normandeau's *Clair de Terre*, which

[1] Sound mass is a musical aesthetic based on the integration of multiple sound events or sources into perceptually dense and homogeneous auditory units. The study reported in this paper focussed on sound mass music, but the theoretical account of extramusical meaning offered here is in principle generalizable.

features selected attributes listed in both descriptions. Participants either drew the same musical attributes into different cross-domain mappings, or selected different musical attributes for different cross-domain mappings. This is a good example of non-arbitrary plurality: the same music may be interpreted as either 'volatile' or 'kaleidoscopic' depending on which musical and extramusical attributes enter into the cross-domain mapping. For both categories, there are plausibly homologous attributes between the musical and extramusical domains (e.g., unpredictable change/motion, heterogeneous timbre/colour), but the semantic categories do not constitute the "content" of the music in any meaningful sense.

In some cases, there was a direct mimetic relation between the category and the sounds: for example, electroacoustic excerpts—especially Xenakis's loud, noisy *Mycenae Alpha*—were rated highly for 'Machinery.' Often it was possible to relate excerpts to ratings on a case-by-case basis, drawing on different musical attributes and also different attributes of the extramusical domain. For example, for the category 'Water,' the relation between the musical excerpt and its extramusical association varied considerably from excerpt to excerpt, appearing to draw upon different homologies depending on the interpretive context (e.g. pizzicato sounds associated with raindrops in one example, Shepard-tone-like patterns associated with waterfalls in another). In some cases, the ratings were easier to interpret in terms of presumed topical significance than homologies between attributes of musical and extramusical domains. For example, the highest rated example for 'Galaxies' was from Jonathan Harvey's *Mortuos Plango, Vivos Voco*, which is musically nothing like the kind of "galaxy" invoked by Xenakis in relation to his stochastic procedures, but is similar in its sinusoid-based timbres to soundtracks from early science fiction movies.

To examine the results for underlying factors that may have guided the ratings across categories, Principal Component Analysis (PCA) was conducted on the Block 2 and Block 3 categories. The five semantic clusters revealed by this PCA, which we named after strong contributing components, are largely intuitive groupings: "Liquid-Crystal," "Busy-Crowd," "Formless-Machinery," "Voluminous-Solid," "Wind-Gas." They also map consistently onto identifiable musical properties. For example, "Busy-Crowd" was characterized by the uncoordinated activity of many parts, while "Voluminous-Solid" involved loud dynamics, spectral saturation, low register and broad compass.

Participants also had the option to add categories of their own for each excerpt in Block 2 and Block 3. The categories participants chose to add implied that their extramusical associations are also affected by topical significance, especially that of movies, television, and video games. Many of the participant-added categories suggest the imagery of science fiction, fantasy, horror, and fairy-tale movies (e.g., "blood," "aliens," "Alice in Wonderland's confusion"). The influence of movies on listeners' extramusical associations has been demonstrated with other styles of music (Margulis, 2017), and it is intuitive that association with movies would also exert a strong influence on listeners' extramusical associations with contemporary music since film soundtracks are likely the only contexts in which many participants (especially nonmusicians) have experienced it.

Conclusions

This research contributes to a growing body of studies that provide a strong counter-argument to the charge of arbitrariness often leveled at accounts of extramusical meaning (e.g., Wallmark 2018, Margulis 2017, Huovinen and Kaila, 2015). There is little doubt that listeners can and do experience extramusical associations with music—even presumably unfamiliar contemporary music—and that the selected set of semantic categories from composers' and theorists' discourse are germane to their experiences. The results further demonstrate that extramusical associations are often significantly consistent across listeners (at least in an experimental context) and that they often relate to identifiable, homologous musical attributes. That the same excerpts may be rated highly along different categories may be explained in terms of attribute selection: the same musical attributes may map onto attributes of different extramusical domains, or different musical attributes may provide the basis for different cross-domain mappings. Such mappings are frequently homologous, but may also involve non-homologous relations such as topical significance. Extramusical meaning emerges not as content delivered by the music, but as a dynamical, selective, many-to-many mapping that arises in the act of interpretation. Plurality in extramusical associations poses methodological challenges for researchers, but it should not deter researchers from exploring this pervasive and fascinating aspect of musical experience, nor to dismiss it as arbitrary.

References

Cook, N. (2001). Theorizing musical meaning. *Music Theory Spectrum, 23*(2), 170-95.

Hanslick, E. (1986). *On the musically beautiful*. (G. Payzant, Trans.). Indianapolis: Hackett Publishing Company.

Huovinen, E. & Kaila, A.-K. (2015). The semantics of musical topoi: An empirical approach. *Music Perception, 33*(2), 217-243.

Jakelski, L. (2014). Witold Lutosławski and the ethics of abstraction. In S. Latek & M. Trochimczyk (Eds.), *Lutosławski – music and legacy* (pp. 71-96). Montreal: Polish Institute of Arts and Sciences in Canada.

Kaczynski, T. (1984). *Conversations with Witold Lutosławski*. (Y. May, Trans.). London: J. & W. Chester.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Ligeti, G., Várnai, P., Häusler, J. & Samuel, C. (1983). *Ligeti in conversation with Péter Várnai, Josef Häusler, Claude Samuel, and himself*. London: Eulenburg.

Margulis, E. H. (2017). An exploratory study of narrative experiences of music. *Music Perception, 35*(2), 235-248.

Noble, J. (2018) Perceptual and semantic dimensions of sound mass. (Unpublished doctoral dissertation). McGill University, Montreal, Canada.

Nussbaum, C. (2007). *The musical representation: Meaning, ontology, and emotion*. Cambridge, Mass.: The MIT Press.

Wallmark, Z. (2018). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music*, 1-21.

Lutosławski, W. (2007). *Lutosławski on Music.* Lanham, Maryland: The Scarecrow Press.

Xenakis, I. (1971) *Formalized music: Thought and mathematics in composition*. Bloomington: Indiana University Press.

Zbikowski, L. (2002). *Conceptualizing music: Cognitive structure, theory, and analysis*. Oxford: Oxford University Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Constructing the Conceptual Meaning in Music: Imaginative Dimensions and Linguistic Descriptions

Marcos Nogueira

Department of Composition, Federal University of Rio de Janeiro, Brazil

mvinicionogueira@gmail.com

Abstract

The theoretical-methodological contribution of the so-called embodied cognitive sciences became absorbed by musicological research in the 1990s. Throughout the 20 years of development of a musical enactivism, several researchers have faced the challenge of overcoming the representational model of the academic tradition to explain what happens to the mind when we interact with music creatively. This paper argues for the validity of the hypothesis according to which the attentional focus of the descriptor regulates the linguistic descriptions of his or her musical understanding. In the process of producing meaning the listener emphases one or another of the imaginative dimensions that this study recognizes as categorization of movements, production of formal images, and the establishment of symbolic predicates. Moreover, the present study argues that events that elicit more meaningful orienting of attention trigger a cognitive device called *orienting response* which regulates the attentional focus of the listener. Before this, the development of a strictly enactivist model for the investigation of the modes of conceptualization of the musical understanding expressed in the linguistic descriptions of the listeners offers unprecedented access to the path that goes from the concept towards the meanings (mostly unconscious) that are not yet concepts. These are the meanings with which we invent our musical worlds before conceptualizing the world musically.

Introduction

The *orienting response*, also known as "orienting reaction" or "orienting reflex," is an immediate action of the organism in response to a particular change it perceives in its environment. A striking feature of the orienting response is that in noticing the event that extrapolates a threshold of discrepancy in the surrounding environment, the individual directs his or her attention to the event before even identifying it. We can understand the orienting response as a set of responsive body indicators that signal the perception of a stimulus that stands out as relevant. The phenomenon would, however, be a response to the "non-aversive" novelty. For example, in the auditory domain, a sonic event perceived as novelty will elicit a spontaneous response unless it presents itself at such an extraordinary level of sound intensity that it overcomes the auditory comfort zone and inflicts on the individual a kind of threat. Also, this last situation provokes another type of reaction, generally understood as the "defensive reflex" of the individual, which emerges as a "blocking" action to the threatening event. In short, simple orienting responses occur when the perceived change in the surrounding environment does not cause another type of reaction known as the "startle reaction," or "startle reflex," the *scare*.

Ivan Sechenov first described the phenomenon of the orienting response in his *Reflexes of the Brain* (1863/1965), but it was Ivan Pavlov who, in *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* (1927), identified it in the terms in which it has been studied until now—Pavlov also referred to the phenomenon as "a reflection of 'what is this?'." He noted that the perception of *novelty* or even a significant event is the central cause of the phenomenon. For him, in the face of the novelty, the individual immediately interrupts what he or she is doing and directs his or her cognitive resources to the source of stimulation—it is, therefore, a behavioral component of orientation. Years later, Evgeny Sokolov (1960, 1963) systematically devoted himself to the phenomenon, describing the primary object of his investigations: habituation as a process of gradual familiarization with a new event that becomes repetitive. According to him, the repetition of the event provides a progressive reduction of the activation of orientation responses. Thus, the original introduction of a change in the currently active neuronal model, that is, the one in which the individual is focused, results in an orientation response. However, when one becomes familiar with the event, the individual gradually evaluates it as inconsequential and unimportant and no longer allocates effort of attention to it. Nevertheless, the orienting response toward novel and salient stimuli may not constitute a unitary process (Barry, 1979). Currently, changes in the currently active neuronal model has been interpreted as an event-related potential (ERP) sign of the orienting response, sharing the antecedent conditions of the traditional theory as well as the orienting response sensitivity to habituation (Barry, MacDonald, & Rushby, 2011).

This paper deals with the process of musical meaning formation. The literature in music theory that dialogues with the results of experimental psychology, especially from Leonard Meyer's *Emotion and Meaning in Music* (1956) to David Huron's *Sweet Anticipation* (2006), emphasizes the role of emotion and expectation (especially the ability to anticipate the occurrence of specific events imaginatively) and its implications as primordial experiences that condition the emergence of the musical meaning. I propose a change of this focus in the investigation of the meaning produced in the act of listening to music, recovering the "behaviorist" concept of the orienting response addressed, however, in the context of enactivism-based cognitive semantics. The background hypothesis of this research is that throughout the experience of music, we do not grasp it properly "by expectations," that is, we do not remain motivated to engage with music simply regulated by imaginative beliefs and anticipations of future events of the piece of music we experience. This study admits that such imaginative acts are more or less conscious and largely conditioned by the listener's expertise, driven by the desire for a reward of meaning. Instead, I believe that in the unfolding of the listening experience, the contextual conditioning—cognitive *schemes* of memory—with which we are facing a new musical experience, should not be reduced to stylistic expectations and anticipations. I argue that, throughout musical listening, the experiences of stylistic expectation are in fact punctual (Nogueira, 2016a, 2016b). Otherwise, I understand that the events perceived as "novelty," those we evaluate, in the act of listening, as ruptures or variants of the *stream* (Bregman, 1990) stasis we infer at each moment, usually regulate the production of perceptual hierarchies of the listener.

Therefore, I propose to discuss the validity of the hypothesis according to which the attentional focus of the listeners regulated by their orienting responses configures the linguistic descriptions of their musical understanding. I also want to believe that the orienting responses of the listeners trigger the production of meaning in three non-exclusive experiential domains of imaginative production: the categorization of movements, the production of formal images, and the establishment of symbolic predicates.

Novelty and Attentional Process

I intend to argue that the process of construction of the musical meaning is inexorably started in contingent situations of the interaction of a complex of potentially relevant events of the musical work—*affordances* (Gibson, 1977, 1979)—and a listener in whose cognitive apparatus the processes that we can understand how musical orienting responses unfold. I understand that this cognitive device, operating in a preconceptual condition (strictly unconscious), determines the selection of the events of the musical stream that will regulate the imaginative production of the listener and the formation of its meaning. It should be admitted that the formation of meaning in music develops as in any other knowledge domain. The search for the origin of meaning in the experience of music must therefore consider, first and foremost, how the listener selects, on the musical surface, the events that will condition his or her construction of meaning. It is this original stage of apprehension of the musical stream that will determine in which configurations the stream will participate in the preconceptual process of formation of the musical meaning. I am referring to inherent schematizations and cross-domain mappings (Lakoff, 1987; Lakoff & Johnson, 1980; Johnson, 1987) with which we engage in musical listening, even before we engage in conceptualization acts that will be revealed in the linguistic descriptions of our musical understanding.

Novelty as something that represents some significant change (discontinuity) in the environment—or, particularly, in the musical stream—is not always something that is easy to identify or classify. Besides, the significance of an event is an issue that also imposes some difficulty for a theory of the musical orientation response. In auditory environments where multiple sonic events compete for attention, the challenge is to find relevant information and to ignore events that are unrelated to current task goals. To understand the attentional process in aural experience is useful to consider how it operates in real time. In the visual search, Eimer (2014) observed that each of the stages of attentional selectivity is temporally and functionally distinct and performs a specific function. *Preparation* is related to "representation of the search goals in working memory"; the *guidance* stage is a "parallel accumulation of information about presence of task-relevant features"; *selection* implies "allocating visual processing resources to candidate target objects at specific locations"; and *identification* is related to "maintaining select objects in working memory" (p.528). I intend to test this model in the investigation of attentional control in the music experience. However, the approach in the present study focuses on the stages Eimer calls "guidance" and "selection": a stage of orientation that does not yet involve the selection of the musical resources assumed by the listener's cognitive apparatus as relevant and the moment when selection is properly assumed.

Although it is essential for organisms to detect novelties in their means, there is no way to predict when an event will be relevant as a modifier of environmental conditions. In the particular case of listening to music, we often find situations in which we initially attribute relevance to specific events of a piece of music that we hear for the first time and which, throughout the experience, are irrelevant. However, it is plausible to consider that the higher the density of "new" events in a given musical stream—which I shall call commutative events, by enabling and even promoting the interruption of a given stream stasis, by replacing the object of the listener's attention—the greater the possibility that any of these become significant in the act of listening. Thus, musical stretches with few potentially "new" events would be virtually less difficult to assimilate since they present themselves as fluidly and consistently coherent configurations. As such, they elicit less attention from listeners in their ongoing process of understanding. Finally, if the musical stream does not provide reorientation challenges to the listener, it will require less attention—which I will understand here to be less interest. On the other hand, if everything in the musical stream to seem like a novelty to the listener, music will present itself as something insuperably confusing.

Considering the direct relationship between commutative events which promote discontinuities, and the attentional process, it is necessary to point out attention here can be understood as mental "allocation of resources" to treat a specific stimulus. The attention should be treated as a limited capacity to process the data of consciousness, and this allocation can be intentionally controlled. Studies developed in the last decades focus on the attention device in different paradigms. "Task-defined" attention is a purely descriptive method of attention, demonstrated when the subject could satisfactorily fulfill a task that required the selection of a specific stimulus over other present stimuli (Brignani, Lepsien, & Nobre, 2010; Desimone & Duncan, 1995; Folk & Remington, 1999; Yantis & Egeth, 1999). That is, attention is inferred based on the quality of the fulfillment of a task that would have required the subject to isolate the given stimulus. An understanding of attention as a psychological process accurately describes it as a "process-oriented" mind-focusing activity (Luck & Vecera, 2002), an active device for selecting one of many possible sensory stimuli or threads of thought, with the purpose of optimizing the data quality focused and the effectiveness of the mental process. Thus, attention would be needed when faced with an extraordinary density of stimuli and tasks mental processes require to operate satisfactorily.

Finally, what must be stressed is that attention is captured whenever the system detects the presence of novelty, that is, what motivates the capture of attention is, in theory, the phenomenon of discontinuity. There are significant clues that this phenomenon reflects an "intentional" orientation of attention, so this makes way for the hypothesis that a complete understanding of the factors that determine whether an event automatically captures the individual's attention may be virtually unreachable, for example, habituation factors and cultural conditioning (Atchley, Kramer, & Hillstrom, 2000; Folk & Remington, 1999). Desimone and Duncan (1995) proposed a framework for the conceptualization of attentional control. It is the *biased competition model*, in which two classes of parameters influence the attentional control: the *bottom-up* parameters, based on the medium stimulation, and the *top-down*, directed at the target (in the environment), revealing distinct flow strategies in the perceptual process (Yantis, 2000). According to the model, bottom-up parameters include the sudden onset of a stimulus—which I want to understand here as a *sensory transient*—or discontinuity in the environment. Top-down parameters, in turn, include both a target pattern, that is, a mental representation of the target being intended (Desimone & Duncan, 1995; Duncan & Humphreys, 1989), and the individual's intentions about the stimulus spatiotemporal configuration, resulting from his or her conditioning.

In this study, I am understanding that model in enactivist terms. Thus, I must warn that a “mental representation” means a conceptual representation of the "target experience," that is, a process that starts from one or more cognitive schemes unconsciously activated in the act of perceiving the target to give meaning to the current experience. The linguistic statement of the understanding then formed is a mental representation "of that experience."

Orienting Response and the Musical Meaning

It is essential, however, to distinguish the process that triggers the attention-directing response—the orienting response—from the process that maintains attention—which begins at the *selection* stage as mentioned above. This is because there are pieces of evidence that may be distinct. Human sensitivity to changes in the environment is well known. The exposure to deviant stimuli in many sensory modalities can result in consequences on the attentional focus (Theeuwes and Chen, 2005; Vachon, Labonté, & Marsh, 2017). Research focusing on the auditory modality has shown that the presentation of a commutative sonic event that deviates from the recent auditory past can induce changes in attentional focus and like this diverting attention from the ongoing task. This disengagement of attention from the current task toward a commutative element in the musical environment refers to an orienting response. Such vulnerability to changes in the auditory environment can be explained by the application of an orienting response theory to music semantics and should be understood as the initial stage of the process of musical meaning construction.

We may admit the validity of theories that explain the musical understanding as a consequence of the comparison of the referential elements of the musical stream selected by the listener in the act of listening and his or her stylistic expectations. If so, we are operating at a level of memory that the researchers recognize as “maintenance of the attention.” The present study focuses, instead, strictly on the mechanisms that trigger the allocation of attention in the act of listening to music. The maintenance of attention, which is not discussed here, is most likely a distinct process that requires its approach and which would be more related to the experiences of listener’s expectancy and anticipation. Instead, I advocate the need for deepening the research on an issue I recognize as neglected by contemporary musical semantics: what is the nature of the bottom-up and top-down factors that trigger what I am acknowledging as musical orienting responses. Also how do commutative events regulate the resource allocation involved in the act of listening? In this direction, I want to believe that the investigation of the linguistic descriptions nature of our musical understanding can reveal important data about the processes that conceptually translate this understanding—but also about which and how the events of the musical surface would have guided and regulated our understanding.

Reisenzein, Meyer, and Schützwohl (1996) understood the orientation response as a probabilistic syndrome of responses provoked, in particular, by the novelty, including several behavioral and physiological components. Addressing what they understood to be the conditions under which the experience of novelty occurs, Reisenzein, Meyer, and Niepel (2012) proposed by way of example that a simple musical event, a note, presented "for the first time" to the listener, would be a novelty in the first sense. A note perceived as an occasional deviation, a "strange" note that prevents the occurrence of a given expected melodic sequence would be a novelty in a second sense. Lastly, a melodic contour that does not follow the pattern of continuity established by the previous contours, contradicting them, would be a novelty in a third sense. Several researchers have investigated how the perceptual system interacts with the environment to structure it (Bregman, 1990, 1991; McAdams & Drake, 2002). In the present study, I am considering that the perceptual organization of concurrent sound events in an auditory scene can take three configurations. It may result in *perceptual fusion* when two or more potentially distinct events are perceived as a single event. It may also result in a *perceptual grouping*, when components of the auditory scene perceived as independent events are grouped perceptively, forming contours and distinguishable regions. Moreover, it may result in *perceptual segregation* when the complex components of the scene (groupings) are perceived as a competing “substream” of a complex stream. An auditory event refers here to a sonic unit with limited temporal extension, experienced when physical actions make vibrating sound sources; it is necessary to observe that the same source can produce sound effects perceived as simple or complex.

Recent research on the perceptual organization of the auditory scene has been reiterating that there is a relatively limited number of types of acoustic cues that signal the constitution of the scene in its components. These cues indicate various attributes of the scene, regulating the possibilities of their apprehension by the individual. The understanding of what happens in the scene will also result from the perceptual intent and cognitive and cultural resources that the perceiver involves in the process. The central question here is how acts of apprehension—musical orienting responses—can signal the presence of cues (be it a simple event, a point grouping or the starting point of a more extensive sequential grouping), from what the processes of selection and segregation of the musical auditory scene will begin.

What types of acoustic cues provoke orienting responses in the music experience? Considering both the task-defined and process-oriented paradigms, I propose a primary classification of fundamental attributes of musical acoustic clues in four categories. The *tonal* category considers the sound contents related to the perception of determined (or relatively determined) pitch and the conditions of compatibility between the tonal materials of the components of the music scene due to continuity, homogeneity, regularity, and symmetry. The *textural* category considers both the envelope of sound intensity and the spectral behavior (timbre) of the auditory scene events and the density of the stream, concerning the complexity of segregation. It occurs, because the listener perceives joint and common changes of these parameters as constituting clues for fusion and the sequential grouping of scene components, whereas independent and dissimilar (not "parallel") changes tend to signal potential stream segregation. I should recognize a *temporal* category, since sound events rarely start or disappear at the same moment, and therefore, the perceptual system assumes that components of the scene perceived as precisely synchronous must be fused as elements of the same event. On the other side, asynchronous components tend to generate "movement" and be grouped sequentially. Lastly, the *topographic* category, considering the emphasis given to the spatiality of the scene components, a mapping of positions in the phenomenal space of the musical stream, which emerges in the act of listening when the attentional process overlaps the event spatial position to the event.

Conclusion

What I am discussing is not a conceptual model that allows us to anticipate—in the form of rules—"what" will be identified by musical orientation responses as more significant events in a given musical stream. The model developed here has the objective of revealing which events of the musical stream would have stood out as acoustic cues for the perceptual system and examining the reasons why such events would achieve this prominence. Besides, it investigates how the processes of perceptive organization develop from that and result in the musical understanding declared by a listener. Whether the model may or may not offer significant resources for compositional or interpretive elaboration is the subject of further investigation.

In the experience of music, the listener has his performance restricted by the transience of musical events that flow in sound complexes in general of remarkable potential of stimulation. The hypothesis that underlies the conceptual model under development in the present study is the musical orienting response can be considered the trigger for the construction of the music understanding in the act of listening. I think the cognitive operations provoked by the following changes of attentional focus throughout the listening experience constitute a crucial point for the investigation of the musical understanding. This understanding will be completed with the selection of the events for the confrontation with memorized cognitive schemes, thus conditioning the production of meaning. The musical meaning is intimately associated with tensions, whose perception is regulated in the interaction of affordances of the musical text and skills of the listener. I argue that the process of conceptualizing our understanding maps the tensional narrative of the music we experience. This narrative consists of acoustic cues related to tonal, textural, temporal, and topographic factors of the musical auditory scene. The listener's orienting response to specific attributes of the musical stream implies an excludable choice of acoustic events that will determine a most stable process of selecting the configurations. So these configurations will be schemed, generating the preconceptual understanding of the experience. I believe that the conceptual descriptions of this understanding can reveal both the script of orienting responses occurred in the course of listening and the interdependent perceptual dimensions that predominated in the overall understanding of the piece of music or each of its sections.

The development of a research protocol that aims to attest the different dimensions of imaginative production involved in the hypothesis presented here should consider: (1) attentional processes based on "event" (musical object) and on "movement" (in the phenomenal space of the stream; (2) the class of bottom-up parameters (commutative events) that regulate attentional control; (3) the connectivity principle governing the partitioning of a musical stream by "contours" and "regions"; and (4) the sensorial transients of the musical stream related to changes of tonal, textural, temporal, and topographic content of the sonic components of the stream.

If the hypothesis discussed in the present study is valid, it is essential to investigate how the acoustic cues mentioned above are conceptualized in terms of what I have been debating since *O Ato da Escuta e a Semântica do Entendimento Musical* (*The Act of Listening and the Semantics of Musical Understanding*) (Nogueira, 2004). When we experience the musical stream, we appropriate it in three concurrent dimensions of imaginative production: a) categorizing its distinctive sonic traces in the form of "movements" from the variability of the sonic state of the stream; (b) producing formal images profoundly stylistic, resulting from habituation and, hence, from the recognition of invariance, recurrence, and contrast of patterns; and (c) establishing a communicative exchange between the imaginative aspects of the mind and the virtual symbolic predicates of the object of listening, which implies tensions, contrasts, and intentions.

Therefore, I believe that a conceptual model based on the overlapping of linguistic and commutative acoustic cues, that is, on the comparison between the conceptual understanding of a given musical segment and the mapping of the possible orienting responses that condition the understanding stated by the listener, can reveal essential perspectives of the semantic path that he or she performed. This calls for the development of an experimental protocol that involves both processes of identification of perceived and relevant acoustic cues in a given musical experience—regulated by their potential discrepancies in the stream, as well as by the schematic inferences of situated and culturally determined listeners—as monitoring of significant neurophysiological changes during the experiment, in order to corroborate the results of these processes.

References

Atchley, P.; Kramer, A. F.; Hillstrom, A. P. (2000). Contingent Capture for Onsets and Off-Sets: Attentional Set for Perceptual Transients. *Journal of Experimental Psychology: Human Perception & Performance, 26,* 594–606.

Barry, R. J. (1979). A Factor-Analytic Examination of the Unitary OR Concept. *Biological Psychology*, 8(3), 161–178.

Barry, R. J., MacDonald, B., & Rushby, J. A. (2011). Single-Trial Event-Related Potentials and the Orienting Reflex to Monaural Tones. *International Journal of Psychophysiology*, 79 (2), 127–136.

Bernstein, A. S. (1979). The Orienting Response as Novelty and Significance Detector: Reply to O'Gorman. *Psychophysiology*, 16(3), 263–273.

Bregman, A. S. (1990). *Auditory Scene Analysis: the Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Bregman, A. S. (1991). Using Brief Glimpses to Decompose Mixtures. In J. Sundberg; L. Nord; R. Carleson (Eds.), *Music, language, speech and brain* (pp. 284–293). London: Macmillan.

Brignani, D., Lepsien, J., & Nobre, A. (2010). Purely Endogenous Capture of Attention by Task-Defining Features Proceeds Independently from Spatial Attention. *NeuroImage*, 51(2), 859–866.

Desimone, R. & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18,* 193–222.

Duncan, J.; Humphreys, G. (1989). Visual Search and Stimulus Similarity. *Psychological Review, 96,* 433–458.

Eimer, M. (2014). The Neural Basis of Attentional Control in Visual Search. *Trends in Cognitive Sciences*, 18(10), 526–35.

Folk, C. L. & Remington, R. (1999). Can New Objects Override Attentional Control Settings? *Perception & Psychophysics, 61,* 727–739.

Gibson. J. J. (1977). The Theory of Affordances. In: Shaw, R.; Bransford, J. (Eds.), *Perceiving, Acting, and Knowing: Towards an Ecological Psychology*. New Jersey: Lawrence Erlbaum Associates, pp. 67–82.

Gibson, James J. (1979). *The Ecological Approach to Visual Perception.* Houghton Mifflin.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. Chicago and London: University of Chicago Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. Chicago and London: University of Chicago Press.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.

Luck, S. & Vecera, S. (2002). Attention. In H. Pashler. *Stevens' Handbook of Experimental Psychology*, 3rd Ed., (pp.235–286). New York: John Wiley & Sons, Inc.

McAdams, S. & Drake, C. (2002). Auditory Perception and Cognition. In H Pashler & S. Yantis (Eds.) *Stevens' Handbook of Experimental Psychology* (3rd ed, pp. 397–452). Hoboken, NJ: Wiley.

Meyer, L. B. (1956). *Emotion and Meaning in Music.* Chicago: University of Chicago Press.

Nogueira, M. (2004). *O Ato da Escuta e a Semântica do Entendimento Musical* [The Act of Listening and the Semantics of Musical Understanding] (Unpublished doctoral dissertation). Communication School/Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

Nogueira, M. (2016a). Resposta de Orientação Musical: Uma Hipótese para a Origem do Dispositivo de Sentido [Musical Orienting Response: A Hypothesis for the Origin of the Meaning Device]. *Música Hodie*, 16(1), 54–70. Federal University of Goiás, Goiânia, Brazil.

Nogueira, M. (2016b). O Sentido do Inesperado: Resposta de Orientação em Música [The Meaning of the Unexpected: Orienting Response in Music]. *Proceedings of the XII Simpósio Internacional de Cognição e Artes Musicais* (pp. 587–595). Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

Pavlov, I. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. (Translated and edited by G. V. Anrep). London: Oxford University Press.

Reisenzein, R., Meyer, W.-U., & Schützwohl, A. (1996). Reactions to Surprising Events: A Paradigm for Emotion Research. In N. Frijda (Ed.), *Proceedings of the 9th Conference of the International Society for Research on Emotions* (pp. 292–296). Toronto: ISRE.

Reisenzein, R., Meyer, W.-U., & Niepel, M. (2012). Surprise. In V. S. Ramachandran (Ed.), *The Encyclopedia of Human Behavior*, 2nd Ed., vol. 3, pp. 564–570 (Amsterdam: Elsevier).

Sechenov, I. (1863/1965). *Reflexes of the Brain*. Cambridge, Mass.: The MIT Press.

Sokolov, E. N. (1960). Neuronal Models and the Orienting Reflex. In Mary A. B. Brazier (ed.), *The Central Nervous System and Behavior*, pp. 187–276. New York: Josiah Macy, Jr. Foundation.

Sokolov, E. N. (1963). *Perception and the Conditioned Reflex.* Oxford: Pergamon Press.

Theeuwes, J. & Chen, C.Y.D. (2005). Attentional Capture and Inhibition (of Return): The Effect on Perceptual Sensitivity. *Perception & Psychophysics*, 67(8), 1305–1312.

Vachon, F., Labonté, K., & Marsh, J. E. (2017). Attentional Capture by Deviant Sounds: A Noncontingent Form of Auditory Distraction? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(4), 622–634.

Yantis, S. (2000). Goal-Directed and Stimulus- Driven Determinants of Attentional Control. In S. Monsell & J. Driver (Eds.), *Attention and performance: Vol. 18, Control of cognitive processes* (pp. 73–103). Cambridge: MIT Press.

Yantis, S. & Egeth, H. E. (1999). On the Distinction Between Visual Salience and Stimulus-Driven Attentional capture. *Journal of Experimental Psychology: Human Perception & Performance, 25,* 661–676.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

A multimodal analysis of vitality forms in the play Krapp's Last Tape

Alicia Nudler,[1] Silvia Español,[2] Paz Jacquier[3] and Adrián Porcel de Peralta[4]

[1,4]*Universidad Nacional de Río Negro, Argentina*
[2]*IICSAL, CONICET-FLACSO, Argentina*
[3]*Instituto de Formación Docente Continua Bariloche, Argentina*
[1]anudler@unrn.edu.ar, [2]silvia.ana.es@gmail.com, [3]pazjacquier@gmail.com, [4]japorceldeperalta@unrn.edu.ar

Abstract

The aim of this paper is to describe and analyze a theatre fragment utilizing the concept of forms of vitality as proposed by developmental psychologist Daniel Stern. The selected fragment belongs to the play Krapp's Last Tape by Samuel Beckett. Three performances of it, each by a renowned actor, were subjected to an ELAN program. A comparative analysis was made using an ad-hoc observational code containing four categories: forms of vitality in the actor´s movements, voice, scene sounds other than voice, and lighting. Subcategories were created based on attributes described by Stern (i.e. sudden, gentle, effortful), and were assigned along a temporal line. Prevalent forms of vitality per category were identified in each case, plus a comparison between categories. Finally, a description of the effect of these different display and combination of forms of vitality on the experience of spectators is discussed.

Introduction

Samuel Beckett´s play Krapp´s Last Tape, written and performed for the first time in 1958, features one character, an old man obsessed with the tapes he has been recording for many years, reflecting the changes in his life. The play therefore deals strongly with the issue of time. Three performances of this play were selected for this paper: Robert Wilson (under his own direction); John Hurt (directed by Atom Egoyan in a film version) and the Argentinian actor Héctor Bidonde (directed by Augusto Pérez). One section from the play – as Krapp first speaks, while looking for a particular tape – was chosen. The three versions of this short excerpt were analyzed and compared utilizing the concept of *forms of vitality* proposed by developmental psychologist Daniel Stern (1985, 2010). Forms of vitality are perceptual Gestalts created by the integration of movement, force, space, time and direction/intentionality. They belong to not just one sensory modality but to several: vision, hearing, touch. Regardless of content, forms of vitality concern the specific way in which dynamic events unfold in space and time, being crucial both to interpersonal encounters and the performing arts.

Background

The notion of forms of vitality was proposed by Daniel Stern originally under the name "vitality affects" in the context of studying mother-infant interactions (Stern 1985). Later he further specified the concept and broadened its scope to psychotherapy and the arts (Stern 2010).

There are few empirical studies of forms of vitality; they have been conducted in the area of tango music and dance (Shifres, Pereira Ghiena, Herrera and Bordoni 2012), neuroscience (Di Cesare, Di Dio, Rochat, Sinigaglia, Bruschweiler-Stern, Stern, and Rizzolatti, 2013), theater (Stern, 2010), autism (Rochat, Veroni, Bruschweiler-Stern, Pieraccini, Bonnet-Brilhault, Barthélémy, Malvy, Sinigaglia, Stern and Rizzolatti 2013), and developmental psychology (Español, Martínez, Bordoni, Camarasa and Carretero 2014; Español, Bordoni, Martínez, Camarasa and Carretero 2015; Español and Shifres 2015; Martínez, Español and Pérez 2018). A direct antecedent to this paper is Weeks (2013) who also studied Krapp´s Last Tape performed by Wilson and Hurt, utilizing the concept of forms of vitality.

Method

The three performances were subjected to the ELAN, a computer program with a tier-based data model for multi-level, multi-participant annotation of time-based media.

An ad-hoc observational code containing four categories was created: forms of vitality in the actor´s movements, voice, sound scenes other than voice, and lighting. Voice was analyzed as a separate category from other scene sounds because of its obvious importance as carrier of linguistic meaning.

Each excerpt was analyzed in detail using the program. With repeated observations by all four authors, subcategories based on attributes described by Stern were created for each category and assigned along a temporal line, as shown in Fig. 1.

Prevalent forms of vitality per category were identified in each version. Also, a comparison to understand the relation between categories was made.

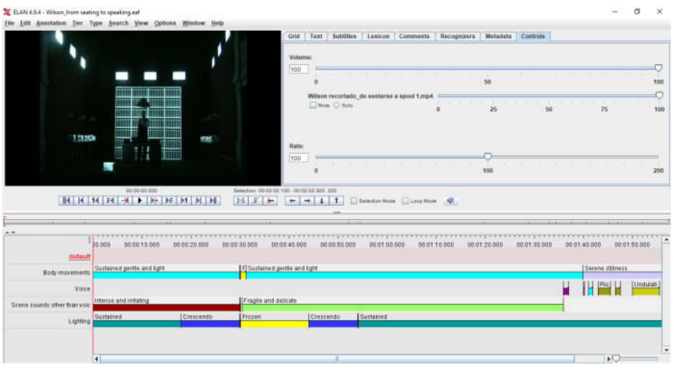


Figure. 1. A view of the tiers in the ELAN program analyzing Wilson´s version of the excerpt. In the upper part of the graphic with colours the timeline can be seen. The four categories are shown on the left, in black font: body´s movements, voice, sound scenes other than voice, lighting. Subcategories assigned to each moment are shown in different colors: sustained, gentle, frozen, etc.

Results

Wilson´s version is particularly refined in all four categories. His movements are sustained, gentle and light most of the time, with only one abrupt and one frozen moment. His voice has clear profiles which are either effortful, directed, sustained or undulating. The rest of sound scenes can be divided in two long moments: intense and irritating first, very fragile and delicate later. These two moments are clearly separated by one clear cut abrupt sound, and then a long moment of silence which in perceptual experience feels frozen and intense as if it were a sound itself. His use of lighting displays crescendo, sustained and frozen forms of vitality with only one moment of abruptness. Regarding the relation among the four categories, sometimes forms of vitality coincide in light, body movement and sound, this coincidence being clearly emphasized. One of these moments is created by an abruptness in movement, sound and lighting, followed by a coincidence in the frozen quality of the three. At some specific moments the four categories are in contrast with one another, and these contrasts are also deliberately emphasized. One of these instances happens at the beginning of the excerpt (in the whole performance it starts at the beginning of the play, to which the moment being analyzed is contiguous), when body movements are sustained, gentle and light along with a sustained lighting, but, in contrast, the sound is intense and irritating.

Bidonde´s body movements are mainly energetic and directed, with some moments of sustained, release, gentle and stillness. His voice is also mainly energetic and directed, effortful, sustained and releasing in short moments. The rest of scene sounds, scarce, are clear cut and intense irritating. Lighting does not play a specific role but, closer to everyday lighting, is sustained.

In Hurt´s version, one long moment of frozen movement and two other moments of almost frozen are underlined. There are also several moments of effortful movements, abrupt and energetic release. His voice is, in convergence with movements, also effortful, releasing, with just one moment of sequential suspension and one of sustained quality. The rest of the sounds are, like in Bidonde´s case, everyday sounds, but in Hurt´s version the sound of a gentle rain is present, so there is a sustained and light sound quality all through the fragment, only interrupted by the sudden noises of some of the character´s actions.

As can be seen in fig. 2, in Wilson´s time line there are long phrases in movement, sounds and lighting. Contrasts and coincidences are carefully elaborated in multimodal compositions. In the first phrase, the intensity of sound opposes the gentleness of movement. In the second phrase, the fragile and delicate sound goes along with the sustained, gentle and light movement and the crescendo and sustained qualities of lighting. The frozen moments are carefully simultaneous.

Figure 2. Wilson´s time line analysis.

In Bidonde´s version, the lighting is constant. There are subsequent sequences of energy and release which are not elaborate in multimodal compositions with lighting like in Wilson´s but sometimes are combined with sounds.

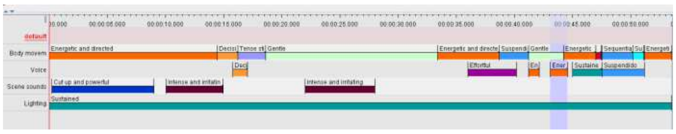


Figure 3. Bidonde´s time line analysis.

In Hurt´s case, the frozen quality of his movements is prevalent, surrounded by some energetic ones. The multimodal compositions are scarce. Several of the phrases in the first two categories are short.

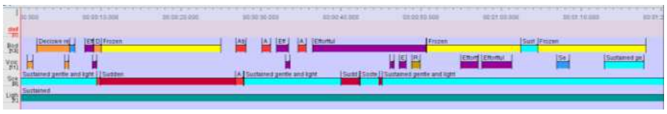


Figure 4. Hurt´s time line analysis.

These differences between versions create quite diverse effects in perceptual experiences, described below in the discussion.

Discussion

Wilson's version of this play provokes an intense aesthetic experience, while also generating a sense of distance between the spectator and the inner world of the character. Bidonde´s version draws the spectator much closer to the psychology of the character: it portrays a grumpy but also energetic, impulsive and somewhat clumsy Krapp who brings forth a feeling of simpathy in the audience. Hurt´s performance, instead, embodies a depressive, nostalgic, reflecting Krapp, eliciting feelings closer to hopelesness and dispair.

Among other differences between the productions, it is to be noted that the rain effect, present in both Hurt´s and Wilson´s versions, is used differently in each case. In the first, the rain causes a very gentle and light sound, rendering a nostalgic atmosphere, in tone with the character´s personality portrayed here and, also, as a reference to constant rain in Becketts´ homeland, Ireland. In Wilson´s case, instead, the rain is present from the very start of the play, as a heavy, all encompassing constant strong sound that becomes intense and irritating. This rain in the actual performance lasts for a long time (approximately 23 minutes) and stops with a startling sound followed by prolonged silence. This rain, a storm here, creates a different experience in the spectator from that in Hurt´s version: by being encompassing, overwhelming, it generates a silencing effect in the actor´s movements which stresses the distance with him, creating at the same time a powerful aesthetic experience. This strong prolonged sound is also an interesting prelude to the upcoming silence, which then becomes intense too.

Two main conclusions can be drawn from the previous results. The first stresses the importance and effect of forms of vitality already stated by Stern. All three versions of the play analyzed here are loyal to Beckett´s written theater piece. However, the different styles with which actions and movements are done and staged generate markedly different experiences in the spectator, conveying diverging psychological traits of the character. In other words, forms of vitality, in their different modalities and in their combination, generate different meanings and sensorial experiencies, even though the basic narrative and the words being spoken are the same.

The other conclusion is related to the theatrical traditions to which these versions belong. Bidonde´s and Hurt´s versions suscribe to a naturalistic canon, which accounts for several of the differences with Wilson´s version. In the latter case, the performance may be related to certain aspects of contemporary visual art, performance art, and choreography (Goldberg 2004:65), and the whole excerpt appears as a composition of forms of vitality in which the coincidence of two or three sensorial modalities frames and highlights specific moments of contrast. In the other two performances, forms of vitality are not subject to a special refinement and work: they appear closer to the ways in which forms of vitality display in everyday life.

Stern states that forms of vitality are shown in a purified form by the arts because dynamic features are usually amplified, refined and repeated in performances (Stern 2012: 75). However, he limits his study of these forms to the time-based arts, which for him are music, dance, cinema and “certain theater”. He thinks that these arts take place in real time, while language based-based art, like traditional theater, fiction and poetry are usually driven by the narrative process and take place in both real time and narrative time, thus complicating the situation.

We can see in the results of this paper that, in line with Stern´s statements, in the more “time-based” performance by Wilson, forms of vitality are clearly refined and elaborated as multimodal compositions, while in the more “traditional” theatrical performances by both Bidonde and Hurt, forms of vitality are present, but their occurence is similar to everyday life: there does not seem to be a specific or clear elaboration of them.

This paper is meant to contribute to the systematic study of the sensorial non-verbal experiences elicited by theater. It also aims at developing a method for the empirical study of forms of vitality as dynamic events, of which time and movement are key aspects. Finally, it intends to deepen our comprehension about how multimodality and the relation between the different senses work.

References

Di Cesare, G.; Di Dio, C.;Rochat, M.; Sinigaglia, C.; Bruschweiler-Stern, N.; Stern, D., and Rizzolatti, G. (2013). The neural correlates of vitality form recognition: an fMRI study. *Social Cognitive and Affective Neuroscience Advance Access*.

Español, S.; Martínez, M.; Bordoni, M., Camarasa, R.; Carretero, S. (2014). Forms of vitality play in infancy. *Integrative Psychological And Behavioral Science*, 1-27.

Español*, S.*, Bordoni, M, Martínez M., Camarasa, R. y Carretero, S. (2015). Forms of vitality play and symbolic play during the third year of life. *Infant Behavior and Development. 40*, 242-251.

Español, S. y Shifres, F. (2015). The artistic infant directed performance: A mycroanalysis of the adult´s movements and sounds. *Integrative Psychological and Behavioral Science*, *49*, 371-397.

Goldberg, R. (2004) *Performance: Live Art Since the 60s*. London: Thames and Hudson.

Martínez, I. C., Español, S. A., & Pérez, D. I. (2018). The Interactive Origin and the Aesthetic Modelling of Image-Schemas and Primary Metaphors. *Integrative Psychological and Behavioral Science*, 1-26.

Rochat, M.; Veroni, V.; Bruschweiler-Stern, N.; Pieraccini, C.; Bonnet-Brilhault, F.; Barthélémy, C.; Malvy, J.; Sinigaglia, C.; Stern, D.; Rizzolatti, G. (2013). Impaired vitality form recognition in autism. *Neuropsychologia*,*51*, 1918–1924.

Shifres, F.; Pereira Ghiena, A.; Herrera, R. y Bordoni, M. (2012). Estilo de ejecución musical y de danza en el tango: atributos, competencia y experiencia dinámica. *Cuadernos de Música, Artes Visuales y Artes Escénicas, 7, 2*.

Stern, D. (1985).*The interpersonal world of the infant. A view from psychoanalysis and developmental psychology.* New York: Basic Books.

Stern, D. (2010). *Forms of vitality: Exploring dynamic experience in psychology, the arts, psychotherapy, and development.* Oxford: University Press.

Weeks, N. (2013). The dynamics of embodied gestures in the theatre of Robert Wilson. *Interdisciplinary Network Conference on Performance,* 1-14.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Igor Stravinsky and György Ligeti Strings Music Textural Similarities

Carlos Alberto Bastos Oliva

Institute of Arts, UNESP, Brazil

pollacooliva@gmail.com

Abstract

This poster/communication is a partial result of a master dissertation developed under guidance of Prof. PhD Marcos Mesquita in the research group Cogmus. The text compares two string music excerpts which denote some similarities perceived through hearing experience. The first one is in Igor Stravinsky's Firebird Suite finished in 1910; the second is in György Ligeti's String Quartet nº 2 which was written in 1968. A segment of both pieces presents a similar sounding structure that is to be compared using music parameters such as rhythmic direction evolvement; chromatic voicing; harmony and texture.

Introduction

A little over half a century apart both, Igor Stravinsky's Firebird Suite performed first time on June 25th 1910 and György Ligeti's 2nd String Quartet whose first performance was on December 14th 1969, present a string music passage which apparently shows a similar sonority effect that might be traced through an analysis using various musical parameters. Articulation, dynamics, chromatic harmonic approach, texture and the evolvement from a large vertical axis to a short one, seemingly are factors that make their musical hearing result much alike. Both excerpts present a somehow chromatic harmony approach; similar metronome marks; equal dynamics and articulation besides a large vertical axis at the beginning which shortens while evolving to the end of sections.

On Firebird's B*erceuse* movement from m.10 through m.15 (end of the movement) the string section of the orchestra performs a *sul tasto* passage evolving a chromatic chord progression that brings out a musical effect that shows some kind of similarity to Ligeti's 2nd Quartet 1st movement from 2nd half of m79 to 1st half of m.84. See Figures 1 e 2.

Figure 1. Firebird Suite's Berceuse movement last 6 measures.

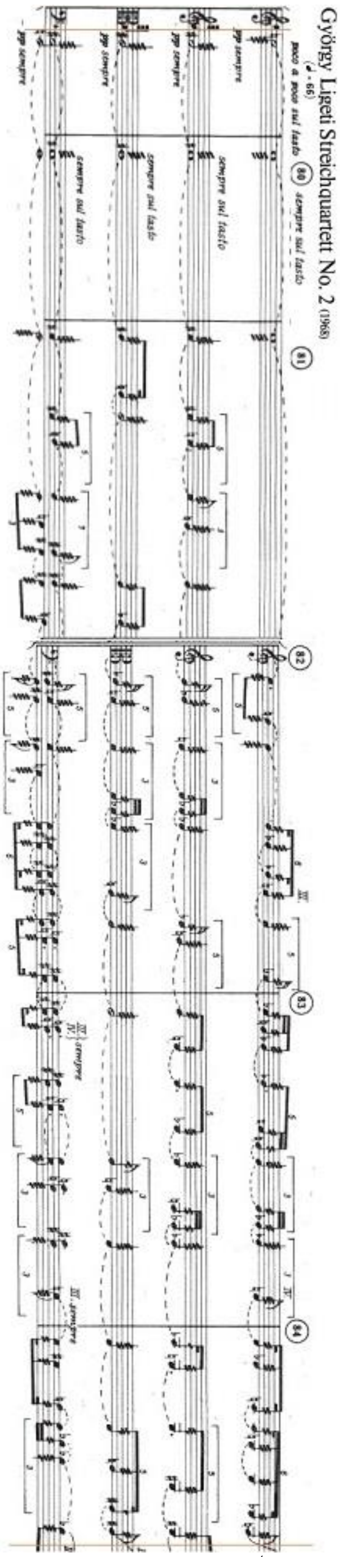


Figure 2. Ligeti's Quartet nº 2, 1st movement m.79 to m.84. With permission of SCHOTT MUSIC, Mainz - Germany

Unidirectional evolvement

The On Firebird's excerpt the pitch range runs from G5 to G1 starting on m.10 narrowing to B4 to G#3 at the end. All the voices move straight forward performing an equal rhythmic value creating a vertical chord progression. See Figure 3:

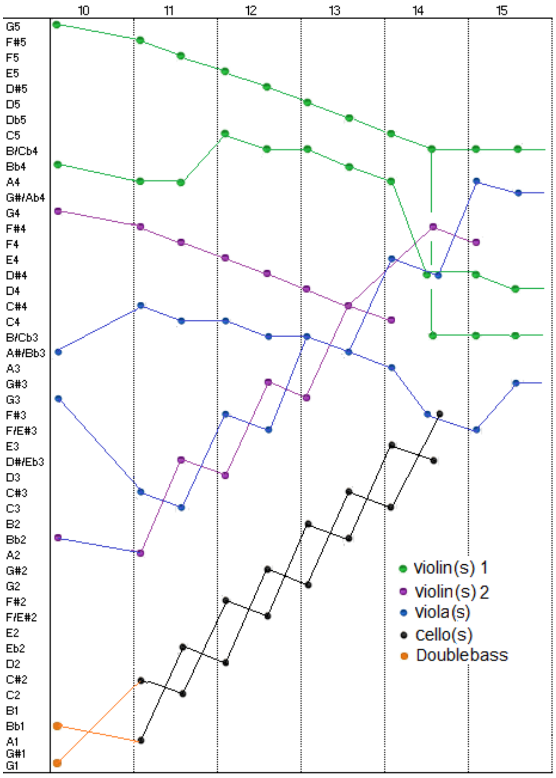
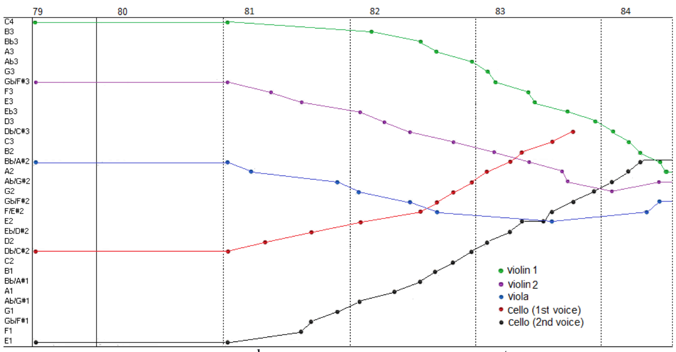


Figure 3. Firebird's *Berceuse* movement pitch range from m.10 to m.15.

Ligeti's passage also shows a large pitch range starting on the 2nd half of m.79 covering from C4 to E1. Likewise it narrows at the end between Bb2 and Gb2 on the 3rd beat of m.84 with all the voices in unidirectional evolvement. See Figure 4.



Ex.4 Pitch range from 2nd half of m.79 through 1st half of m.84 on Quartet's 1st movement.

Rhythm

On Firebird's excerpt the voicing evolvement displays a chordal type of progression with homorhythmic disposal (Berry, 1987). The rhythmic development of all voices occurs in parallel mode, i.e. starts with a whole note for each voice on measure 10 followed by two half notes per voice on the remaining measures, see Figure1.

While examining Ligeti's 2nd Quartet we perceive that the rhythmic evolvement of the voices remains parallel from the start point; 2nd half of m. 79 up to ¾ of the first beat in m. 81when the viola goes down a half step from A#2 to A2. See Figure 5:

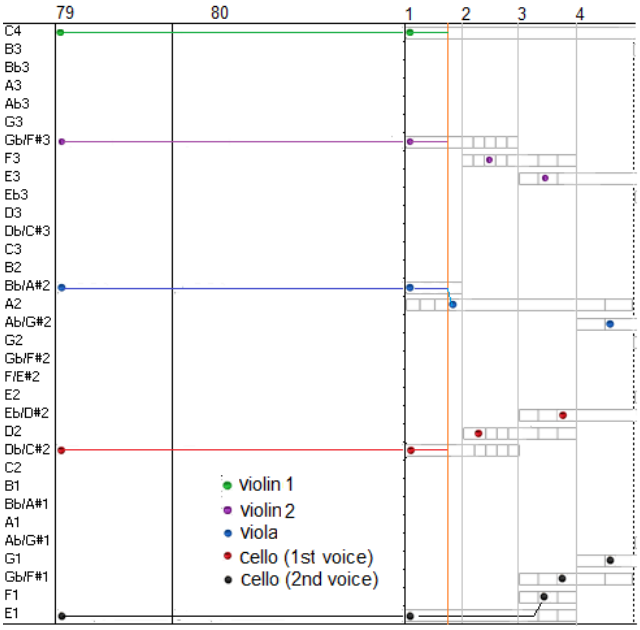


Figure 5. Parallel voice movement up to ¾ of measure 81.

From that point on each voice performs note attacks on different parts of the beat within rhythmic variations dividing the beat in 2, 3, 4, 5 and 6 parts (Vitale, Claudio 2013), see Figure 6.

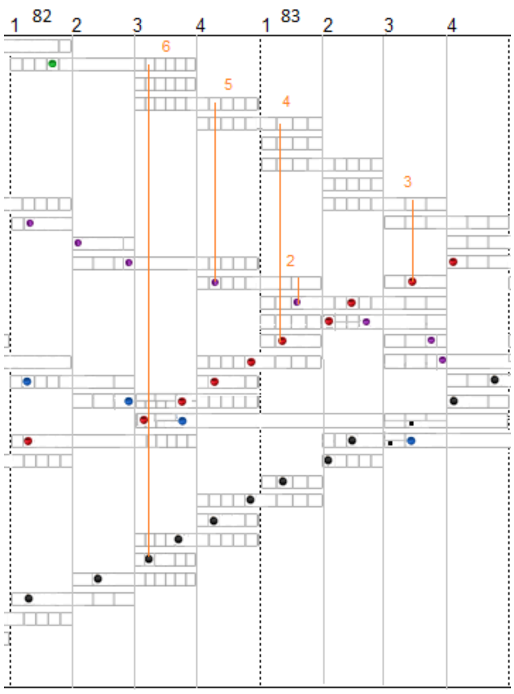


Figure 6. Measures 82 and 83 with note attacks in different parts of the beat.

Chromatic voicing

Firebird Suite's excerpt was written in the beginning of 20th century, most precisely between 1909 and 1910. It presents a rather tonal harmonic progression although, mostly with chromatic movement from the upper voices. See Figure 7:

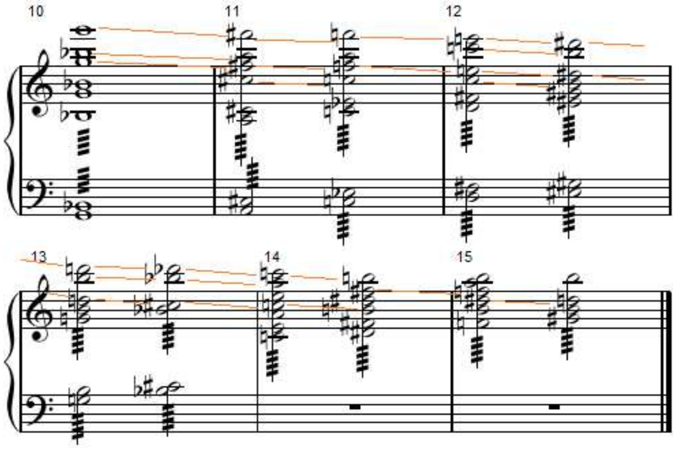


Figure 7.1. Chord progression with the upper voices descending chromatically.

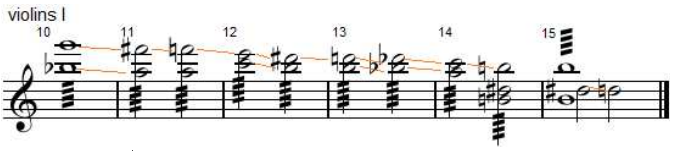


Figure 7.2. 1st violins lines.

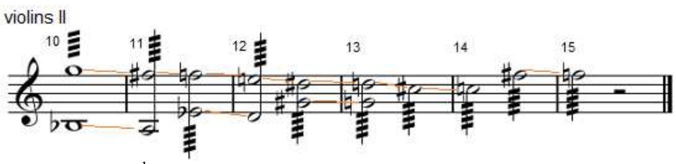


Figure 7.3. 2nd violins lines.

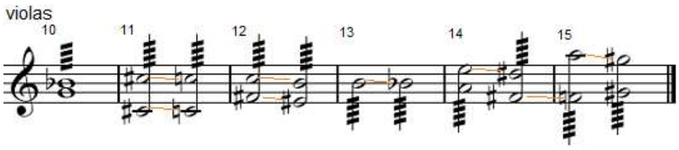


Figure 7.4. Viola lines.

György Ligeti used to voice the parts in a chromatic way thus achieving the micropolyphonic effect (Bernard, 1994). See Figure 8:

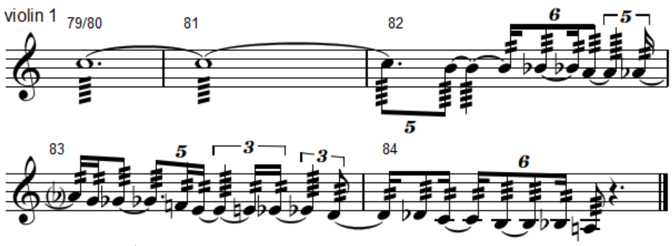


Figure 8.1. 1st Violin.

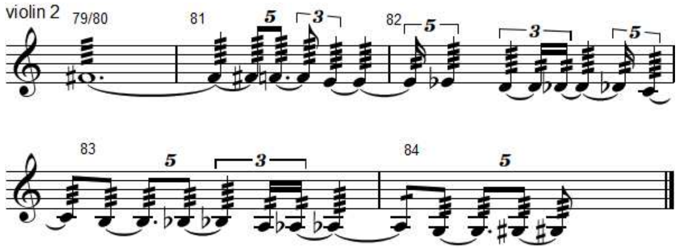


Figure 8.2. 2nd Violin.

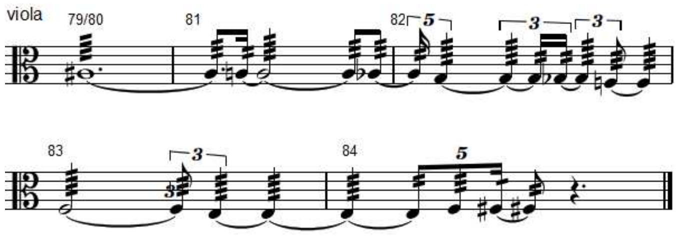


Figure 8.3. Viola.

Figure 8.4. Cello.

Harmony

Looking at the harmonic progression on Stravinsky's strings excerpt it is possible to observe an intense tonal chord change movement although, it configures a chromatic harmonic passage rather than a key centered one. Along with the vertical alignment of the voices forming noticeable chord structures, it is also possible to observe the resulting horizontal chromatic lines occurring in the upper voices. See Figure 9:

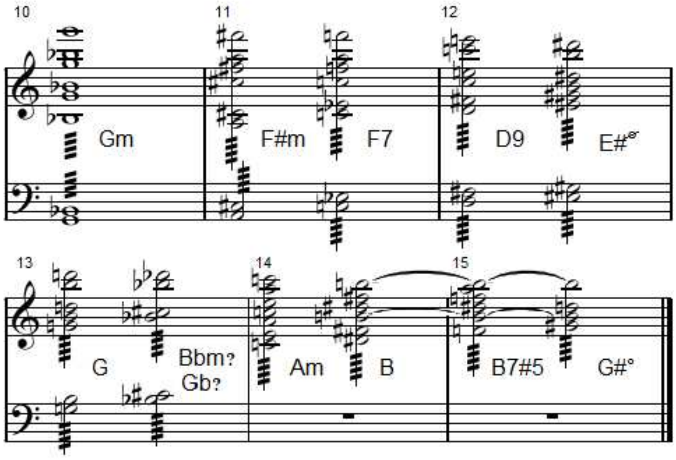


Figure 9. Chords formed from the vertical alignment of the voices.

On Ligeti's piece along with the fact that note attacks are in different parts of the beats creating a "spread out" fully chromatic movement of all voices, it is possible to perceive some vertical alignments among the voices performing detectable harmonic blocks. The metronome mark indicated by the composer is one of sixty six beats per minute. With this speed along with a 4/4 time signature, one quarter note will sound for about 1.1 seconds. The vertical alignment of certain notes from 2nd half of m.79 up to the 3rd beat of m.82 may be understood as creating harmony block chords of dense harmonic sounding result. Despite the shortness of time remaining in each vertical formation, it is possible to recognize some kind of chord progression for about 13.2 seconds. Distances among the voices are also numbered in half steps. The first block/chord formed on the 3rd beat of m.79 last 7.43 seconds up to the 3rd quarter of first beat on m.81. See Figures 10.1 to 10.6:

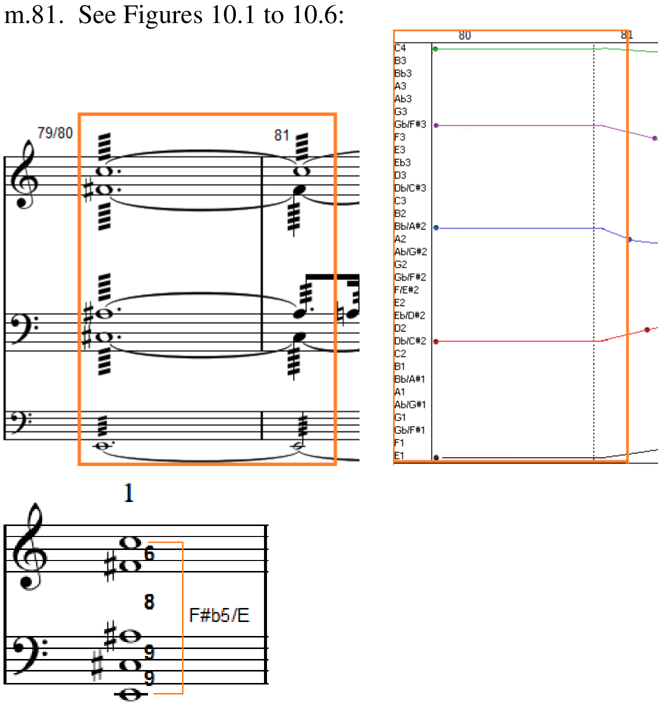


Figure 10.1. Harmonic block/chord C / F#/E formed in first 7.43 seconds.

In m.81 from 3/5 of 2nd beat up to 1/3 of 3rd beat a new chord formation happens once 2nd violin drops from F#3 to F3 and cello goes up a half step from C#2 to D2. It results in C – F – Dm7 and this block/chord remains for a whole second.

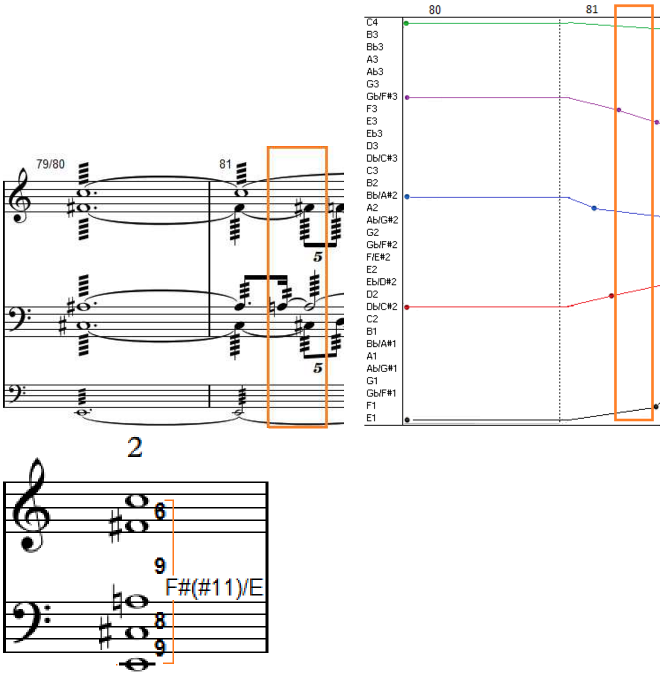


Figure 10.2. Block/chord F#(#11)/E.

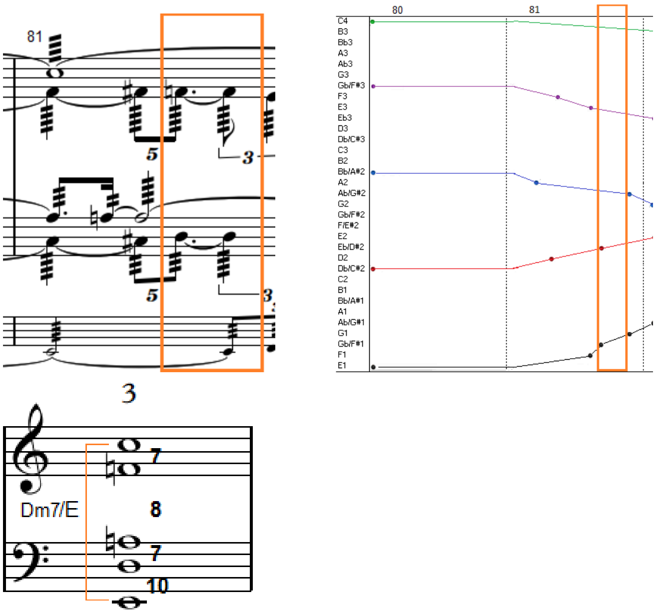


Figure 10.3. Block/chord C – F – Dm7.

Starting on the 2nd half of the last beat on m.81 a new harmonic formation occurs once the viola drops from A3 to Ab3 and the bottom voice of the cello goes up a half step from F#1 to G1. This block/chord sounds for 0.72 of a second and results in C – Ab – E augmented.

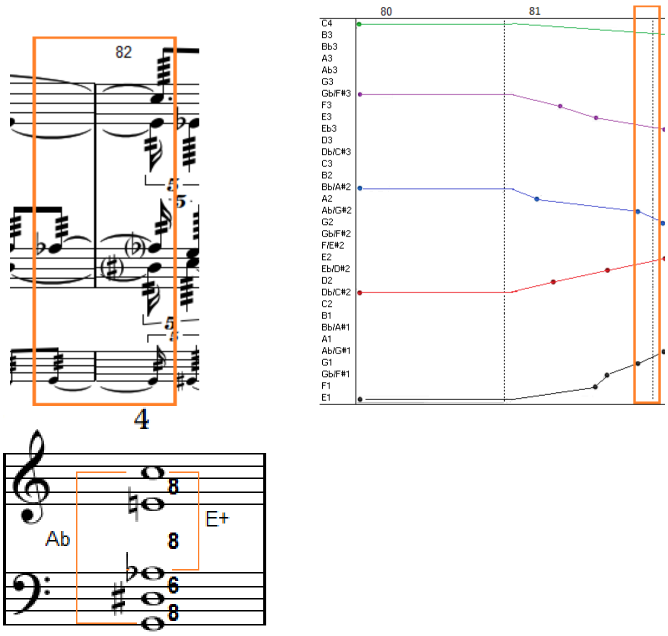


Figura 10.4. Block/chord C – Ab – Eaugmented.

The 5th block/chord happens when all the voices move except 1st violin. 2nd violin descends from E3 to Eb3; viola from Ab2 to G2; cello upper voice ascends from D#2 to E2 and bottom voice from G1 to G#1. This lasts for 0,88 of a second with 1st violin going down from C4 to B3 in about 0,44 second of the chord duration.

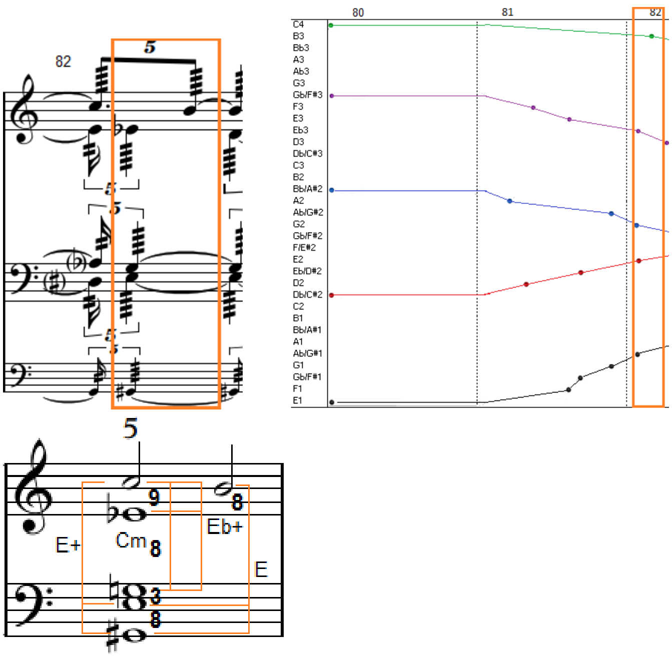


Figura 10.5 Block/chord Eaug – Cm for 0,44 of a second and E – Ebaug for another 0,44s.

The last block/chord detected before voice movement become much intense is the one that happens from the down to the 5/6 of the 2nd beat on m.82. This one lasts for 0,92 of a second and also has two states when bottom voice of cello goes up from G#1 to A1 on 2/3 of the beat.

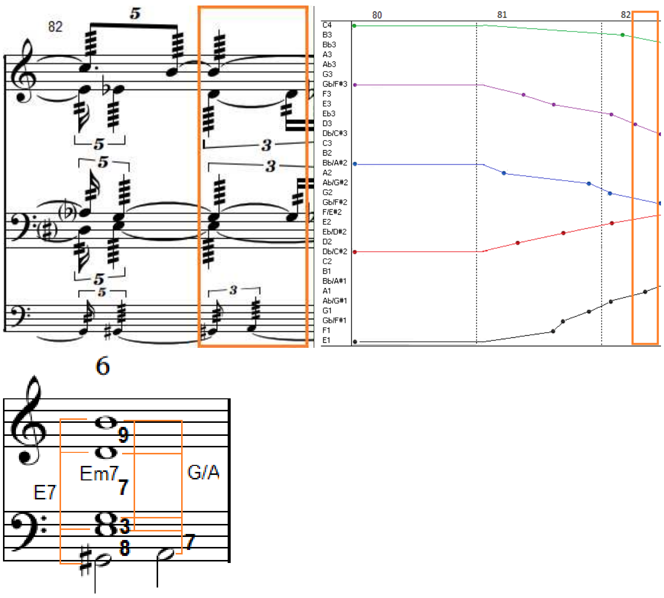


Figure 10.6 Block/chord E7 – Em7 that turns G/A (or A7sus4).

Conclusion

Due to the great amount of similarities in musical parameters such as instrumentation, dynamics, articulation, speed, chromatic voicing, a large pitch range at the start point ending on a much shorter one, still some similar harmonic features in the chord changes using chromatic movement within the voices, both excerpts present a very close sounding texture. The overall sonority of both excerpts is determined by similarly textural conditions. Use of strings with the same articulation; linear and contra directional character of voice evolvement; high density and compression levels at start points due to a large vertical axis created by the distance among the extreme voices which shortens towards the end (Berry, 1987); (Guigue,2004). This last feature is more evident on Firebird's excerpt once some voices start to perform same notes near to the end. See Figures 11.1 and 11.2:

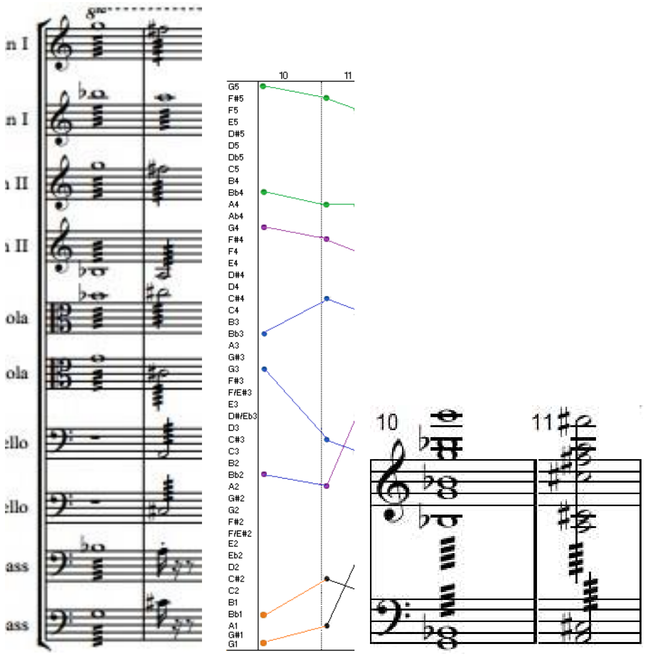
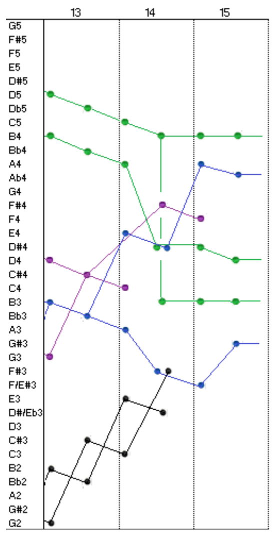


Figure 11.1. Start point of Firebird's excerpt on m.10.



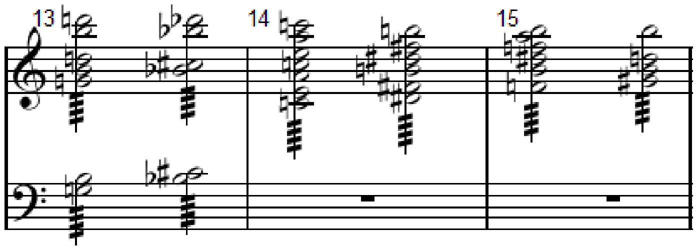


Figura 11.2 Last three measures of Firebird's excerpt.

On Ligeti's 2nd Quartet excerpt same thing happens as far as the vertical movement. It starts with a rather large axis at the 2nd half of m.79 and shortens towards the end on 3rd beat of m84. The big difference here is the gradual change of direction with the violins going down and the viola and cello going up both chromatic; and for the note attacks of the four instruments being in different parts of the beats in a *legato* articulation, that assures the continuity of the sonority as a whole (Vitale Claudio, 2013). See Figures 12.1 and 12.2:

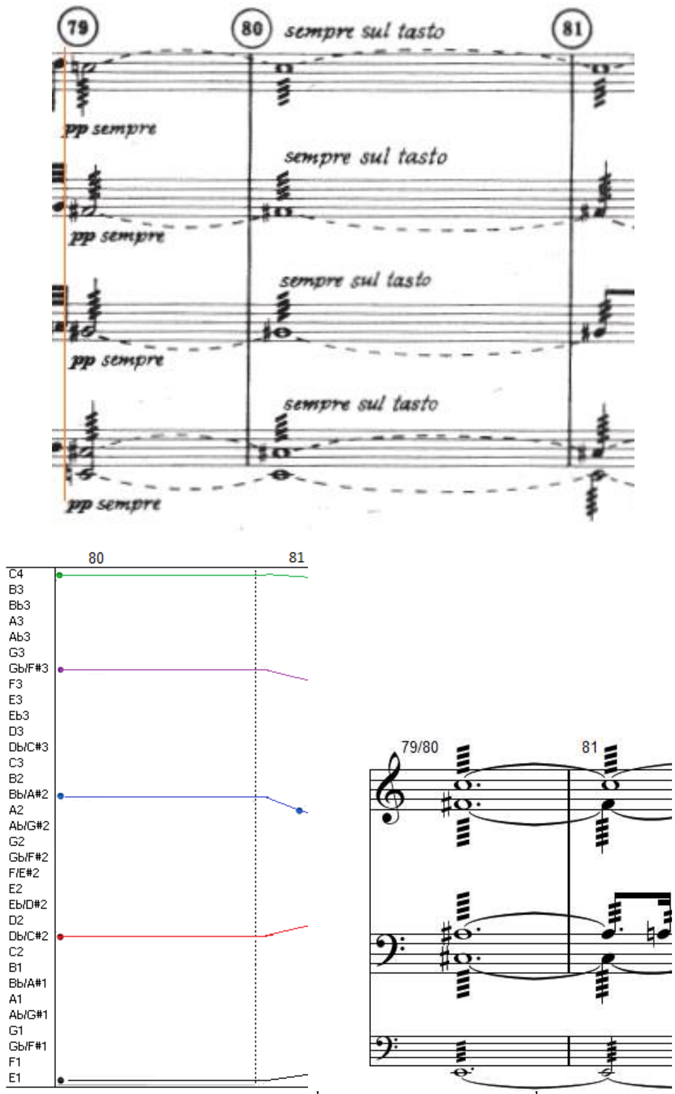


Figure 12.1. Starting point of 2nd Quartet excerpt on 2nd half of m.79.

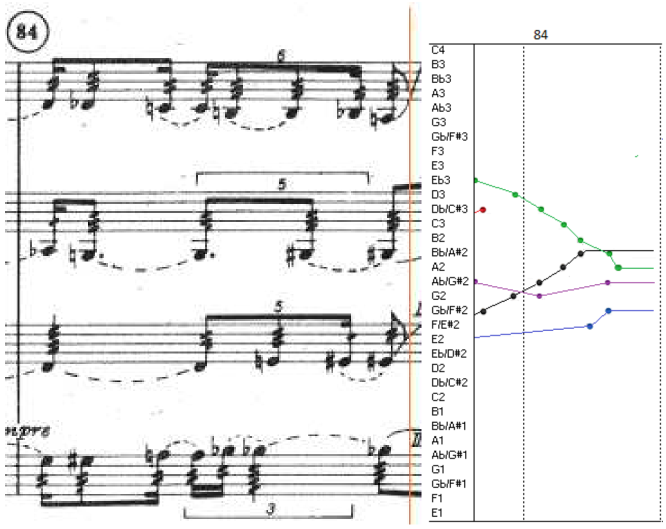


Figure 12.2 2nd Quartet's excerpt last measure.

After a detailed analysis approaching formal and traditional analytical tools such as pitch relation; rhythmic evolvement and harmonic features, it is possible to prove and assert that the hearing experience might bring out sounding parameters that are not directly written on the music paper but can be perceived in an attentive listening fruition.

György Ligeti achieved a kind of continuous sound texture out of the string quartet with this timewise odd note attacks. Different than that, Igor Stravinsky wrote similarly in direction but with a parallel rhythmic evolvement. Although separate in time for over 50 years and presenting much different compositional procedures both segments end up resulting a very similar sound effect.

The analysis testifies that the sounding similarities perceived by the hearing perception may be demonstrated positively. This might be useful as a possibility of bringing the hearing cognition to a deeper level in the study of music.

Acknowledgements.

Marcos José Cruz Mesquita, phd

All members of research group COGMUS

References

Bernard, Jonathan W. Inaudible structures, audible music: Ligeti's problem, and his solution. EUA: Music Analysis, Vol. 6, Nº. 3, 1987.

Bernard, Jonathan W. Voice Leading as a Spatial Function in the Music of Ligeti. EUA: Music Analysis, Vol. 13, No. 2/3, 1994.

Bernard Jonathan W. Ligeti's restoration of interval and its significance for his later works. EUA, Music Theory Spectrum, Vol. 21, No. 1, 1999.

Berry, W. Form in Music. New York: Pearson, ed. 2, 1985.

Berry, W. Structural functions in music. New York: Dover Publications, ed. 2, 1976.

Cope, David. Techniques of the Contemporary Composer. EUA: Cengage Learning, ed. 1, 1997.

Cope, David. Fundamentals of Statistical Analysis. EUA: Foundation Press, ed. 1, 2005.

Guigue, Didier. Esthétique de la sonorité: l'héritage de Debussy dans la musique pour piano du xxe siècle. [Sounding Aesthetics: Debussy's heritage in the piano music of 20th century]. France: Editions L'Harmattan, 2004.

Schoenberg, A. Fundamentals of composition, London: Faber and Faber, ed. 2, 1970.

Vitale, Claudio H. A gradação nas obras de György Ligeti dos anos sessenta [Rhythm graduation in Geörgy Ligeti works in the 1960's]. (Published doctoral dissertation). Brazil: Universidade de São Paulo, 2013.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Induced emotion within musical experiences: testing the universal paralanguage theory via musical sounds.

Joseph S. Plazak,[1] Zachary A. Silver,[1]

[1] *School of Music, Illinois Wesleyan University, United States*

[1]zsilver@iwu.edu, jplazak@iwu.edu

Abstract

In two related experiments pertaining to the theory of universal paralanguage, we aimed to test if manipulations of a psycho-mechanical sound property (i.e. sound source size) relative to a listener could predict emotional responses. Both studies utilized the TANDEM-STRAIGHT vocoder (Kawahara, 2006) to alter sounds recorded from each participant in order to produce new sounds with psycho-mechanical properties consistent with larger or smaller sound sources, relative to the listener. Our first study asked human participants about their perceptions of these sounds, and also measured their electrodermal activity in response to the stimuli. Our second study examined a cross-species response pattern within domestic canines, and utilized looking time duration as an operational measure for the perceptual salience of these timbral manipulations. Results from both studies found evidence consistent with the hypothesis that size-manipulations may have some predictive power for the perception of sound source size, but the results were not consistent with the hypothesis that these sound source size manipulations could explain induced emotional responses to musical stimuli.

Introduction

The theory of a universal paralanguage posits the existence of a shared "sound code" that is utilized within speech, music, and animal communication for the purposes of transmitting & receiving emotional information (Gussenhoven, 2001). Recent research suggests that the direct perception of psycho-mechanical properties of a sound source may contribute towards a better understanding of this theoretical universal paralanguage, and implicates musical timbre perception as an important aspect of the theory (van Dinther & Patterson, 2006; Plazak & McAdams, 2017). Crucially, through the direct perception of such timbral features, including a sound source's mass, size, materiality, and energy, it becomes theoretically possible to relate external sound sources to oneself (i.e. relating the source to the receiver), thus hypothetically facilitating the ability to engage appropriate emotional responses to music, and other types of sound (Plazak, 2016).

The ability to detect the source size of musical instruments has been documented in previous literature (van Dinther & Patterson, 2006; Chiasson, Traube, Lagarrigue, & McAdams, 2016), which provides support for the notion that listeners are not only sensitive to manipulations of musical sound source size, but are also capable of comparing two sounds based on this psycho-mechanical property. As an extension of this capacity, the question remains as to whether manipulations of sound source size can be perceived relative to the size of the listener. In a pilot study (Plazak & Silver, 2016), we reported a methodology for utilizing listener-normalized stimuli in order to specifically test this capacity, as well as to investigate the role of sound source size within induced emotional responses. The pilot study results were consistent with the hypothesis that listeners were capable of making sound source size judgments relatively, yet warranted further investigation to determine if induced emotional responses from sound source size manipulations could be measured via electrodermal activity (EDA) responses.

The mechanism through which listeners extract size-related information from auditory stimuli may also be shared by other mammalian species. If visual resources are scarce or unavailable (e.g. in poorly lit or densely packed areas), animals may rely on their ability to extract size related information from auditory cues in order to identify a possible threat such as large or threatening conspecifics. A variety of species (e.g., domestic dogs: Taylor, Reby, & McComb, 2010; red deer: Reby et al., 2005; koalas: Charlton et al., 2011; panda bears: Charlton, Zhihe, & Snyder, 2010; alligators: Reber et al., 2017) perceive the size of a conspecific based on either natural or synthetic vocalizations manipulated to give the impression of a larger or smaller conspecific. We utilized a convenience sample of domestic canines from an established dog research laboratory at our university for this aim. The method of stimulus presentation (discussed below) was similar to the paradigm used in our experiment with human listeners. Since it was not feasible to easily capture arousal levels in this population, we elected to measure perceptual salience via a looking-time paradigm.

Below, we report two separate studies related to the role of sound source size within the theory of a universal paralanguage, one using human participants, and another using domestic canines. In the first study, we aimed to find evidence that human listeners can correctly distinguish listener-normalized manipulations of sound source size, and more boldly, that these manipulations could explain common markers of induced emotion within electrodermal activity patterns. Specifically, we expected to find that the perception of relatively larger sound sources would induce larger arousal responses compared to stimuli perceived as relatively smaller, consistent with the idea that larger sound sources could be perceived as a threat to the listener. In our second study, we aimed to show that these same listener-normalized manipulations of sound source size could be distinguished within animal perception. Using a looking-time paradigm, we expected to find that stimuli manipulated to sound like a larger sound source would result in longer looking times, which might be indicative of more salient or higher priority information within animal communication.

Experiment #1

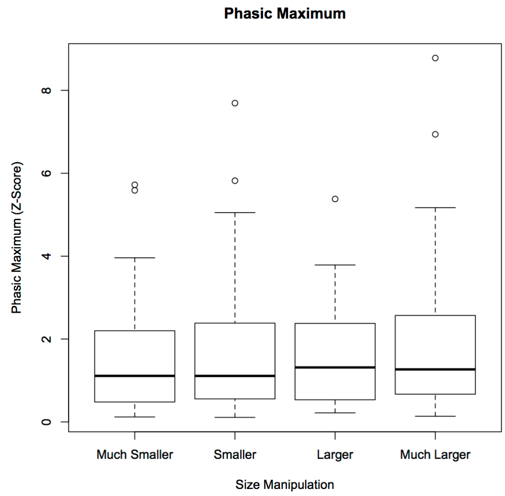
Method

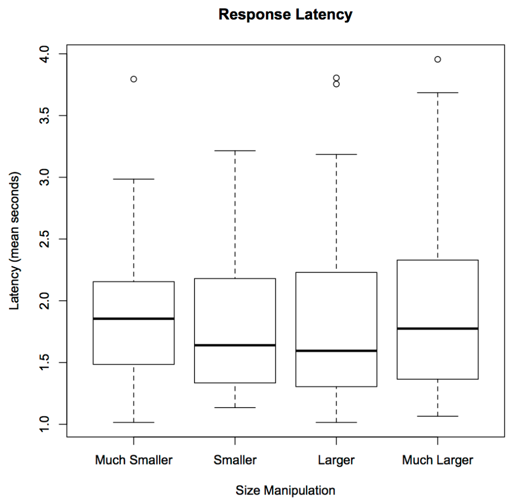
The methodology utilized in experiment 1 was identical to that reported in our 2016 pilot study, and thus, we provide only an abridged description of the methodology here. Twenty-four musician and non-musician participants from Illinois Wesleyan University were recruited via email for this study. At the beginning of each session, participants were asked to sing and record a sustained pitch (approximately 3 seconds long) near the middle portion of their vocal register, and thereafter the spectral envelope of the recording was re-synthesized using the Matlab-based TANDEM-STRAIGHT vocoder (Kawahara, 2006; Kawahara, Takahashi, Morise, & Banno, 2009) to one of four unique spectral envelope ratio (SER) manipulations: very small (SER x 0.7143), small (SER x 0.833), large (SER x 1.2) , and very large (SER x 1.4) stimuli, relative to the listener. Skin conductance measurements were collected via the same apparatus, and following the same paradigm, described in our 2016 pilot study (Plazak & Silver, 2016). Unlike our previous experiment, we used a small room heater to raise the ambient temperature of the room to a range conducive for recording EDA. At the end of each session, participants again listened to their manipulated sound recordings and were asked to determine if the sound sources that they heard were manipulated to sound relatively larger or smaller than their own size.

Results

We first investigated the perceptual responses provided by participants at the end of each session, in which they reported whether each sound could have originated from a larger or smaller sound source. In total, participants were 92.39% accurate (SD = 9.78%) in identifying whether the sounds they heard were modified to sound larger or smaller than themselves, which differed significantly from a chance level of 50% ($t(22) = 20.779$; 95% CI [88.16, 96.62]) and was consistent with the hypothesis that sound receivers can perceive sound source size information in relation to themselves.

Using the same method described in our 2016 study, we used Ledalab to analyze each EDA signal, with a particular interest in skin conductance response latency, response phasic maximum, and the response global mean. We ran separate within-subject ANOVAs for each of the three features of interest listed above, utilizing size manipulation (much smaller, smaller, larger, much larger) as the independent variable. In each case, we found a significant effect of subject, and no significant effects of size manipulation. Due to the significance differences between subjects, we elected to normalize each participant's data and express their signals as z-scores. Once all signals were normalized, we again repeated the Ledalab analysis process. While the normalization process did reduce some of the variability in our dataset, it did not meaningfully change our results. The repeated measures ANOVA found no main effects of size manipulation on EDA latency ($F[3, 23]=0.371$, $p=.77$), phasic maximum ($F[3, 23]=0.782$, $p=.51$), or the global mean ($F[3, 23]=0.411$, $p=.75$), although we did still find significant differences between subjects with regards to latency ($F[3, 23]=1.907$, $p<.05$) and phasic maximum ($F[3, 23]=4.254$, $p<.01$). These results are presented in figure 1a-c.





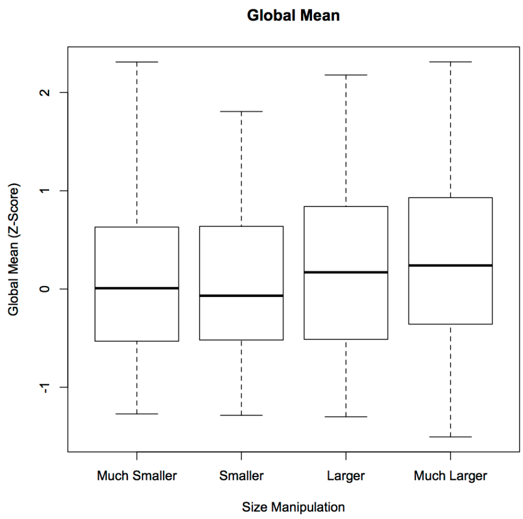


Figure 1 a-c. Mean values taken from normalized signals for EDA phasic maximum, EDA response latency, and EDA response global mean as a function of size manipulation category. All values were taken from a window between 1 - 4 seconds after stimulus onset.

Experiment #2

Method

Twenty-four domestic dogs were recruited via email (Mean age = 5.16 years, SD = 3.14; Mean weight = 52.19 pounds, SD = 28.33; 12 males) from a pre-existing database of dog owners.

In advance, dog owners submitted audio recordings of their dog's bark, and we extracted three individual barks from each recording and combined them into a loop with one second of silence between each bark. We then created synthetic barks using the TANDEM- STRAIGHT vocoder (Kawahara, 2006; Kawahara et al., 2009), through which we morphed the spectral envelope ratio of each dog's bark as to give the impression that the bark was produced by a larger or smaller dog. We used the same four spectral envelope ratio (SER) manipulations as experiment 1.

For testing, dogs were positioned approximately 5 meters away from two speakers that were 2 meters apart from each other. An experimenter was positioned behind a curtain out of the dog's view. With the dogs in the testing area, we played their synthesized barks back to the dog subject. The testing session unfolded as a continuous flow of experimental and habituation trials. Experimental trials consisted of three manipulated barks, which were all modified to the same size, played on a loop with one second of silence between barks. Habituation trials consisted of a loop of unaltered barking sounds between each of the experimental trials. An experimenter initiated playback of the subsequent loop once the dog subject looked away from the sound source for three consecutive barks. The order of the four size manipulations was randomized, and dogs heard each size manipulation twice for a total of 16 trials (8 experimental trials and 8 habituation trails). Behavioral observations were recorded via Sony Handycam HDR-CX220 camcorders, and analyzed at 30 frames per second using the open source software MPEG Steamclip. We recorded the duration of time that the dog looked at the speaker from which the sound originated. Testing sessions lasted between 5-13 minutes.

Results

Across all dogs, mean looking times for all trials combined was 11.78 seconds (SD = 12.47 seconds; 95% CI [6.79, 16.77]). A repeated measures ANOVA revealed that dogs spent significantly different amounts of time looking at sounds corresponding to larger dogs, smaller dogs, and habituation sounds ($F[2, 22] = 4.724$, $p = 0.020$) such that dogs looked at the source of the sounds longer in response to sounds manipulated to sound like a larger dog (M = 17.35, SD = 21.03, 95% CI [8.93, 25.76]) compared to sounds manipulated to sound like a dog smaller than the subject (M = 11.69, SD = 14.41; 95% CI [5.92, 17.46]; $p = 0.053$). Compared to the habituation sounds (M = 6.30, SD = 4.89; 95% CI [4.34, 8.26], dogs also looked at both the larger and smaller sounds for a longer duration of time (relatively larger: $p = 0.005$; relatively smaller: $p = 0.032$).

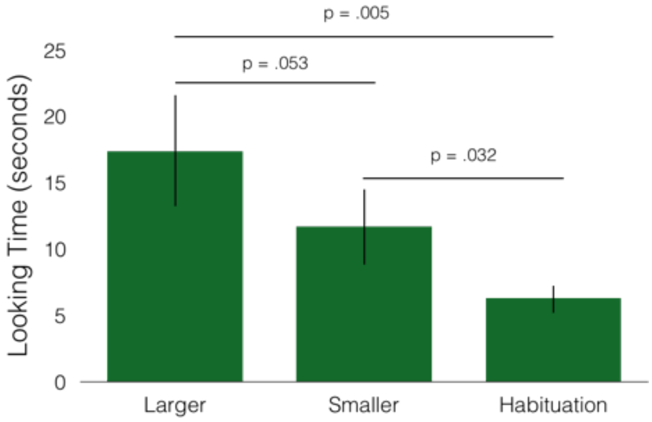


Figure 2: Mean looking times and standard deviations for each of the three stimuli types. Dogs looked in response to larger sounds for a longer duration of time compared to smaller sounds (p = 0.053) and habituation sounds (p = 0.005). Dogs also looked longer in response to smaller sounds longer compared to habituation sounds (p = 0.032).

General Discussion

Consistent with existing literature, our results provided evidence that listeners are sensitive to manipulations of sound source size. We furthered the existing literature by demonstrating that human listeners were able to perceive these sounds relative to their own size and were highly accurate in determining if a size manipulation resulted in a sound that was larger or smaller than themselves. Further, we found similar results in a cross-species experiment, with domestic canines exhibiting marginally significant differences between sounds manipulated to sound relatively smaller and relatively larger than themselves. However, we have no way of knowing if the dog participants actually perceived these sounds in a manner that is similar to human listeners.

Our collection of EDA data was motivated by the hypothesis that listener-normalized stimuli could explain individual differences to induced emotional responses from sound stimuli, but the resultant data was not consistent with this hypothesis. Instead, we found significant differences between subjects when analysing raw EDA scores, and similar between subject differences after normalizing the data. The amount of variance prevented us from drawing any conclusions regarding the relationship between listener normalization and induced emotional responses, and we report these null results with the hope that related research may investigate other metrics of induced emotion (pupil dilation, brain activity patterns, etc.) besides electrodermal activity.

Along with previous research, these results were consistent with the hypothesis that the ability to perceive the size of a conspecific through sound cues alone occurs across species, and further, that responses to larger and more energetic sound sources may be prioritized across species. Ultimately, these cross-species findings may guide us towards a better understanding of induced emotional responses to music, particularly with regards to how emotional responses resulting from various orchestrations (or in the case of electronic music, various types of sound modelling) can result in reliable patterns of induced emotional responses for some listeners.

References

Charlton, B. D., Ellis, W. A. H., McKinnon, A. J., Cowin, G. J., Brumm, J., Nilsson, K., & Fitch, W. T. (2011). Cues to body size in the formant spacing of male koala (phascolarctos cinereus) bellows: Honesty in an exaggerated trait. *Journal of Experimental Biology, 214*(20), 3414-3422. doi:10.1242/jeb.061358

Charlton, B. D., Zhihe, Z., & Snyder, R. J. (2010). Giant pandas perceive and attend to formant frequency variation in male bleats. *Animal Behaviour, 79*(6), 1221-1227. doi:10.1016/j.anbehav.2010.02.018

Chiasson, F., Traube, C., Lagarrigue, C., & McAdams, S. (2017). Koechlin's volume: Perception of sound extensity among instrument timbres from different families. *Musicae Scientiae, 21,* 113–131. http://dx.doi .org/10.1177/1029864916649638

Gussenhoven, C. (2001). Intonation and biology. In Liber Amicorum Bernard Bichakjian (pp. 59–82). Maastricht, the Netherlands: Shaker.

Kawahara, H. (2006). STRAIGHT, Exploration of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoustical Science and Technology, 27,* 349–353. http://dx.doi.org/10 .1250/ast.27.349

Kawahara, H., Takahashi, T., Morise, M., & Banno, H. (2009). Development of exploratory research tools based on TANDEM-STRAIGHT. In *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference* (pp. 111– 120). Sapporo, Japan: Hokkaido University.

Plazak, J., & McAdams, S. (2017). Perceiving changes of sound-source size within musical tone pairs. *Psychomusicology: Music, Mind, and Brain, 27*(1), 1.

Plazak, J. (2016). Listener-normalized musical affect: conceptualizing "felt" musical emotions. In V. Johnson, D. Zhao, H. Conrad (editors), *"Proceedings of the 14th International Conference for Music Perception and Cognition."* pp. 304-306.

Plazak, J. & Silver, Z. (2016). Inducing emotion through size-related timbre manipulations: A pilot investigation. In V. Johnson, D. Zhao, H. Conrad (editors), *"Proceedings of the 14th International Conference for Music Perception and Cognition."* pp. 276-278.

Reber, S. A., Janisch, J., Torregrosa, K., Darlington, J., Vliet, K. A., & Fitch, W. T. (2017). Formants provide honest acoustic cues to body size in american alligators. *Scientific Reports, 7*(1). doi:10.1038/s41598-017-01948-1

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society B: Biological Sciences, 272*(1566), 941-947. doi:10.1098/rspb.2004.2954

Taylor, A. M., Reby, D., & McComb, K. (2010). Size communication in domestic dog, canis familiaris, growls. *Animal Behaviour, 79*(1), 205-210. doi:10.1016/j.anbehav.2009.10.030

van Dinther, R., & Patterson, R. D. (2006). Perception of acoustic scale and size in musical instrument sounds. *The Journal of the Acoustical Society of America, 120*, 2158–2176. http://dx.doi.org/10.1121/1.2338295

Acknowledgments

The authors would like to extend their sincere gratitude to Ellen Furlong (Department of Psychology, Illinois Wesleyan University) for her assistance in the collection and analysis of the canine data.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Modeling Rhythmic Complexity in a Corpus of Polyrhythm Examples from Europe and America, 1900-1950

Ève Poudrier[1] and Daniel Shanahan[2]

[1] *School of Music, University of British Columbia, Canada*
[2] *School of Music, Ohio State University, United States*

[1]eve.poudrier@ubc.ca, [2]daniel.shanahan@gmail.com

Abstract

Rhythmic complexity, as represented by *polyrhythm* (the superposition of two or more contrasting rhythms, meters or speeds), is often identified as one of the central features of twentieth-century Western art music. This project uses computational analysis to explore the development of polyrhythm in a corpus of 719 examples extracted from 450 works by composers from Europe and North America from 1900 to 1950 (Suter, 1980). The current study aims to develop complexity metrics to examine the use of polyrhythm in this period, test competing claims about its development, and explore the cognitive processing of complex rhythms. Corpus examples and associated metadata were processed to be analyzed using the *Humdrum Toolkit* (Huron, 1995). Exploratory analysis was conducted using a stratified sample dataset (N = 80) that includes four randomly selected examples for each composer (N = 20). Correlational analysis using global complexity measures (average nPVI, event density, entropy) showed that differences in entropy and variability between rhythmic groups within each excerpt could be predictive of genre, and that the entropy of the composite rhythms seems to decline in the 1930s and 1940s, before increasing again in the 1950s. These results are taken as starting points for future avenues of research on the interaction of rhythmic groups, variability, and complexity in the use of polyrhythm in Western art music of the twentieth century.

Introduction

While the cognitive mechanisms that support entrainment to periodic auditory signals have become increasingly well-defined (e.g., Large & Jones, 1999), little research has focused on how musical agents (which may be identified as belonging to one or more of the three categories traditionally described as listeners, composers, and performers), in various cultures, actually perceive and aesthetically evaluate complex musical rhythms. Such rhythmic complexity is often identified as one of the central features of twentieth-century Western art music. In particular, the use of *polyrhythm*, that is, the superposition of two or more contrasting rhythms, meters or speeds, seems to have increased significantly both in prevalence and scale. Before the turn of the twentieth century, apart from a few isolated cases (e.g., Mozart's simultaneous use of three different orchestras, each playing in a different meter in a scene from *Don Giovanni*, 1787), polyrhythms were generally used locally or as a means to create special accompaniment textures (e.g., Chopin's Etude No. 1 from *Trois Nouvelles Études*, 1840).

Several competing ideas have been offered to explain this stylistic development. Composers' growing interest in experimentation, including the representation of multiple temporal perspectives simultaneously, has been interpreted as stemming from a desire to express more intense human experiences brought about by urbanization and social unrest. Alternatively, the resulting musical structures have been characterized as mechanistic, reflecting the rapid development of industrialization and technology. The bulk of evidence presented in support of these claims is mostly anecdotal in nature, often taking a composer's statement as a point of departure and mapping it loosely onto socio-cultural trends or historical events. Yet another approach, common in music-theoretical research, has been to use close analysis to generate a set of observations, which are then formalized based on some guiding principles borrowed from a variety of disciplines such as acoustics, mathematics, linguistics, or phenomenology. Common to these approaches is the reliance on close reading, subjective interpretation, and generalization from a small number of examples believed to be representative (e.g., Krebs, 1999; Poudrier, 2009; Roeder, 1994), without the possibility to test the validity of the claims, that is, to measure the probability that the findings are not due to chance or researchers' bias.

The current study is part of a larger project that seeks to develop a framework for the computational analysis of polyrhythm, and to explore the development of polyrhythmic techniques in a corpus of 719 musical examples extracted from 450 works by composers from Europe and North America in the first half of the twentieth century, when the use of these techniques has been noted to increase significantly. The primary research objectives of this project are to: (1) identify trends in the development of polyrhythm over the focus period within the context of musical production (composer's national origin, date of composition, and location of first public performance as well as performers involved); (2) identify structural features and associated musical parameters, such as timbre, register, dynamics, and performance manner (e.g., staccato vs. legato); and (3) explore psychometrics associated with polyrhythm perception. The current study focuses on the first objective and aims to develop complexity metrics to examine the use of polyrhythm in this corpus, test competing claims about its development, and explore the cognitive processing of complex rhythms.

The Suter (1980) Corpus

The source corpus was compiled in print format by Louis-Marc Suter (1980). Suter's study aimed to identify trends in stylistic development as well as to classify the specific structures and effects resulting from the use of polyrhythmic techniques in the works of twenty representative composers in the period of 1900 to 1950, from Leoš Janáček (1854-1928) to Benjamin Britten (1913-1976). While Suter's study provides evidence for an increase in the use of polyrhythm, both in terms of prevalence and complexity, the large number of works surveyed and the methodology he adopted (close reading of a sub-set of examples) place important limitations on the reliability and specificity of his findings, and thus, many of his claims must be regarded as speculative.

So far, computational analysis of temporal structures has been limited to rhythmic patterns that are integrated in a single metric hierarchy, including the disruption of an established meter (Huron & Ommen, 2006; Temperley, 1999; Volk, 2008). Suter's sampling of polyrhythm examples in 913 representative works presents a unique opportunity to expand the scope of rhythm research, and to test the validity of claims about the development of polyrhythm in this specific style and period. The encoding of this corpus will also facilitate the design of ecologically valid stimuli for listening experiments aimed at exploring the psychophysical mechanisms and cognitive constraints at work in the processing of complex rhythmic structures, and could provide a basis for comparative studies of polyrhythm production and perception in different cultures (e.g., such as those found in African diaspora music).

Methods

Sampling of source materials. Suter's original corpus is comprised of 913 works by twenty different composers from Europe and North America (see Figure 1), ranging from Leoš Janáček (1854-1928) to Benjamin Britten (1913-1976). To allow for generalization across the various styles and genres represented in the focus period, Suter selected these twenty composers based primarily on the timing of their output (i.e., majority of works composed between 1900 and 1950) and even distribution in terms of composers' age during the focus period.

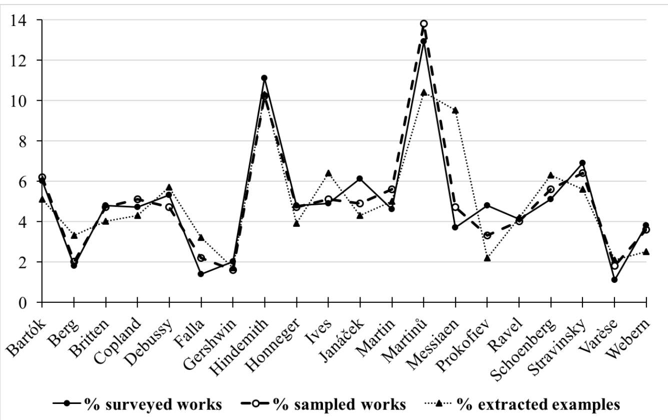


Figure 1. Composer percentage representation in the Suter (1980) Corpus based on surveyed works, sampled works, and extracted examples. A total of 913 of works were surveyed; 719 examples were extracted from 450 of these works. Composers' average representation within type is 5% (*SD* = 2.6 to 2.9), with works and examples by Martinů (10.4 to 13.8%) and Hindemith (10.2 to 11.1%) being the most represented, and those by Varèse (1.1 to 2.1%), Gershwin (1.6 to 2%), and Falla (1.4 to 3.2%) being the least represented.

Among the group of available composers that satisfied these basic criteria, specific composers were selected based on their recognized influence on the music of the period ("individual character") and the diversity of genres represented in their output ("stylistic diversity") rather than on the prevalence of polyrhythm in their works. There was no attempt at equal representation in terms of national origin. To allow for characterization, surveyed works were then selected to provide a representative sample of the full range of a given composer's creative output, both in terms of style and genre, rather than proportional representation across composers or genres.

These selection criteria resulted in an unbalanced dataset in terms of national origin (e.g., there are four French composers, but no Italian composer), genre (larger proportion of orchestral and instrumental works in comparison to vocal or stage works), and composer representation (number of works/examples per composer, as shown in Figure 1). Examples are also quite varied in terms of length, ranging from 1 to 44 notated bars (M = 6, SD = 4), while source work full units (e.g., movement or scene from which an example was extracted) range from 9 to 1,044 notated bars (M = 231, SD = 192), with extracted examples representing an average of 5.5% of the source work full units (SD = 9.2). The current corpus is comprised of the examples for which score excerpts were included in Suter's dissertation; these were extracted from 450 works, with 151 of these works being represented by two or more extracted examples.

Encoding of examples. A *reserve dataset* (RDS) was created using a stratified sample of the full corpus (designated as FDS) to allow for both exploratory analysis and hypothesis testing. The RDS is comprised of four randomly selected examples for each composer (N = 80), the remaining 639 examples forming the *testing dataset* (TDS). Examples from Suter (1980) that featured non-adjacent segments were split into separate examples resulting in a total number of 719 examples for the current corpus. The encoding of the corpus involved transcribing the score excerpts using Sibelius so that they could be converted in kern representation using Humdrum. For works that featured texted parts, all musical data were transcribed except for the text, and to facilitate score reading, transposing instruments were transcribed to sounded pitch (in C). Transcribed examples were then proofread and edited manually as needed, including correction of translation errors. To date, all of the RDS examples have been encoded, and 101 examples from the TDS are still in process, including eight examples that could not be transcribed due to software limitations and some examples for which a published score has not been located. It has not been determined yet whether these examples can be manually encoded, and whether they will be excluded from the corpus or will be retained for the purpose of metadata analysis.

Metadata gathering. For the purpose of exploring correlations between structural features and aspects of musical production, metadata related to the twenty represented composers as well as to each example's source work and first public performance were collected and linked to the primary data (see Table 1). The primary sources for metadata collection consisted of in-print as well as online authoritative scholarly sources such as scholarly editions of a composer's complete works, Oxford Music Online, and biographies as well as other easily accessible resources such as composer-dedicated websites, IMSLP, CD booklets from Naxos Music Library, and Wikipedia. Metadata collected from non-scholarly sources were subjected to a data validation process requiring confirmation by at least one additional source. Additional metadata pertaining to each example's temporal attributes were collected using the most easily accessible printed edition of each work. Recorded duration was collected using randomly selected recordings from the Naxos Music Library (one for each example). To date, metadata gathering for the full dataset (FDS) is at 93.6% average completion, with most fields having reached above 90% completion, except for premiere organization (46.7%), premiere venue (67.2%), and premiere ensemble (84.3%).

Table 1. Corpus metadata based on composers, works, premieres, and examples. Sub-type categories are shown in parentheses.

Composers	• Birth/death place (city, country) • Birth/death year • Nationality
Works	• Genre (orchestral, instrumental, vocal, stage) • Sub-genre (symphonic, concerto, programmatic, keyboard, chamber, song, choral, opera, ballet, theatrical) • Composition year (first, last) • First publication year
Premieres	• First public performance year • Location (organization, venue, city, country) • Performers (conductor, ensemble, soloists)
Examples	• Source work full unit[a] • Notated length • Recorded duration • Time signature • Tempo (expression, notated beat, metronomic rate)

[a]For multi-movement works (and stage works), this corresponds to the specific movement (or scene) where the excerpt is located.

Mapping the corpus. To facilitate access to the corpus metadata and visualization of the corpus works in their spatio-temporal context, each example was geocoded based on the location of the source work full unit's first public performance using ArcGIS and Carto. Whenever possible, the specific venue's geographical coordinates (latitude and longitude) were used; if venue could not be determined, the city's coordinates were used. By clicking on the premiere location, the user is able to access a portion of the metadata associated with each example as well as a score representation of the example and the randomly pre-selected recording of the source work full unit. Works' premiere locations can also be visualized in time based on a specific composer or genre or a sub-set of composers and genres. A sample map of the reserve data set with timeline animation is available here: https://epoudrier.carto.com/builder/adfbe0b9-093c-4eb7-b3ef-bb2b4b05a99f/embed.

Overall, it is worth noting that the most common premiere location for the full corpus is Paris, France (21.1% representation), with New York being the second most common (12.5%). On the timeline representation, there is a noticeable shift from Paris to New York when comparing the premiere locations from before to after the second world war, which may be related to several of the composers represented in the corpus having emigrated to the United States (e.g., Paul Hindemith and Arnold Schoenberg) at that time.

Corpus Characteristics

While the use of a corpus of examples collected by a third party shifts the sampling bias away from the researchers, generalization of findings will need to take into account the specific characteristics of the corpus. The following sub-sections present some descriptive statistics pertaining to the examples in the full corpus ($N = 719$).

Timeline. While most of the corpus examples were composed, premiered, and published in the time period from 1900 to 1950, there are a number of examples that fall outside of this period, with 22 works having been first completed before 1900 and 69 works having been first completed from 1950 onward. Thus, the actual period represented by the three phases in the creation of these musical works corresponds to a time range of 1877 to 2009 (see Figure 2).

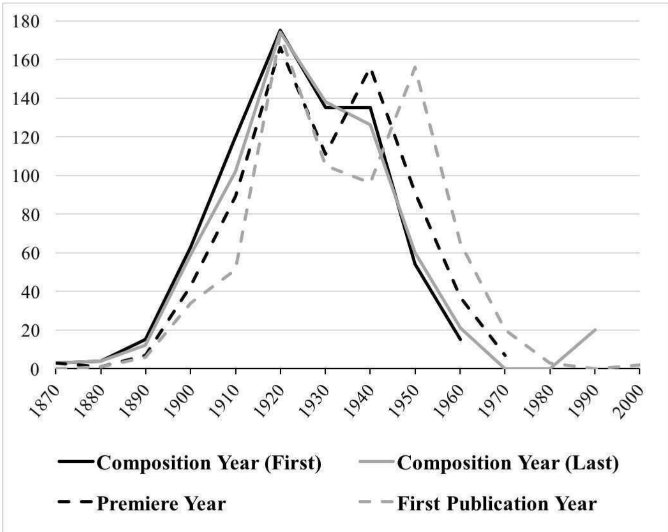


Figure 2. Distribution of corpus examples by decade based on composition year, premiere year, and first publication year. To account for composers' revisions of source works, metadata pertaining to composition year includes both first completed year and last completed year. Note that the spike in last completed composition year is represented by a single work, Olivier Messiaen's *Turangalîla Symphony*, from which 20 examples were extracted.

In general terms, the average first completed composition year represented by the corpus is 1929 ($SD = 16$) and the last completed composition year, which accounts for composers' revisions of the source works, is 1932 (SD = 18.6); average year for premiere and first publication are only slightly later, 1934 ($SD = 17$) and 1939 ($SD = 18.5$), respectively. However, while composition ($N = 719$), premiere ($N = 711$), and first publication ($N = 712$) followed each other closely, there appears to be a noticeable delay between premiere and first publication in the 1940s. Of special interest in this corpus is the potential disruptive effect of the two world wars (1914-1918) on the process of musical creation, from composition to premiere and publication, and this observation suggests that it may be a fruitful avenue of research.

Genre and sub-genre. The corpus source works were categorized into four genres based on the type of ensemble for which they were written: orchestral, instrumental, vocal, and stage works. Overall, orchestral and instrumental works account for 38% and 36% of all source works, respectively, while both vocal and stage works account for only 13%. Each of the four genres was further characterized based on style and specific instrumentation (see Figure 3). When taking sub-genre into consideration, the most represented works are for chamber ensemble (N = 175), accounting for 24.3% of the corpus works (and extracted examples), and the least represented are ballets and theatrical works (e.g., Arthur Honegger's *Jeanne d'Arc au bûcher*, a dramatic oratorio for spoken roles, soloists, children's choir, mixed chorus, and orchestra, which was composed in 1935 and premiered in Basel, Switzerland in 1938).

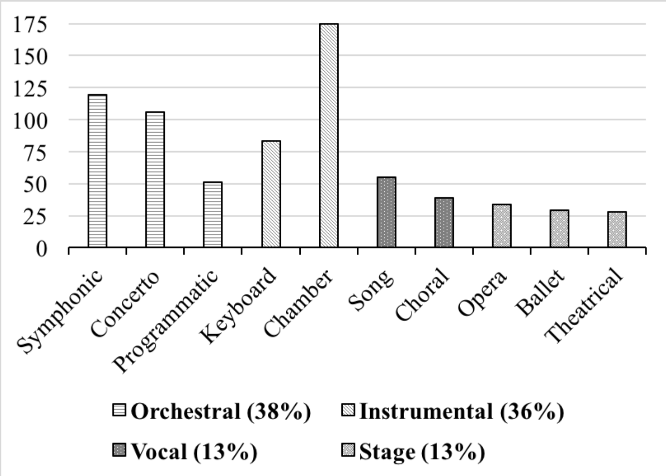


Figure 3. Genre and sub-genre representation. Percentage representation for each of the four genres is given, and sub-genre distribution is illustrated by the corresponding number of works in the full corpus.

Meter and tempo. Metadata pertaining to start time signature and tempo for both source work full units and extracted examples were collected. Overall, the most common start time signatures used are 4/4 and 3/4, which together account for 45.9% and 38.1% of source works full units and extracted examples, respectively. Given the wide range of time signatures used, these were further categorized based on the specific number of beats in the notated bar (duple, triple, quadruple, and odd, which feature uneven subdivision of the notated bar) or the relationship between notated beats and subdivision units (e.g., compound and polymetric); the type "none" refers to works in which rhythmic duration is notated without a reference to a time signature (see Figure 4).

In addition to quadruple and triple time signatures, duple time signatures (especially 2/2 and 2/4) are also relatively common, accounting for 19.3% and 22.2% of source work full units and extracted examples, respectively. In contrast, odd and polymetric time signatures are the least frequently used, respectively accounting for only 3.4% and 2.6% of start time signatures for source work full units (N = 24 and 18), and for 5.8% and 6.9% of start time signature for extracted examples (N = 40 and 48). However, while start time signature representation for source work full units and extracted examples are fairly similar overall, quadruple time signatures (which include 4/2, 4/4, 4/8, and 4/16 as well as 8/8 and 8/4) and polymetric time signatures (e.g., 4/2 and 3/4 superposed) are somewhat more frequent at the beginning of the extracted examples than at the beginning of the source work full units. The relatively higher prevalence of polymetric time signatures at the beginning of the extracted examples is consistent with the nature of the extracted examples, all of which feature some form of polyrhythm, but the increased representation of quadruple time signatures has no theoretical basis and is worth exploring further. Finally, one aspect of the extracted examples that is not represented here is that 14% of these excerpts feature changing time signatures.

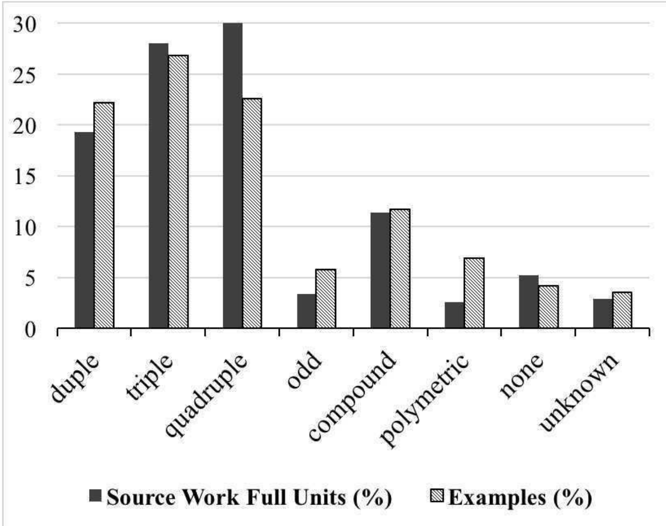


Figure 4. Percentage representation of start time signatures for source work full units and extracted examples based on metric type. Start time signatures could not be identified for 20 source work full units and 24 extracted examples, due to difficulties in locating a published score.

Start tempo metadata (expression, notated beat, and metronomic rate) were also collected. Expression markings (e.g., Allegro) are featured at the beginning of 93% of source work full units (based on metadata gathering at 97.2% completion), while specific metronomic rates for the notated beats are featured at the beginning of only 73% of the source work full units (and only 70% of the extracted examples, based on metadata gathering at 96% completion). When provided, the average notated start tempo is 95 bpm for the source work full units (SD = 34.2), while that of the extracted examples is 103 bpm (SD = 37.0), a difference that calls for further analysis. Because of the limited availability of metronomic rates as well as the difficulty involved in computing tempo based on the notated examples alone, the recorded duration of each example was also collected using randomly selected recordings from the Naxos Music Library.

Measuring Complexity

Using our reserve dataset, we performed a number of exploratory tests focusing on the interactions of tempo, genre, composer, year, and various aspects of rhythmic complexity, such as entropy and rhythmic variability. Specifically, we examined the composite rhythms of each excerpt, both as a single (global) feature and as a pair of interacting rhythmic groups (e.g. "3 against 2"). Three research assistants were tasked with placing parts (N = 3 to 30) into one of two groups;

examples that featured only two parts were automatically divided into two groups, and there was one example that featured a single part, which was excluded from the procedure. Research assistants were presented with the score and asked to assign each part to one of two groups based on how rhythmically and metrically similar the parts were to each other, with the parts that were least similar to each other having to be placed in different groups. These coders (all music graduate students) were largely in agreement. When two of the three graduate students were in agreement, the majority opinion decided the separation. In the rare instances where there was no agreement, we examined the score and made a judgment about the appropriate grouping. All of these cases were due to duplicate or missing parts. The analyses presented here therefore examine the single composite rhythm of excerpts as well as the interaction of two composite rhythms, each of which results from one of the two contrasting rhythmic groups within each excerpt.

Measurements

As a starting point, we decided to examine aspects of rhythmic entropy and variance. *Entropy* (a general measure of uncertainty and variability) was calculated as the Shannon entropy of the composite rhythm of the entire excerpt (the global composite rhythm), as calculated with the *Humdrum Toolkit* (Huron, 1995; Sapp, 2005; specifically, the *infot* and *beat* tools). This allows for a measurement of some degree of "unexpectedness" in the rhythm. Similarly, we calculated the *average variability* of each instrument, as calculated with the normalized pairwise variability index (see Daniele, 2017; Daniele and Patel, 2003; Grabe and Low, 2002). Rather than examining a composite rhythm, this measurement examines each rhythm individually, providing an average metric for variability.

Additionally, we examined the event density of each musical excerpt (as notated in the score), as well as the density in "clock time." *Event density* was calculated simply as the number of events per notated measure, whereas *clock time density* was measured on the performed time extracted from a randomly selected recording. Tempo was also calculated based on the recorded performance following these steps: the number of beats per measure were multiplied by the number of measures in the notated example, and the resultant number was divided by the time elapsed in the recording. For example, if 4 measures of 4/4 were performed in 5 seconds, the tempo was calculated as (4*4)60/5, or 16*12, resulting in an extrapolated tempo of 192 bpm.

Finally, in addition to examining these metrics in isolation, we also examined the difference between rhythmic groups, as represented by their composite rhythms.

Change over time. With the reserve dataset, it seems that there is no change in any of the aspects we examined over time, but a number of relationships point to possible avenues for future research. There was also no significant change in tempo, nPVI, entropy, or the difference in entropy and nPVI between rhythmic groups. Figure 5 indicates a gradual slowing of tempo over time in our reserve dataset, in keeping with much of the research on performance practice during this period (see Cook, 2013; Philip, 1992). These tempi, however, were deduced from contemporary performances of the pieces, meaning that rather than reflecting performance practice, they reflect a difference in contemporary approach from pieces composed in different periods.

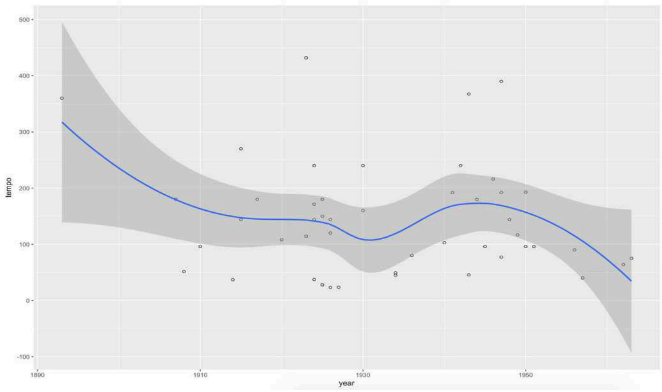


Figure 5. Performance tempi plotted against first composition year.

Interestingly, there seem to be a number of measurements in which there is a mid-century dip, including the entropy of the composite rhythm entropy (see Figure 6), and the difference in entropy between the two rhythmic groups. The time period over which decreasing entropy can be observed seems to correlate to the period during which a delay between premiere and first publication year was previously observed. As noted earlier, as this period also corresponds with the emergence of the second world war, this is a result that calls for closer examination.

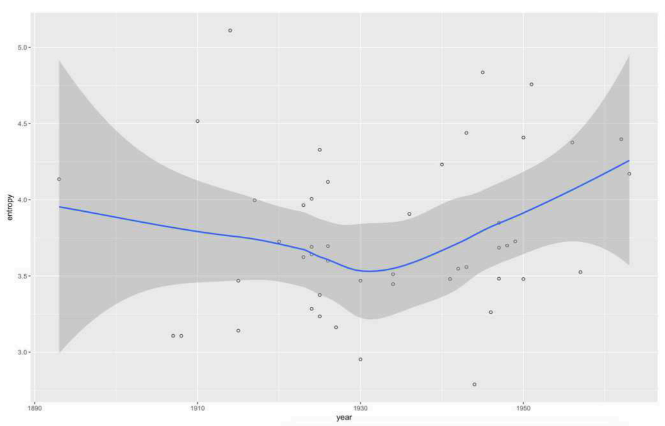


Figure 6. Entropy of the composite rhythm over time.

Predicting genre. The interaction between rhythmic characteristics and genre is also quite interesting. Figure 7 presents the variable importance of certain characteristics when used to train a generalized linear model classifier (measured here with the absolute value of the *t*-statistic of each feature) to predict genre. Although the classifier itself performed significantly better than chance (71% accuracy), it would likely do much better given more data. These results suggest that the difference in entropy between rhythmic groups plays a large role when predicting vocal works, whereas the difference in nPVI between groups is possibly more predictive of instrumental works. Further analysis is needed to pick these points apart, as the two measurements likely contain a certain amount of covariance. Finally, tempo and event count also appear to be important in predicting orchestral works, which similarly warrants further inspection.

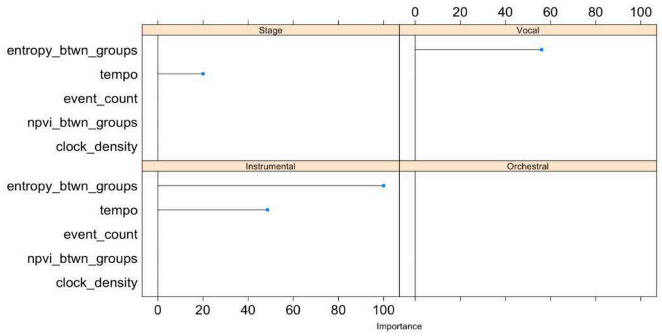


Figure 7. Log Odds Ratios resulting from a mixed effects logistic regression model predicting metric type. None of the features were significant predictor.

Conclusion

So far, claims about a significant increase in rhythmic complexity over the 50 years represented by the Suter (1980) corpus have not been substantiated, at least from the perspective of global measures of variability (entropy and nPVI) and density applied to a stratified sample (reserve dataset) of examples that feature a wide range of polyrhythmic structures. This may be understood to attest to a certain stylistic homogeneity in the music of this period, although this observation remains to be tested. Nonetheless, some trends based on national origin and genre have been identified, and the wide variance observed in measures of rhythmic regularity based on composer, nationality, genre, and pre- vs. post-war composition year suggests that there are other factors at play, which future analytical work will address.

Although there were no significant differences observed, it would seem that there are a number of avenues in which further research would be warranted. Firstly, polyrhythmic gestures change over time, but not necessarily in a linear way. Our preliminary results suggest that it is likely that the middle of the twentieth century differs from earlier and later periods in terms of polyrhythmic practice. Secondly, it would seem that the types of features used in these polyrhythmic compositions might be genre-dependent. Future work would ideally explore to what extent this might be the case, what other features are involved, and how the changing nature of complex rhythms might inform our understanding of twentieth-century music.

Acknowledgements. This project was made possible with funding support from the Whitney Humanities Center at Yale University and the Hampton Research Endowment Fund at the University of British Columbia, as well as by the efforts of many research assistants, including Annelies Andries, David Baker, Austin Culler, Connor Davis, Natalie Dietterich, Marina Gallagher, Carl Gardner, Stephen Guerra, Felicia Hartono, Michael Kirchmayer, Melissa Liang, Amita Mahey, Sarah Pratt, Hailey Wu, and Jasper Yoo.

References

Cook, N. (2013). *Beyond the score: Music as performance*. New York: Oxford University Press.

Daniele, J. R. (2017). A tool for the quantitative anthropology of music: Use of the nPVI equation to analyze rhythmic variability within long-term historical patterns in music. *Empirical Musicology Review*: EMR. Retrieved from http://emusicology.org/article/view/4893

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology, 7*, 515-546.

Huron, D. (1995). *The Humdrum Toolkit: Reference manual*. Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Huron, D. & Ommen, A. (2006). An empirical study of syncopation in American popular music, 1890-1939." *Music Theory Spectrum, 28*(2), 211-232.

Krebs, H. (1999). *Fantasy pieces: Metrical dissonance in the music of Robert Schumann*. New York, NY: Oxford University Press.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review 106*(1), 119-159.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87*. https://doi.org/10.1016/S0

Philip, R. (2004). *Early recordings and musical style: Changing tastes in instrumental performance, 1900-1950*. Cambridge University Press.

Poudrier, È. (2009). Local polymetric structures in Elliott Carter's *90+* for piano (1994). In B. Heile (ed.), *Modernist legacy: Essays on new music* (pp. 205-233). Farnham, England: Ashgate.

Roeder, J. (1994). Interacting pulse streams in Schoenberg's atonal polyphony. *Music Theory Spectrum, 16*(2), 231-249.

Sapp, C. S. (2005). Online Database of Scores in the Humdrum File Format. In *Proceedings of the International Society for Music Information Retrieval* (pp. 664–665).

Suter, L-M. (1980). Les polyrythmes dans la musique de vingt compositeurs de la première moitié du vingtième siècle, 1900-1950 [Polyrhythm in the music of twenty composers from the first half of the twentieth century, 1900-1950]. (Unpublished doctoral dissertation). University of Bern, Bern, Switzerland.

Temperley, D. (2004). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

Volk, A. (2008). The study of syncopation using Inner Metric Analysis: Linking theoretical and experimental analysis of metre in music. *Journal of New Music Research, 37*(4), 259-273.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Investigating Musical Pattern Ambiguity in a Human Annotated Dataset

Iris Yuping Ren,[1] Oriol Nieto,[2] Hendrik Vincent Koops,[1] Anja Volk,[1] and Wouter Swierstra[1]

[1] *Department of Information and Computing Sciences, Utrecht University, Netherlands*
[2] *Pandora Media, Inc., Oakland, CA, USA*

y.ren@uu.nl

Abstract

Many musical structures, such as musical motifs or patterns, are inherently ambiguous and lead to different equally plausible interpretations by listeners. Accordingly, annotations of these musical structures by different listeners give rise to disagreement between different annotators. In Music Information Retrieval (MIR) tasks such as automatic musical pattern extraction, however, this disagreement poses particular difficulties for evaluation. To provide data for further research on the ambiguity of musical patterns, we present a new annotated patterns dataset. This collection comprises six musical excerpts each annotated by 12 annotators. We observe from the data that disagreement amongst annotators is common. We therefore propose to perform two treatments on annotations to achieve higher pairwise annotator agreement: by using extra rankable metadata on the annotations such as relevance/importance scores, and by adjusting the time resolution of the annotated patterns' time span. We hypothesise that, by using the top-ranked annotations and lowering the time resolution of the annotations, we may obtain more pairwise annotator agreement in the dataset. We perform computational analyses and provide supporting evidence that patterns rated as highly relevant or/and with lower time resolution tend to have more agreement amongst the annotators, in contrast to those rated with lower relevance and higher time resolution. Our analyses could be useful for the development and evaluation of new musical pattern discovery algorithms.

Introduction

In Music Information Retrieval (MIR), automatic musical pattern discovery is an active area of research where algorithms are designed, employed on music data, and evaluated to extract musical patterns computationally for different applications (Collins, 2011). However, ambiguity of musical structures poses difficulties on how one should evaluate the output of such algorithms: if different equally valid interpretations of musical patterns or motifs[1] are possible, what should we take as the ground truth for evaluating algorithms?

Music ambiguity per se is a widely studied topic in music perception and cognition research. Not only musical patterns, but also other dimensions of music can be ambiguous in their interpretation: polyrhythms, for example, can be heard and interpreted with beats at different levels; chords of a harmony sequence can be interpreted with different functions (Randall, 1999, Schoenberg, 1983). Furthermore, creators of music often employ elements of ambiguity to their compositions, and listeners often experience uncertainty where multiple simultaneous interpretations are possible (Bernstein, 1976).

In this paper, we put focus on the ambiguity of monophonic melodic patterns. While musicologists use the term "motif" often intuitively in their analyses of musical compositions, there exists no generally agreed upon definition of what constitutes a motif or pattern: it can be a short musical idea, a salient recurring figure, musical fragment, or succession of notes that has special importance in or is characteristic of a composition. Related concepts also include musical sequence, imitation (Benward, 2014), melody type (Hiley, 1993), musical cell (Nattiez, 1990), phrase (Burkhart, 2005; Sadie, & Tyrrell, 2001), and subject (Scholes, 1970). These different notions of what might constitute a musical pattern pose challenges for creating annotations as reference data for evaluating pattern discovery algorithms.

One attempt to systematically evaluate pattern discovery algorithms is the Music Information Retrieval Evaluation eXchange (MIREX) Discovery of Repeated Themes & Sections task (Collins, Janssen, Ren & Volk, 2017). This task uses a single reference annotation compiled from music theoretic analyses. In the task, a pattern is defined as a set of time-pitch pairs that occurs at least twice in a piece of music. However, in (Ren, Koops, Volk, & Swierstra, 2017) it is shown that pattern discovery algorithms do not agree with each other on what patterns should be extracted from the pieces, and they agree even less with the patterns from the reference annotation. This raises the question on the suitability of one reference annotation for evaluating pattern discovery algorithms. Due to the lack of a clear music theoretic notion of what constitutes a pattern, we therefore propose to take a data-driven approach towards the understanding of the notion of musical patterns, by gathering multiple annotations for the same piece and analysing the disagreement and agreement between different annotators on the discovered patterns.

Therefore, we present a new musical patterns dataset: HEMAN (Human Estimations of Musically Agreeing Notes) where multiple perspectives on six musical excerpts are made available. We asked 12 subjects to annotate patterns in six musical excerpts. HEMAN is a digitised, open source version of the dataset introduced in (Nieto, Farbood, 2012). While we show that there exists considerable disagreement between annotators on what constitutes the patterns of a piece, we demonstrate that the disagreement can be reduced by 1) considering only the patterns that annotators have rated as highly relevant and 2) by lowering the time resolution.

[1] In this paper we use the terms "patterns" and "motifs" interchangeably.

The two main contributions of this paper are:

- Releasing digitised pattern annotation data in the JAMS (JSON Annotated Music Specification) format (Humphrey et al, 2014) and time interval format for facilitating future research.
- Analyses on the HEMAN dataset including two methods for alleviating pairwise annotator disagreement: use annotations that are rated as highly relevant and reduce the time resolution of the annotated intervals.

Experimental setup

The annotation process for obtaining the dataset was conducted at New York University (NYU). Subjects were all graduate students at NYU and had an average of 10 years of formal musical training (Standard Deviation = 2.3). Detailed information on the subject's music experience background can be found in the next section.

The HEMAN collection comprises 6 music excerpts as listed below:

1. Bach – Cantata BWV 1, Movement 6, Horn:
2. Bach –Cantata BWV 2, Movement 6, Soprano:
3. Beethoven –String Quartet, Op. 18, No. 1, Violin I
4. Haydn –String Quartet, Op. 74, No. 1, Violin I
5. Mozart –String Quartet, K. 155, Violin I
6. Mozart –String Quartet, K. 458, Violin I

Some of these excerpts were chosen because they were particularly hard for humans to analyse given the structural ambiguity and creative variations of the musical material. For example, the Bach chorale had very little rhythmic variation or clear grouping cues aside from phrase ending points. The other type of excerpts contains many evident and rigid repeated patterns.

Each piece was annotated by the 12 subjects. Unlimited time was given to the subjects. We did not reveal the name of the pieces on the annotation sheet. The following instructions were given to the annotators:

"Please, analyze the following musical excerpts and mark all the musical motives you can find. A musical motive is defined as a short musical idea, a salient recurring figure, musical fragment, or succession of notes that has some special importance in or is characteristic of a composition. It shouldn't be longer than a musical phrase. If you find a motive that is similar to another (or multiple versions of a motive), choose the one that you think is the most representative. Even though all motives are relevant, please rate each one of them from 1 to 3:

- 1 = Not as relevant
- 2 = Relevant
- 3 = Highly relevant

You can listen to the music excerpts as many times as you like. You can find them here. http://urinieto.com/NYU/Research/MotivesExperiment/"

We deliberately offered several possible interpretations as to what defines a "musical motif" and allow the subjects to both analyse the pieces and use their musical intuitions on what constitutes a musical motif in the process. We did not ask them to laboriously label all occurrences of the same pattern for us. In this way, we obtain at least the prototypes of musical patterns the subjects perceived in the piece.

The HEMAN Dataset

In this section, we present the content of the dataset, how we digitised it, and some difficulties we encountered and decisions we made during the digitising process.

The musical background of the subjects

In Table 1, we report the results of the musical background questionnaire of our subjects.

Table 1. Musical background of annotators. The header Inst and Inst2 denote the main and the secondary instrument the participant plays, followed by how many years of experience. The headers Theory and Overall denote the self-rated musical theory background level and the self-rated musical background level. A professional background has a score of 5 and no background has a score of 1. The header Abs stands for absolute pitch/perfect pitch, with 0 stands for self-identified as no perfect pitch, 1 otherwise.

Inst	Year	Inst2	Yea r	Theory	Overall	Abs
Piano	10	Violin	4	3	3	0
Oboe	5	Piano	1	2	3	0
Piano	5	None	0	3	3	0
Piano	12	None	0	5	5	1
Piano	14	Trumpet	8	4	4	0
Guitar	13	Bass	0	5	5	0
Piano	13	Violin	4	4	4	0
None	0	None	0	0	0	0
Guitar	8	Piano	4	5	5	0
Guitar	10	Kazoo	4	5	5	0
Violin	8	Guitar	2	3	3	0

As we can see, we have annotators with different levels of musical background. Most of them are trained with western musical instruments. In the following subsections, we detail how the making of the dataset started with paper-based annotation, how we calculate the time intervals from the photos and convert the musical patterns to a numerical format, and finally how we convert the numerical format to the JAMS format.

Digitising paper-based annotations

Figure 1 shows an example of a raw annotation. From this notation, we first calculate the time intervals [start time, end time] of the pattern annotations. We take the unit of a crochet as one in this time interval format. For example, the first annotated pattern in the time interval format is [0,2], and the second pattern is [8,10].

There are advantages and disadvantages with using a paper-based setting. On the one hand, this could preserve the most natural mindset on perceiving patterns in music; on the other hand, this poses some risks for the correct interpretation of the markings. For example, in Figure 1, it is not immediately apparent whether to include the last quiver in bar 3 into the musical pattern. The same situation applies to the crochet in bar 14 in the patterns. The rule we followed here is to take the midpoint of the gap between the two notes in questions, and depending on which side the annotation start/end with respect to the midpoint, we take include/exclude the note in the patterns. For example, in both bar 3 and bar 14, we include the quiver and the crochet. Furthermore, we noticed that some annotators forgot to mark the relevance even though their time interval annotations look reasonable. In this case, we give all patterns a relevance score of 1.

As mentioned in the last section, since the annotators were not forbidden to mark the occurrences in addition to the prototype patterns, we do find such occurrences in the photo. We take those occurrences into account as long as a relevance score is given.

Finally, although transcribing from the photo format to the time interval format is relatively time consuming and prone to human errors, we do not know of any matured technology which could be used to automatically convert between the two formats. Optical recognition techniques are promising, but could give low accuracies with the imprecise markings.

In the future, it would be ideal to have a digital system which could be as natural a process for the annotators as paper and pen. For a large scale online experiment, it would only be viable with developing such an annotation system.

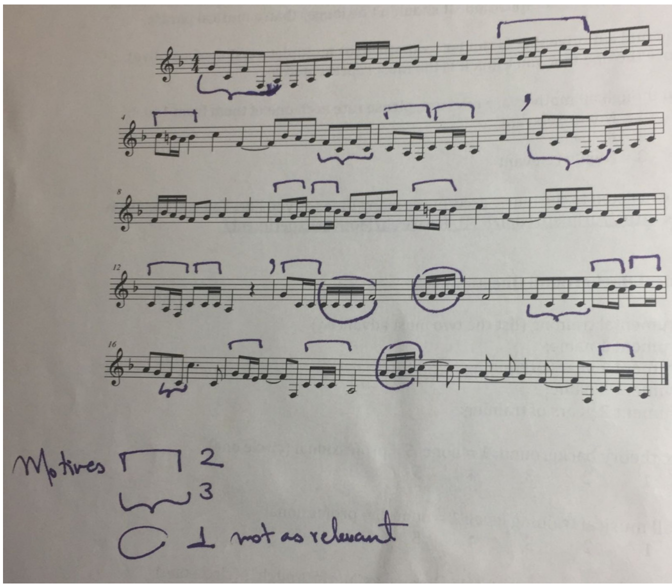


Figure 1. Example of the raw data of annotation

Visualising the time interval data

After converting to the time interval format, we can visualise the data taking the same approach as in (Ren et al 2017) as shown in Figure 2. In this visualisation, we abstract away the actual notes, just preserving the temporal markings in the excerpts. We can see that the disagreement amongst the annotators is prevalent.

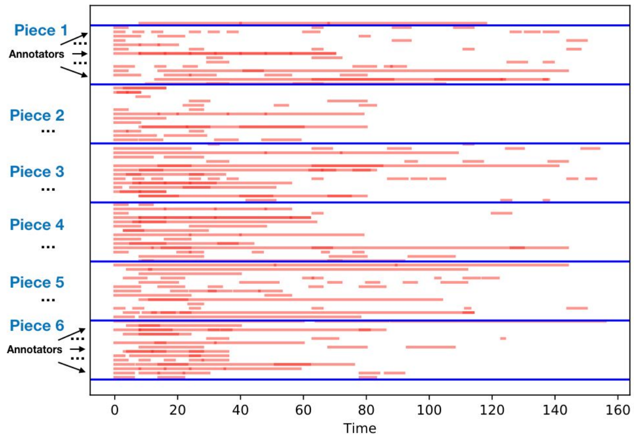


Figure 2. Time intervals in the six musical excerpts from 7 different annotators. Red bars indicate that there exist a pattern, and the absence of red bars indicates the absence of patterns. The blue horizontal lines separate different pieces. The x-axis represents time in the unit of a crochet. The y-axis represents different annotators. Due to limited space, the time intervals of only seven annotators are shown here. We can see different degrees of agreement and disagreement amongst the annotators.

Time intervals to numerical format

After obtaining the time interval data, we took the symbolic music data from a python toolbox, music21. By segmenting the music21 data using the time intervals, the numerical format of a (onset, duration, pitch) triplet was created. Each pattern is thus represented by a succession of triplets and its associated relevance score. We further organise each individual pattern into a "Excerpt -> Annotator -> Relevance -> Pattern" hierarchy in a python dictionary.

Numeric format to JAMS

From the hierarchical structure of the numerical symbolic musical patterns, we further convert the data to the JAMS format. JAMS provides a simple, structured, and sustainable approach to representing rich information in a human-readable, language agnostic format. This format fits our purpose well because it supports multiple types of annotations, multiple annotations for a given task, and rich file level and annotation level metadata. The dictionary format is then converted to the JAMS format using the JAMS python library and a HEMAN parser script, which can be accessed at the official repository of JAMS: https://github.com/marl/jams-data/blob/master/parsers/heman_parser.py. The time-interval representation, the numeric format, and the JAMS files of the dataset can all be accessed online: https://github.com/irisyupingren/HEMANanalysis.

Method

In this section, based on intuition and observation on the dataset and findings about relevance values in the context of segmentation (Bruderer, Mckinney & Kohlrausch, 2009), we investigate our hypothesise that, by taking the most relevant patterns, and lowering the time resolution, we obtain more pairwise annotator agreement in the dataset.

To conduct our computational analysis and verify our hypothesis, we take the time interval representation of the annotations. We do not need to consider the actual notes when analysing pairwise annotator agreement because the only disagreement possible is the starting and/or ending time given a specific piece.

To measure pairwise agreement, we take each individual annotator as the reference and use the standard precision, recall, and F1 score as measurements of agreement (Goutte & Gaussier 2005). A formal definition is given below:

$$\text{Precision} = \frac{\#\ \textit{matched annotations}}{\#\ \textit{annotations of the referenced annotator}}$$

$$\text{Recall} = \frac{\#\ \textit{matched annotations}}{\#\ \textit{annotations of the current annotator}}$$

$$\text{F1} = \frac{\textit{Precision} \times \textit{Recall}}{\textit{Precision} + \textit{Recall}} \times 2$$

The notion of matched annotations is defined on different levels of time resolution: {Annotation1=[begin1, end1], Annotation2=[begin2, end2]} $\in$ {Matched annotations}, if |begin1- begin2| + |end1 - end2| $\leq$Threshold. The essence of lowering the time resolution of the annotations is taking a larger tolerance on identifying whether two annotations are agreeing (matched) or disagreeing (not matched). In this way, we can see how much disagreement there is on different scales of time resolution.

We use precision, recall, and F1 score instead of the kappa agreement measures like in (Balke et al, 2016) because this approach simplifies the calculation and avoids taking the average across the musical piece.

In addition to numerical methods, we examine the annotations analytically and categorically. In Figure 3, we show an example of disagreement amongst three annotators on the same piece.

We categorise the possible types of disagreement as follows (Annotation1 = [a1,b1], Annotation2 = [a2,b2], symmetrical cases with switched order of the annotators are omitted):

- On the individual pattern level
  - When there is a match
    - a1 < a2, b2 > b1
    - a1 = a2, b2 > b1
  - When there is no match
- On the piece level
  - The number of annotations
  - Whether there are overlaps of patterns
  - Occurrences
    - Exhaustive occurrences
      - with variation
      - without variation
    - Non-exhaustive occurrences
      - Arbitrary
      - Only Prototypes

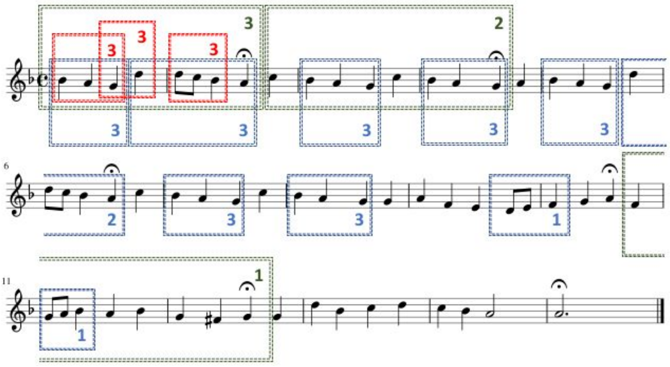


Figure 3. Examples of disagreement. Different colours indicate different annotators. The relevance values are marked with respect to each pattern.

Results

In this section, we show the numerical values of precision, recall, F1 score, and analysis of the annotations. Our initial exploration shows supporting evidence for our hypothesis that by using the top-ranked annotations and lowering the time resolution of the annotations, we may obtain more pairwise annotator agreement in the dataset. We first discuss effects of the relevance score and the time resolution threshold separately, and then see their effects together.

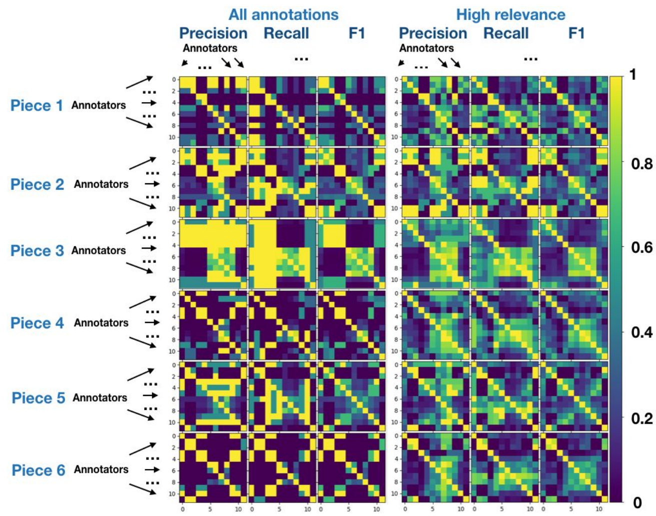


Figure 4. Precision, recall and F1 score across pieces. Different rows denote different pieces and different annotators. Different columns denote different metrics and relevance scores. Within each subfigure, we show the results of 12×12 annotators on one metric $\in$ {Precision, Recall, F1}. Yellow side of spectrum denotes high agreement.

Relevance

In the left subfigure of Figure 4, we show the measurement of agreement using precision, recall and F1 score computed using all annotations; in the right subfigure, we show the precision, recall and F1 score computed using the annotations which are rated as the most relevant.

When considering all annotations, we see a *grouping* phenomenon where subclusters amongst annotators with high agreement are formed. By considering the high relevance annotations, the grouping phenomenon is reduced and we can see an overall increase of agreement by comparing the results in Figure 4.

Time resolution thresholding

For showing the effects of the time resolution thresholding step, we only show the average across the six excerpts instead of looking at the excerpts individually. The averaged results already show a clear sign of increased agreement with a lower time resolution.

By comparing row by row, we observe in Figure 5 that, with a lower time resolution, that is, a more relaxed threshold, the precision, recall and F1 scores increase. The relaxation of the threshold is effectively loosening the notion of "matched annotations", as mentioned in the last section. For example, if Annotation1 = [a1,b1], Annotation2 = [a1, b1 + 2ε], with threshold = ε, the two annotations are not matched; with threshold = 2ε, the two annotations are matched. With different degrees of loosening, we can examine the pairwise agreement amongst annotators on different scales of time resolution.

In the case where all annotators have the same annotations, taking different time resolution values does not have an effect on the precision, recall and F1 score. In the case of the different annotations, the metrics will reach a stable value as the threshold increases. In our case, we see an increasing trend in the metrics as the threshold increases. Therefore, more pairwise annotator agreement is reached on lower time resolutions.

Relevance ranking and time resolution thresholding combined

In Figure 5, we show the combined effects of relevance ranking and time resolution thresholding. As expected, the agreement increases with the effects from both steps. We therefore conclude that, by using extra rankable metadata on the annotations such as relevance/importance scores, and by adjusting the time resolution of the annotated patterns' time span, we achieve a higher degree of pairwise annotator agreement in the HEMAN dataset.

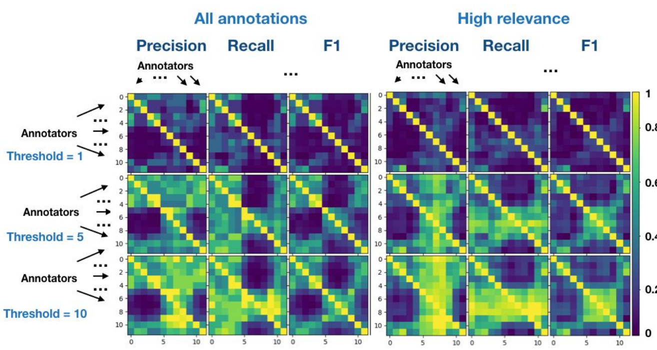


Figure 5. The effects of threshold change and relevance value. The precision, recall and F1 score across pieces are averaged. The threshold values are given as multiples of quarter note length. Other specifications are the same as Figure 4: Different rows denote different pieces and different annotators. Different columns denote results denote different metrics and relevance scores. Within each subfigure, we show the results of 12×12 annotators on one metric $\in$ {Precision, Recall, F1}.

Conclusion

Based on our data, we draw the tentative conclusion that motifs or patterns rated as highly relevant are more trustworthy than those rated with lower relevance values. In addition, by lowering the time resolution of annotations, we gain more agreement amongst the annotators. Therefore, if we choose the patterns of high relevance and threshold the annotations, we could establish an evaluation measure with small irreducible errors for automatically extracted patterns.

In the future, we would like to extend the work by collecting more data from a larger number of subjects using a web interface and verify our current conclusions with more data and further similarity analysis. We could also design other related experiments where we would provide candidate patterns and ask the subjects to choose. Another exploration could be to ask the subjects to annotate and rank the patterns in complete musical pieces rather than excerpts. Finally, for algorithms, we will put the dataset into use for computational pattern extraction evaluation tasks, and see if the current state-of-the-art algorithms can reproduce human annotations with high agreements. The agreement values could also be used as training data for novel computational pattern extraction models, such that the models could give a confidence value when predicting the presence of patterns.

Acknowledgements. We thank the participants of the experiment and everybody who stimulated thought-provoking discussions.

References

Balke, S., Driedger, J., Abeßer, J., Dittmar, C., & Müller, M. (2016). Towards Evaluating Multiple Predominant Melody Annotations in Jazz Recordings. Proc. of the 15th International Society for Music Information Retrieval Conference (pp. 246-253). New York City, United States.

Benward, B. (2014). Music in Theory and Practice Volume 1. McGraw-Hill Higher Education.

Bernstein, L. (1976). The unanswered question: Six talks at Harvard (Vol. 33). Harvard University Press.

Bruderer, M. J., Mckinney, M. F., & Kohlrausch, A. (2009). The Perception of Structural Boundaries in Melody Lines of Western Popular Music. Musicæ Scientiæ, 13(2), 273–313.

Burkhart, C. (2005). The phrase rhythm of Chopin's A-Flat Major Mazurka, Op. 59, No. 2. Engaging music: Essays in music analysis, 3-12.

Collins, T., Böck, S., Krebs, F., & Widmer, G. (2014, January). Bridging the audio-symbolic gap: The discovery of repeated note content directly from polyphonic music audio. In Audio Engineering Society Conference: 53rd International Conference: Semantic Audio. Audio Engineering Society.

Cross, I. (2005). Music and meaning, ambiguity and evolution. Musical communication, 27-43.

Collins, T. (2011). Improved methods for pattern discovery in music, with applications in automated stylistic composition (Doctoral dissertation, The Open University).

Collins, T., Janssen, B., Ren, I. Y. & Volk, A. (2017). Discovery of Repeated Themes & Sections, 2017. http://www.musicir.org/mirex/wiki/2017:Discovery_of_Repeated_Themes_%26_Sections. Accessed on 20 May 2018.

Goutte, C., & Gaussier, E. (2005, March). A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In European Conference on Information Retrieval (pp. 345-359). Springer, Berlin, Heidelberg.

Hiley, D. (1993). Western plainchant: a handbook. Oxford University Press.

Humphrey, J.E., Salamon, J., Nieto, O., Forsyth, J., Bittner, R., Bello, J.P., JAMS: A JSON Annotated Music Specification for Reproducible MIR Research. Proc. of the 15th International Society for Music Information Retrieval Conference (pp. 591-596). Taipei, Taiwan.

Lerdahl, F., & Jackendoff, R. S. (1985). A generative theory of tonal music. MIT press.

McFee, B., Nieto, O., Farbood, M. M., & Bello, J. P. (2017). Evaluating Hierarchical Structure in Music Annotations. Frontiers in psychology, 8, 1337.

Nattiez, J. J. (1990). Music and discourse: Toward a semiology of music. Princeton University Press.

Nieto, O., & Farbood, M. M. (2012). Perceptual evaluation of automatically extracted musical motives. In Proceedings of the 12th International Conference on Music Perception and Cognition (pp. 723-727).

Randall, D. M. (1999). The Harvard concise dictionary of music and musicians. Harvard University Press.

Ren, I. Y., Koops, V., Volk, A., & Swierstra, W. (2017). In Search Of The Consensus Among Musical Pattern Discovery Algorithms. Proc. of the 18th International Society for Music Information Retrieval (pp. 671-677). Suzhou, China.

Thom, B., Spevak, C., & Höthker, K. (2002, September). Melodic segmentation: evaluating the performance of algorithms and musical experts. International Computer Music Conference.

Sadie, S., & Tyrrell, J. (2001). Dictionary of music and musicians. New York: Oxford University Press. Yónatan Sánchez.

Schoenberg, A. (1983). *Theory of harmony*. Univ of California Press.

Scholes, P. A. (1970). The Oxford Companion to Music. Oxford University Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Formant Distances and the Similarity Perception of Wind Instrument Timbres

Christoph Reuter[1], Isabella Czedik-Eysenberg[1], Saleh Siddiq[1], Michael Oehler[2]

[1]*Musicological Department, University of Vienna, Austria*
[2]*Musicological Department, University of Osnabrück, Germany*
[1]christoph.reuter@univie.ac.at, isabella.czedik-eysenberg@univie.ac.at, saleh.siddiq@univie.ac.at, [2]michael.oehler@uni-osnabrueck.de

Abstract

Do shorter distances between the formant positions of two sounds in a Euclidean space correlate with the impression of greater timbre similarity (and vice versa)? Is the numerical vector of formant positions a suitable basis for a precise computational classification of the involved instruments? Plotting the formants of 586 wind instrument sounds at all reachable pitches in ***ff*** and ***pp*** into a X/Y scatter diagram (X-axis: first formant, Y-axis: second formant), it is possible to visually distinguish musical instruments and their registers by their formant positions. In a listening test, 22 participants rated the (dis)similarity of 40 loudness-adjusted timbre combinations of wind instruments. Half of the stimuli contained sounds with extremely close/overlapping formant regions while the other half contained sounds with very distant formant regions. Moreover, with the help of machine learning methods, we tested the precision of the instruments' classification by means of their formants using 5-fold cross validation.

It turned out that timbres with very close formant positions are perceived as very similar while timbres with very distant formant positions are deemed very dissimilar. Based on these results we calculated a prediction model and compared it to a timbre similarity prediction model based on MFCCs.

When matching the formant positions of the examined timbres with their respective musical instruments we achieved a classification precision of 46.1% (cubic KNN). By adding a set of further timbre descriptors, we could increase the classification precision to 84.6% (quadratic SVM). The resulting confusion matrix corresponded very well with human mismatching of musical instrument timbres. In summary, the perception of timbre similarity in wind instruments was shown to be closely related to the sounds' formant positions. Regarding computational musical instrument classification, formant positions alone do not seem to be a sufficient criterion compared to other more powerful feature combinations, but regarding the evaluation of perceived timbre similarity, the distance between formant positions can serve as a helpful explanatory tool.

Background

It is a well-known fact in timbre research that the timbral character of a musical instrument as a whole cannot be sufficiently represented by the sound attributes of a single tone. Rather, musical instruments' timbres and the timbre of a single tone should be treated as two different concepts (Stumpf, 1926, p. 393; Siedenburg, Jones-Mollerup, & McAdams, 2016, p. 15). Nevertheless, in most publications about timbre research a single tone is treated as a prototypical representation of its source instrument. Only few concepts for a wider timbral description of an instrument through its entire dynamics and pitch range do exist. Beside Mel Frequency Cepstral Coefficients (MFCCs, e.g. Loughran, Walker, O'Neill, & O'Farrell, 2008) and Modulation Power Spectrum (MPS, e.g. Elliott, Hamilton, & Theunissen, 2013) formant areas were shown to be very useful for a more comprehensive description of musical instruments' timbres (since 1926, e.g. Stumpf, 1926; Schumann, 1929; Mertens, 1975; Meyer, 2009 etc.). Therefore, based on the formant areas of 586 wind instrument timbres in all reachable pitches and two dynamic levels we compiled a two-dimensional formant map (X-axis: formant 1, Y-axis: formant 2). The calculations were done using Praat (Boersma, & Weenink, 2013). In this map, the musical instruments (represented as point clouds in different colors) emerged to be clustered together by instrument, dynamics and register (see fig. 1). The calculated formant positions correspond to the descriptions found in the literature (e.g. Schumann, 1929; Mertens, 1975; Reuter, 1996; Meyer, 2009 etc.).

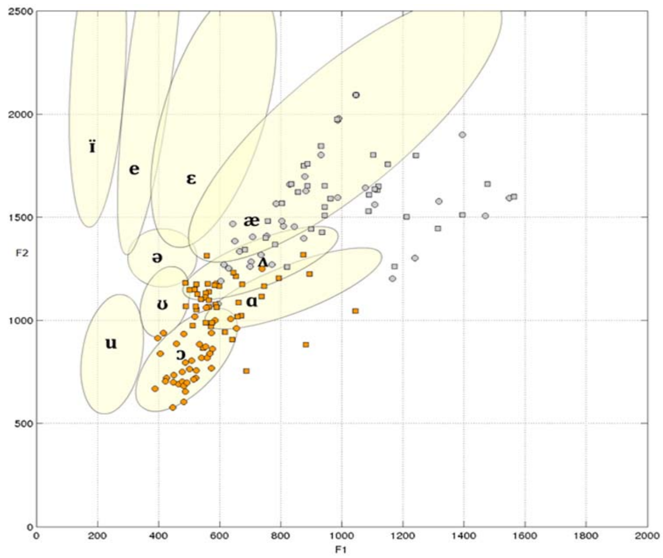


Figure 1. Formant map with the sounds of bassoon (orange) and oboe (grey) in all achievable pitches in *ff* and *pp* (Reuter, Czedik-Eysenberg, Siddiq, & Oehler, 2017).

Aims

These relatively well separated point clouds suggest that formant positions can perhaps serve as a useful and intuitive representation for a timbral description and classification of musical instrument.timbres. In addition, it indicates that sounds with short distances between their formants should be perceived as more similar than sounds with more distant formants. So, the aim of this study is to test the possible advantages and suitability of the formant paradigm for musical timbre similarity prediction as well as for timbre classification by investigating the following questions:

- Do shorter distances between the formant positions of two sounds correlate with the impression of greater timbre similarity (and vice versa)?

- Is the numerical vector of formant positions an suitable basis for a precise computational classification of the involved instruments? (Besides, which additional timbre features can improve the classification result?)

Computational description of perceived timbre similarity: Formants and MFCCs in comparison

In a listening test, 22 participants rated the (dis)similarity of 40 loudness-adjusted timbre combinations including wind instruments like flute, oboe, clarinet, bassoon, trumpet, trombone, French horn and tuba on a scale between 1 and 8 (8 = maximum dissimilarity). Half of the stimuli contained sounds with extremely close/overlapping formant regions while the other half contained sounds with very distant formant regions. Instrumental sounds were taken from the Vienna Symphonic Library (VSL) and adjusted to a matching ANSI-loudness level using the Genesis loudness toolbox in Matlab. The listeners' judgments were tested for correlations with the Euclidean distances between positions in the formant map. Indeed, a very significant correlation between the formant distances and the listeners' similarity judgments was found ($r = 0.759$, t-test with $p < 0.001$, confidence interval: [-3.1381, -1.8960]).

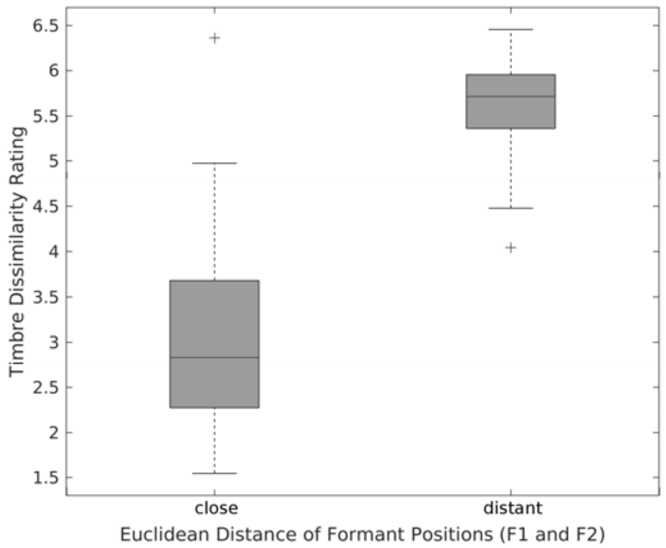


Figure 2. The distance of the formant positions (X-axis: close vs. distant) correlates strongly with ratings of perceived timbre similarity (Y-axis: 1 = very similar; 7 = very dissimilar).

Here, the first two formants (F1 and F2) have an almost equally strong linear relationship with the similarity scores and also have a very strong intercorrelation ($r = 0.9196$, $p < 0.0001$). In comparison, each of the first three MFCCs correlates weaker with the listeners' similarity scores, while the remaining MFCCs (4–13) show no significant correlations to the listeners' judgments at all. (Formants were calculated in Praat (Boersma & Weenink, 2013) and MFCCs in Matlab with the help of the MIRtoolbox (Lartillot & Toiviainen, 2007).)

Table 1. Correlation of individual formant positions and MFCCs with the perceived timbre similarity.

Timbre feature	r	p
F1	0.7514	< 0.0001
F2	0.7477	< 0.0001
F3	0.4227	< 0.0001
MFCC1	0.6384	< 0.0001
MFCC2	0.5959	< 0.0001
MFCC3	0.3513	< 0.05

A similar result can be seen when considering formant and MFCC combinations. In a two- or three-dimensional formant or MFCC space, the Euclidean distances based on the formants 1 and 2 correlate most strongly with the perceived timbre similarity ($r = 0.759$, $p < 0.0001$, see Fig. 3), while the Euclidean distances based on the MFCCs 1–3 give a slightly weaker but comparable picture ($r = 0.695$, $p < 0.0001$, see Fig. 4).

Table 2. Correlation of the Euclidean distances of formant (F1, F2, and F3) and MFCC combinations with perceived timbre similarity.

Timbre descriptors	r	p
F1 and F2	0.7591	< 0.0001
MFCCs 1, 2, and 3	0.6949	< 0.0001
MFCCs 1 and 2	0.6939	< 0.0001
F1, F2, and F3	0.6916	< 0.0001
MFCCs 1–13	0.6812	< 0.0001

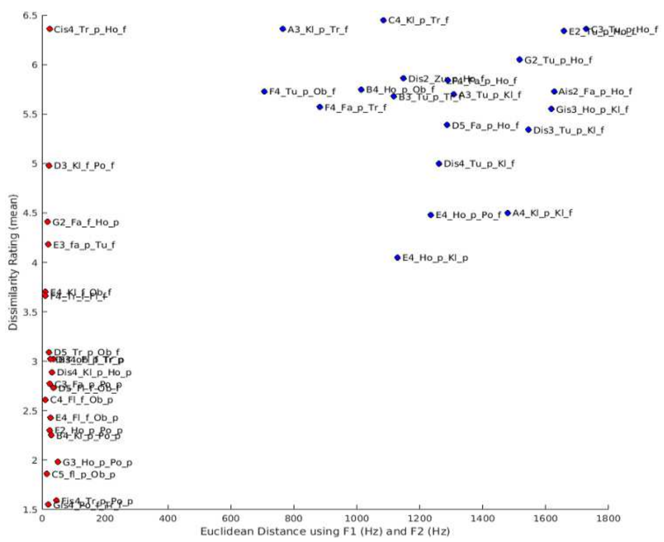


Figure 3. Scatter plot of the Euclidean distances of the first two formant positions (X-axis) and the perceived timbre similarity (Y-axis) (red: close formant positions, blue: distant formant positions).

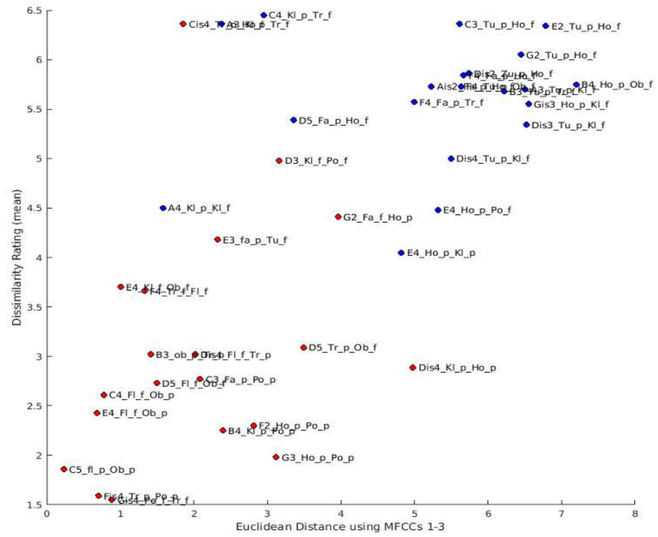


Figure 4. Scatter plot of the Euclidean distances of the first three MFCCs (X-axis) and the perceived timbre similarity (Y-axis) (red: close formant positions, blue: distant formant positions).

In a spider's web representation with the axes F1, F2, and F3 respective MFCC1, MFCC2, MFCC3 the triangles formed between the axes overlap very closely in case of a perceived timbre similarity (see Figure 5) while the distances of the triangle sides increase with increasing timbre similarity (see Figure 6).

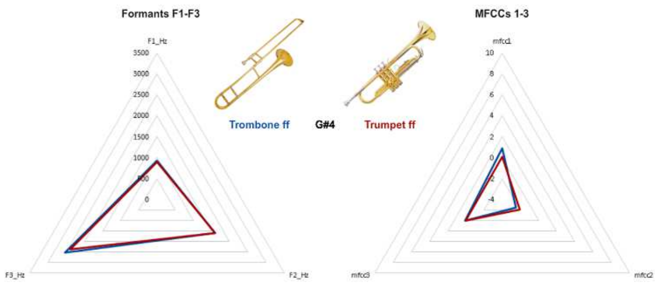


Figure 5. Spider's web representation with the axes F1, F2, F3 respective MFCC1, MFCC2, MFCC3 of the pair with the most perceived timbre similarity (trombone *ff* and trumpet *ff* on G#4).

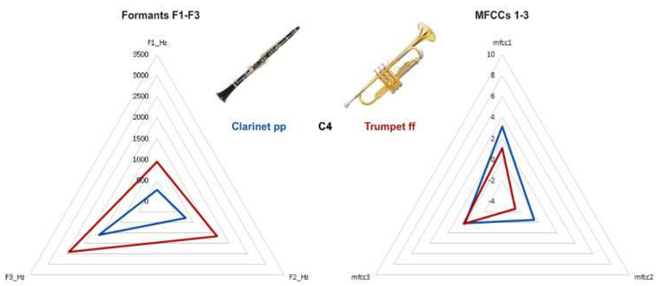


Figure 6. Spider's web representation with the axes F1, F2, F3 respective MFCC1, MFCC2, MFCC3 of the pair with the least perceived timbre similarity (clarinet *pp* and trumpet *ff* on C4).

A comparison of these results indicates that formants and MFCCs are similarly suitable for the determination of perceived timbre similarity. This is also suggested by the model of Darch et al. (2005), where formant areas can be derived on the basis of MFCCs.

Computational predictability of perceived timbre similarity: Formants and MFCCs in comparison

Regression models trained via machine learning (using 5-fold cross-validation) allow the prediction of perceived timbre similarities to a certain degree: Based on formants, the best result with a coefficient of determination (R^2) of 0.53 was found when using the absolute differences of F1 and F2 as features (see Figure 7).

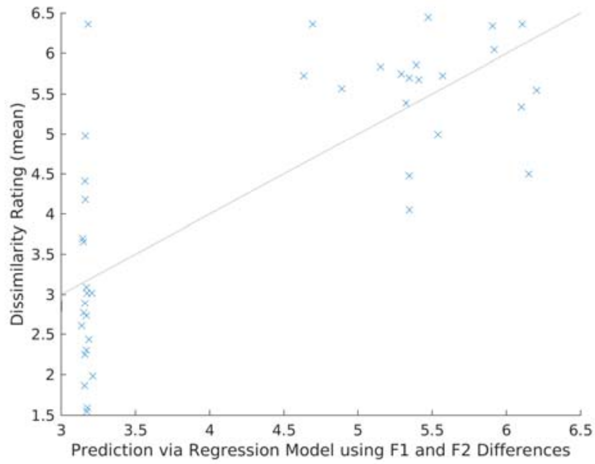


Figure 7. Prediction model based on F1 and F2 (R^2 = 0.53, RMSE =1.08, MSE =1.16, MAE =0.86).

Based on MFCCs a comparable result (R^2 = 0.56) can be obtained when usingMFCC1 and MFCC2 (see Figure 8).

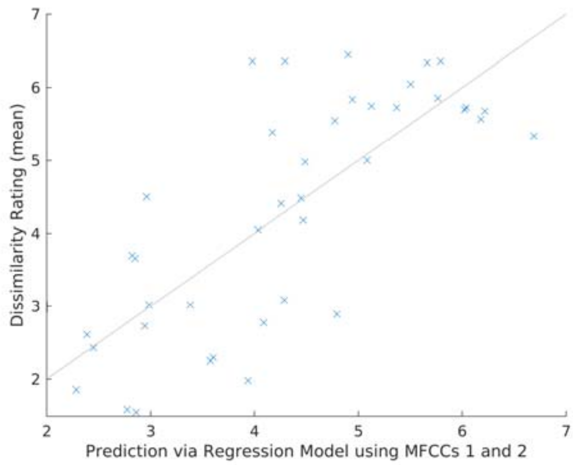


Figure 8. Prediction model based on MFCC1 and MFCC2 (R^2 = 0.56, RMSE= 1.05, MSE = 1.10, MAE = 0.81).

In other words, formants and MFCCs were approximately equally suitable for predicting timbre similarities. However, one has to take into account that other timbre features may also effectively predict perceived timbre similarities: For example, considering the distance of the spectral centroid of one sound to the other, it can be shown that with increasing distance between their values, the sounds' timbres are perceived as increasingly dissimilar (coefficient of determination (R^2) of 0.78, see Figure 9).

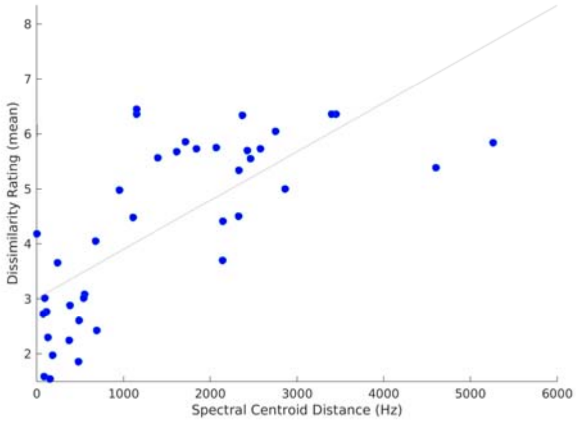


Figure 9. Prediction model based on spectral centroid (R^2 = 0.78, RMSE = 0.75, MSE = 0.56, MAE = 0.58).

Timbre classification based on formants

With the help of machine classification methods, timbre classes can be assigned to musical instruments with a precision of 46.1% based on the first three formants (F1 / F2 / F3, cubic k-nearest-neighbour classification) (see Figure 10).

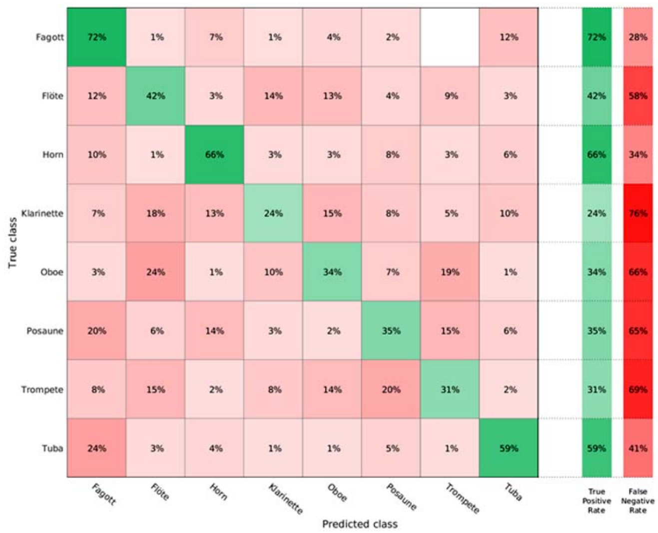


Figure 10. Confusion matrix based on F1, F2, and F3 (cubic KNN, classification precision of 46.1%).

When adding a set of complementary timbre descriptors to the formant positions, the precision of the instrument recognition can be increased to 84.6%. These additional timbre descriptors are attack time, spectral flux, roughness, brightness, spectral entropy, maximum RMS value, spectral centroid and unpleasantness. The last descriptor is implemented based on the results in Reuter, Oehler, & Mühlhans (2013), the rest is calculated with the help of Matlab MIRtoolbox (Lartillot & Toiviainen, 2007). The resulting confusion matrix shows plausible parallels to human perception, in the sense that instrument confusions frequently occurring in humans are also mirrored in the matrix. Examples are the confusions between flute, clarinet and oboe, between trumpet and trombone as well as in between tuba and bassoon timbres (see Figure 11).

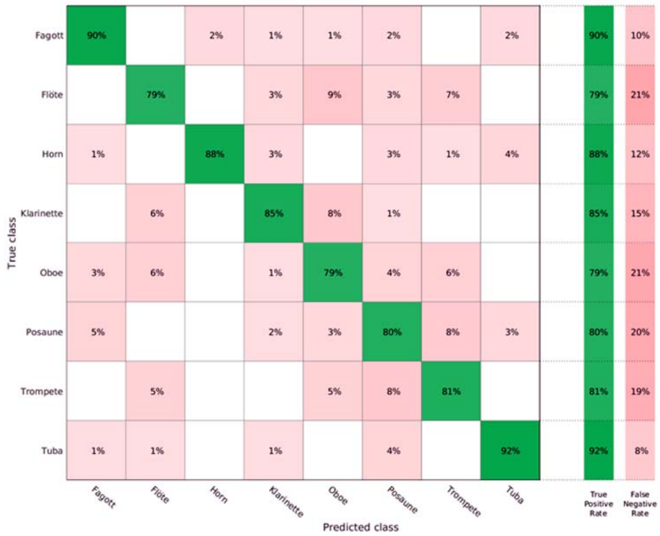


Figure 11. Confusion matrix based on F1, F2, and F3 as well as on attack time, spectral flux, roughness, brightness, spectral entropy, maximum RMS value, spectral centroid, and unpleasantness (quadratic SVM, classification precision of 84.6%).

Conclusion

The timbre similarity of wind instruments corresponds well to the proximity of the first two formants. A similarity prediction approach based on formant positions yielded results comparable to using MFCCs. Constructing classification models, we showed that a formant-driven instrument identification is possible, but does not lead to particularly high recognition rates. However, adding a small number of additional timbre descriptors yields an effective classification model that also shows correspondences to human perception. While formants, as sound descriptors for timbre similarity perception, strongly correlate with each other, individual MFCCs are more linearly independent of each other. However, the advantages of formants over MFCCs are: (1) formants need only two values to compactly describe a distinctive audible and therefore intuitively accessible spectral content and (2) they provide a solid foundation through more than 100 years of research history. For these reasons, formants can be considered as useful alternatives to MFCCs when analysing (wind) instrument timbres in music information retrieval.

Acknowledgements. This project is supported by funds of the Österreichische Nationalbank (OeNB, Anniversary Fund, project number: 16473). Special thanks to the Vienna Symphonic Library for supporting our work with two free "Vienna Super Package" licenses.

References

Boersma, P., & Weenink, D. (2013). *Praat: Doing Phonetics by Computer* [Computer program], Version 5.3.51, retrieved 2 June 2013 URL: http://www.praat.org/.

Darch, J., Milner, B., Shao, X., Vaseghi, S., & Yan, Q. (2005). Predicting Formant Frequencies from MFCC Vectors. In *Proceedings of the 2005 IEEE International Conference on Acoustics, Speech, and Signal Processing, Volume I of V*. (pp. I/941–944). Pennsylvania, USA: Pennsylvania Convention Center/Marriott Hotel.

Elliott, T., Hamilton, L., & Theunissen, F. (2013). Acoustic Structure of the five Perceptual Dimensions of Timbre in Orchestral Instrument Tones. *Journal of the Acoustical Society of America*, 133(1), 389–404.

Lartillot, O., & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction from Audio. In *Proceedings of the 10th International Conference on Digital Audio Effects (DAFx)*(pp. 237–244). France, Bordeaux.

Loughran, R., Walker, J., O'Neill, M., & O'Farrell, M. (2008). Musical Instrument Identification using Principal Component Analysis and Multi-Layered Perceptrons. In *Proceedings of the International Conference on Audio, Language and Image Processing ICALIP 2008* (pp. 643–648). China: Shanghai University.

Mertens, P.-H. (1975). *Die Schumannschen Klangfarbengesetze und ihre Bedeutung für die Übertragung von Sprache und Musik* [The Schumann Laws of Timbre and their Meaning for the Transmission of Voices and Music]. Frankfurt: Bochinsky.

Meyer, J. (2009). *Acoustics and the Performance of Music. Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*. New York: Springer.

Reuter, C. (1996). *Die auditive Diskrimination von Orchester-instrumenten* [Auditive Separability of Orchestral Instruments]. Frankfurt: Lang.

Reuter, C., Czedik-Eysenberg, I., Siddiq, S., & Oehler, M. (2017). Formanten als hilfreiche Timbre-Deskriptoren für die Darstellung von Blasinstrumentenklängen [Formants as helpful Timbre Descriptors for the Depiction of Wind Instruments Timbres]. In *Proceedings of the 43th. German Annual Conference on Acoustics „Fortschritte der Akustik". DAGA 2017* (pp. 190–193). Germany: University of Kiel.

Reuter, C., Oehler, M., & Mühlhans, J. (2014). Physiological and acoustical correlates of unpleasant sounds. In *Proceedings of the Joint Conference ICMPC13/APSCOM5* (p. 9). South Korea: Yonsei University, Seoul.

Schumann. K.E. (1929). *Physik der Klangfarben* [Physics of Timbres]. (Habilitation Thesis). University of Berlin, Berlin, Germany.

Siedenburg, K.; Jones-Mollerup, K., & McAdams S. (2016). Acoustic and Categorical Dissimilarity of Musical Timbre: Evidence from Asymmetries between Acoustic and Chimeric Sounds. *Frontiers in Psychology*, 6:1977.

Stumpf, C. (1926). *Die Sprachlaute* [The Sounds of Voice]. Berlin: Springer.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Identifying the Perceptual Dimensions of Musical Instrument Timbre

Lindsey Reymore,[1] David Huron[2]

[1,2] *School of Music, The Ohio State University, The United States*

[1]reymore.1@osu.edu, [2]huron.1@osu.edu

Abstract

Different musical instruments do not simply exhibit different musical timbres; they also evoke different phenomenological experiences, or "musical instrument qualia." Informal descriptions of instrument sounds seem to employ stereotypical characterizations, such as "airy" for a flute or "heavy" for a tuba. Previous research has sought to identify dimensions of timbre through listener judgments of paired comparisons (Grey, 1977, Kendall et al., 1999). Investigations into the semantic dimensions of timbre demonstrate that a number of descriptive terms map consistently onto particular acoustic correlates of timbre (e.g. Zacharkakis, Pastiadis, & Reiss, 2014).

This three-part project aims to develop a model of the dimensions of musical instrument qualia. Results of the first two studies are reported here.

In the first study, interviews were conducted with professional and semi-professional musicians in which they were asked to open-endedly describe the sounds of different musical instruments. Responses were subjected to content analysis using a pile-sort method, yielding an initial list of categories of musical instrument qualia. To test the reliability and usefulness of the dimensions yielded from the first study, the second study asked participants to rate musical instrument sounds according to the pile-sorted categories. Principal components analysis suggests ways in which categories with significant overlapping variance can be collapsed; preliminary results of this analysis are presented here.

In the last stage of the study (not reported here), participants will rate an expanded number of instruments according to each of the dimensions in the final PCA model; these results will then be used to generate perceptual profiles for the most common Western musical instruments.

Introduction

The human voice is perhaps the most universal of all music-making instruments. However, nearly all cultures supplement the voice with a selection of mechanical instruments such as drums and flutes. Although some cultures employ only a handful of instrumental resources, many cultures provide a broader palette of possible instruments for music making. The Western symphony orchestra provides a notably varied selection of instruments. In addition to percussion instruments, the modern orchestra provides roughly 16 common core instruments: piccolo, flute, oboe, English horn, clarinet, bass clarinet, bassoon, contrabassoon, French horn, trumpet, trombone, tuba, violin, viola, cello, and contrabass. When creating an orchestral work, the composer might draw on any of these instruments individually, and when used in combination, the possibilities that 16 instruments afford are enormous. Less common instruments are also frequently added for a variety of artistic purposes, such as the alto flute or Wagner tuba. Since most musical lines can be played by more than one instrument, why might a composer choose one instrument or instrument combination over another? For example, in the second movement (Funeral march) of Beethoven's third symphony, why does Beethoven assign the solo line to an oboe rather than to a flute or French horn?

Different instruments don't simply exhibit different timbres. Instrument sounds evoke different qualia, associations, and phenomenological experiences. Informal descriptions of instruments often employ stereotypic characterizations, such as when a French horn is described as "noble," or a tuba is described as "heavy." How widespread are such associations, and where do these characterizations come from? In this study we propose to identify the kinds of qualia commonly associated with different musical instruments.

Considering possible dimensions for musical timbre is hardly new. Grey (1977) pioneered the use of multidimensional scaling (MDS) as a tool for investigating instrument timbres. The MDS approach received considerable extension in work by Roger Kendall and Edward Carterette (e.g., Kendall, 1993, 1999). In MDS, similarity judgments are used to construct a multidimensional map of the relationships between sounds. The dimensions themselves remain unlabeled, and it is up to the researcher to infer and interpret latent dimensions that may be revealed through MDS. MDS relies simply on listeners characterizing pairs of timbres according to the degree of similarity. MDS provides no direct approach for understanding the qualia that are evoked by different sounds. Consequently, in our study, the principal aim is to elicit qualia judgments more directly through self-report, rather than attempting to infer them indirectly through paired timbre comparisons.

Perceptual studies have also addressed the semantics of timbre. Many of these studies have found three to four semantic dimensions of timbre space (von Bismarck, 1974, Pratt and Doak, 1976, Kendall and Carterette, 1991). Elliot et al. (2012) triangulated a semantic approach, in which participants rated adjectives on given bipolar scales, with dissimilarity judgments and acoustical analysis, yielding a 5-dimensional space. Zacharakis et al. (2014) investigated timbre description across languages—Greek- and English-speaking participants selected descriptor words from a predefined vocabulary list of 30 words, and results from both languages were reducible to three highly similar dimensions.

These approaches differ from ours in that they are focused on timbre as a property practically separable from pitch and loudness; particularly in recent studies, stimuli are typically carefully controlled and consist of a single pitch. The current study considers a view of timbre in which timbre interacts with pitch and loudness (see Siedenburg and McAdams, 2017), but it also goes beyond timbre as such to include the broader concept of qualia, that is, of the phenomenological experience of sound that may extend beyond acoustical correlates and include cultural, affective, or other types of factors. Although there is variation in timbre across the ranges of instruments, we are looking for qualia that are part of the identity of the instrument—the 'clarinet-ness' of a clarinet, or what makes a clarinet a clarinet.

We began by conducting a series of interviews with musicians during which they were encouraged to describe the sounds of musical instruments as well as their experiences of those sounds. Their natural language descriptions are taken as a starting point for the construction of a model of musical instrument qualia semantics. Zachary Wallmark (2018) carried out a text corpus study of orchestration treatises, which similarly took natural language as a starting point for understanding the perceptual semantic dimensions of timbre, albeit from written text. Wallmark's work demonstrates that language about timbre in orchestration treatises is relatively systematic and consistent.

In this study, our approach to the timbre focuses on experiential 'qualia' which entails both perceptual and cognitive aspects. As with other timbre research, we aim to produce a model where the timbral experience can be characterized according to multiple intersubjectively reliable criteria.

Study 1: Open Interviews

Method

Interviews were conducted with 23 musician participants who were asked to imagine the sounds produced by 20 familiar instruments and to describe their phenomenal experiences of each of the imagined sounds. Spoken descriptions were transcribed *in situ* and the comments later analysed for content.

Participants were prompted with a name of an instrument and first asked to rate their familiarity with that instrument. They were next asked to imagine the sound of that instrument playing a single note and to rate the vividness of their imagined sound. Participants were asked to once again imagine the sound, and then to describe it. This process was repeated for 20 musical instruments. In addition, if a participant's primary instrument was not on the list, they were further asked to imagine and describe the sound of their instrument. Each participant described the instruments in a unique random order. Additional information was collected on participants' age and musical experience for both primary and secondary instruments. Most interviews ranged between 45 and 60 minutes in duration, though several interviews lasted 90 minutes or more.

Participants

Our principal aim was to find potential interviewees with considerable familiarity with different musical instruments. A reasonable assumption is that large ensemble performers, conductors, and composers are familiar with a large variety of instruments, so recruitment efforts were concentrated on professional and semi-professional large-ensemble musicians as well as conductors and composers. As a group, the 23 participants reported an average of 18.9 years of large-ensemble experience (range 7-50 years). The principal instrument for these 23 musicians included 10 different instruments (flute, oboe, alto saxophone, trumpet, violin, viola, cello, double bass, guitar, and percussion). Five participants self-identified as composers, and four others self-identified as conductors. Participants were also asked to report secondary instruments on which they had practiced regularly at any time and on which they considered themselves to be proficient. Each musician reported at least one additional instrument, and the total number of unique primary and secondary instruments reported from the group was 30.

Stimuli

One approach to investigating instrument qualia might be to have different instruments play the same passage, and then to have listeners describe the different qualia evoked. The selection of suitable stimulus passages raises a number of questions. A musical passage will convey a number of qualities independent of the timbre of the performed instrument. For example, the passage itself may be rather quiet or loud, animated or subdued, sombre or joyous, etc. The character of the passage is apt to have a marked impact on how listeners describe the associated phenomenological qualia. In addition, different performers are likely to exhibit idiosyncratic interpretations of the passage, as well as variations in tone or timbre. Even in the case of single instrument tones, there are a number of variables that can have a marked impact, such as the amount of reverberation, mode of articulation, and overall dynamic level. Furthermore, because sound characteristics can change over the range of instruments, choosing a single tone does not necessarily represent the instrument as a whole.

Enculturated listeners are likely to already have stereotypic or cliché images of typical sounds produced by different instruments. These clichés offer a useful opportunity to probe common associations or experiences evoked by these instruments. Accordingly, we resolved to have participants simply imagine the sounds of individual instruments. That is, to start, we encouraged participants to imagine single tones played by various instruments, and then to describe the phenomenal qualia these images evoke. It was specified that the participants should imagine a professional rather than an amateur sound and that they should imagine a typical or common sound rather than an unusual sound an instrument might be able to make.

It might be objected that there can be no stable qualia evoked simply by imagining an isolated sound. All real sounds exist in some sort of context. Arthur (2016), for example, has shown how harmonic context can influence qualia descriptions of listeners. The effect of context notwithstanding, the extant research also suggests that imagining sounds without regard to context can indeed produce stable descriptions with high intersubjective agreement. For example, Huron (2006) used this approach in studying the qualia evoked by isolated scale degrees.

Ideally, our study would aim to solicit qualia descriptions representing the broadest range of possible musical timbre. One way to achieve such a breadth would be to include a large number of contrasting instruments, such as the African mbira, Indonesian peng ugal, and the European hurdy-gurdy. However, since our approach relies on imagined sounds, participant familiarity with the instrument is essential.

Given the aim of familiarity, we began by assembling a list of 44 instruments that are likely to be familiar to Western musician participants. However, in anticipation of our second study (described below) we recognized that 44 instruments would be too many for our planned rating task. Accordingly, it was necessary to reduce the list of 44 instruments to 20. In selecting this subset, our aim remained to select those instruments that are most contrasting in their timbres. Rather than relying on our own judgments of dissimilarity, we recruited 15 independent judges for a short study. The recruitment criteria were the same as described for Study 1. Participants were provided the list of 44 instruments \and asked to identify the group of 20 instruments with the most contrasting timbres or tone colors.

From these judgments, an optimization algorithm was used to determine a subset of 20 (familiar Western) instruments whose timbres were deemed maximally contrasting—listed below. These 20 instruments were used as prompts in the interviews.

Alto saxophone	Kazoo
Bagpipes	Oboe
Banjo	Piano
Bass Clarinet	Piccolo
Bass Drum	Snare drum
Cymbals	Timpani
English horn	Triangle
Flute	Tuba
French horn	Vibraphone
Harp	Wood block

Results

Open interviews with 23 participants asked to describe the above listed 20 musical instruments (along with their primary instrument, if it was not included) resulted in 480 descriptions. The transcribed responses were then parsed into component terms or ideas. For example, in the case of the flute, one participant responded: “A cool tone, soft, not super focused, breath thing—attack is not very emphasized; soft, fuzzy, not super focused, not very warm…I know flute tones do have a center but I don’t often see them that way; sky blue color, cloudy; prissy; higher chakras; puffy and fluffy.” Each response was parsed in order to identify possible component ideas. For example, the above response was deemed to convey thirteen ideas: cool, soft, not focused, breath, fuzzy, not warm, no center, sky blue, cloudy, prissy, higher chakras, puffy, fluffy. Antonyms were not assumed; for example, the above analysis recorded ‘cool’ and ‘not warm’ as separate component ideas.

Personal associations and references to specific musical works were excluded from the content analysis. The modifiers “more,” “most,” “many,” “few,” “less,” “least,” “somewhat,” “not,” “a lot of,” and “very” were also discarded from the descriptions for this stage of analysis. After these exclusions, the list of component ideas across all 20 instruments was reduced to 4,276 items. Elimination of duplicate descriptors reduced this number of unique ideas to 2,487 items; however, this number was still too high for manual sorting. An automated sort routine revealed that 502 items were mentioned more than once. All 502 items were printed on individual slips of paper, which were used in an ensuing pile sort task (De Munck, 2009).

The pile sort task was conducted independently by both authors. No number of categories was pre-established; rather, as many categories were formed as were felt to be necessary to form coherent groups of ideas. Independently, the authors assembled the content into 59 and 70 categories. Following the sorting, the authors met to discuss and reconcile their lists of categories. During this process, they referred back to the original list of terms and took into account the number of times a given term appeared in the transcripts across participants. This process resulted in the distillation of the 502 original pile-sort categories into 75 descriptive categories. In naming the final resolved categories, close attention was paid to the frequency of occurrence of different terms. Hence, for example, in the category, “gentle/calm,” the word “gentle” occurred most frequently, followed by the word “calm.” Other terms deemed to belong to the same category, such as “relaxing” or “peaceful,” occurred less frequently.

Recall that each of the words used in the pile sort task occurred a minimum of twice in the qualia transcripts. As a further check of validity and comprehensiveness, the authors went back to the larger list of 2,487 unique ideas and revisited those 1,985 words that had occurred only once in the transcripts. Each of these words was assigned to the most appropriate category. Using this procedure, they found that nearly all of the words could be reasonably well accommodated within the 75-fold classification taxonomy. Of the remaining unclassified words, the experimenters agreed that most of these orphan words could be included by adding two more categories. Accordingly, our final taxonomy consists of 77 categories:

Aggressive	Piercing/Sharp
Airy/Breathy	Pinched/Constrained
Beautiful	Ping/Ding/Ting
Big	Powerful
Brassy	Precise/Clean
Bright	Pure
Brilliant	Quick Decay
Buzzy	Raspy/Guttural
Clear	Reedy
Colorful	Resonant/Vibrant
Commanding/Assertive	Rich/Complex
Cool/Cold	Ringing/Long Decay
Cute/Innocent	Rough
Dark	Round
Deep	Rumbling/Booming
Direct/Projecting	Sad/Melancholy
Dramatic/Expressive	Salient/Present
Focused/Compact	Serious/Solemn
Folk-like/Pastoral	Shrill/Harsh/Annoying
Full	Simple
Funny/Comical	Singing/Voice-like
Gentle/Calm	Soaring/Floating
Grainy/Gravelly	Soft/Smooth
Happy/Joyful	Sparkling/Shimmering
Heavy	Supportive/Foundational
Heroic/Noble	Sustained/Even
High	Sweet
Hollow	Thick/Fat
Light (in weight)	Thin/Narrow
Loud	Twangy
Low	Unclear/Indistinct
Metallic	Unique/Distinct
Mournful/Wailing	Versatile/Flexible
Muted/Veiled	Visceral
Mysterious/Ethereal	Warm
Nasal	Watery/Fluid
Noisy	Wavy/Undulating
Open	Woody
Percussive	

Study 2: Rating Task

Method

Study 2 aimed to measure the validity and reliability of the 77 descriptive themes in characterizing the sounds of musical instruments. Participants completed a survey where each of 20 instruments was judged according to all 77 descriptors.

Given the 20 instruments used in Study 1 and 77 descriptors, the number of ratings amounts to 1,540, a task which takes two to three hours to complete. In order to reduce the task length, ratings were distributed across multiple participants. Any given participant judged a minimum of two instruments for all 77 descriptors. No participant rated the same two instruments as another participant, so there were overlapping ratings across all instrument pairs. Participants were given the option to rate as many instruments as they liked.

The survey was conducted using the Internet and Qualtrics platform. As in the interviews, participants were first asked to rate their familiarity with the instrument, to imagine a single tone and rate the vividness, and then to imagine the tone again before beginning the rating process. Though participants rated a variable number of instruments, each instrument was always rated on all 77 dimensions. Ratings were on a 7-point Likert scale ranging from "not _____" to "very _____". No bipolar scales were used. Additional background information was collected, including age, primary instrument or voice type, and self-reported musical sophistication.

Participants

Participants were recruited across the Internet as well as from The Ohio State University School of Music subject pool. Subject pool participants are second year undergraduate students pursuing a music major. Although the experiment was implemented as a web-based survey, subject pool participants were nevertheless tested individual in an Industrial Acoustics Corporation sound attenuation room.

The average number of years in which participants reported having engaged in regular musical practice was 18.4 (n=101, range 2-60 years, SD=10.8).

Stimuli

Once again, we elected to rely on participants' imaginations rather than using recorded instrumental excerpts. See "Stimuli" under Study 1.

Preliminary Results

Qualia were sorted in order of average per-instrument variance. Low variance means that participants showed greater agreement in rating instruments according to the target qualia. For example, "reedy" has the lowest per-instrument variance, at 1.08 (recall that these descriptors were rated on a scale from 1 to 7). Thus, participants were strongly in agreement about which instruments did/did not sound "reedy." Descriptors with the next-lowest variances include "low" (1.72), "percussive" (1.83), "brassy" (1.87) and "rumbling/booming" (1.90). Conversely, high variance was observed in response to the word "visceral" (3.81), demonstrating that participants had relatively little agreement as to which instruments are/are not visceral. Descriptors with the next highest variances include "Versatile/Flexible" (3.22), "Quick decay" (3.16), Ringing/Long (3.12), and Hollow (3.01).

Table 1 below compares instruments that received the highest average rating on a given descriptor with those that received the lowest average rating. For example, of all 20 instruments, crash cymbals received the highest average rating on "aggressive," while on average, the harp was rated the least "aggressive." Of the 20 instruments, the alto saxophone is the single instrument to not be included in this list; that is, it was ranked on average neither highest nor lowest on any of the 77 terms.

Table 1. The highest and lowest average rated instruments for each descriptive category are given below.

Descriptive Category	Highest Rated Instrument	Lowest Rated Instrument
Aggressive	Cymbals	Harp
Airy/Breathy	Piccolo	Bass Drum
Beautiful	Harp	Kazoo
Big	Bass Drum	Triangle
Brassy	French horn	Harp
Bright	Triangle	Bass drum
Brilliant	Triangle	Bass drum
Buzzy	Kazoo	Piccolo
Clear	Triangle	Kazoo
Colorful	English horn	Bass drum
Commanding/ Assertive	Snare drum	Harp
Cool/Cold	Snare drum	Harp
Cute/Innocent	Kazoo	Bass drum
Dark	Bass clarinet	Triangle
Deep	Bass drum	Triangle
Direct/Projecting	Snare drum	Harp
Dramatic/Expressive	English horn	Wood block
Focused/Compact	Wood block	Tuba
Folk-like/Pastoral	Bagpipes	Bass drum
Full	Tuba	Triangle
Funny/Comical	Kazoo	Harp
Gentle/Calm	Harp	Cymbals
Grainy/Gravelly	Kazoo	Harp
Happy/Joyful	Banjo	Bass drum
Heavy	Bass drum	Triangle
Heroic/Noble	French horn	Kazoo
High	Piccolo	Bass drum
Hollow	Wood block	Triangle
Light (in weight)	Triangle	Bass drum
Loud	Cymbals	Harp
Low	Bass drum	Piccolo
Metallic	Triangle	Wood block
Mournful/Wailing	English horn	Wood block
Muted/Veiled	English horn	Bagpipes
Mysterious/Ethereal	Harp	Kazoo
Nasal	Kazoo	Harp
Noisy	Cymbals	Harp
Open	Timpani	Kazoo
Percussive	Snare drum	English horn
Piercing/Sharp	Piccolo	Bass drum
Pinched/Constrained	Kazoo	Bass drum
Ping/Ding/Ting	Triangle	English horn
Powerful	Timpani	Harp
Precise/Clean	Triangle	Kazoo
Pure	Harp	Kazoo
Quick Decay	Snare drum	Bagpipes
Raspy/Guttural	Kazoo	Harp

Reedy	Oboe	Bass drum
Resonant/Vibrant	Vibraphone	Wood block
Rich/Complex	English horn	Wood block
Ringing/Long Decay	Cymbals	Wood block
Rough	Kazoo	Harp
Round	French horn	Triangle
Rumbling/Booming	Bass drum	Flute, Piccolo
Sad/Melancholy	English horn	Wood block
Salient/Present	Piccolo	Banjo
Serious/Solemn	English horn	Kazoo
Shrill/Harsh/Annoying	Kazoo	Harp
Simple	Wood block	Bagpipes
Singing/Voice-like	English horn	Wood block
Soaring/Floating	French horn	Snare drum
Soft/Smooth	Vibraphone	Cymbals
Sparkling/Shimmering	Triangle	Bass drum
Supportive/ Foundational	Tuba	Kazoo
Sustained/Even	French horn	Wood block
Sweet	Harp	Bass drum
Thick/Fat	Bass drum	Triangle
Thin/Narrow	Piccolo	Bass drum
Twangy	Banjo	French horn
Unclear/Indistinct	Bass drum	Piccolo
Unique/Distinct	Bagpipes	Bass drum
Versatile/Flexible	Piano	Cymbals
Visceral	Bass drum	Triangle
Warm	French horn	Snare drum
Watery/Fluid	Flute	Wood block
Wavy/Undulating	Vibraphone	Wood block
Woody	Wood block	Cymbals

Notice in Table 1 the predominance of the bass drum, harp, kazoo, triangle, wood block, English horn, piccolo, and cymbals, suggesting that these instruments exhibit many of the most extreme timbral characteristics in our group of 20 contrasting familiar instruments.

Also note that the highest average descriptor ratings offer useful single-word portrayals identifying what might be the most salient or characteristic timbral qualia for each instrument. As reported in Table 2, these descriptions amount to near clichés. On one hand, characterizing the alto saxophone as reedy, the kazoo as buzzy, the wood block as percussive, or the banjo as twangy seems obvious and uninformative. On the other hand, their very obviousness lends credence to the methodology. Perhaps more problematic are descriptions of the English horn, harp, and flute as "beautiful." At first this characterization might seem fatuous, especially if the term "beauty" is interpreted from the broad perspective of philosophical aesthetics. However, if the term is understood in an everyday sense, it connotes a sort of simple gentle sweetness that does appear appropriate for these instruments.

Table 2. Highest rated descriptor, on average, for each of the 20 musical instruments. Descriptors were rated on a scale from 1, 'not' to 7, 'very.'

Instrument	**Descriptor**	**Average**
Alto saxophone	Reedy	5.90
Bagpipes	Unique/Distinct	6.75
Banjo	Twangy	6.52
Bass clarinet	Low	6.44
Bass drum	Rumbling/Booming	6.76
Crash cymbals	Loud	6.77
English horn	Beautiful	6.11
Flute	Beautiful	5.81
French horn	Heroic/Noble	6.47
Harp	Beautiful	6.32
Kazoo	Buzzy	6.94
Oboe	Unique/Distinct	6.55
Piano	Beautiful	5.73
Piccolo	High	6.88
Snare drum	Percussive	6.86
Timpani	Powerful	6.78
Triangle	Ping/Ding/Ting	6.86
Tuba	Big	6.43
Vibraphone	Resonant/Vibrant	6.45
Wood block	Percussive	6.82

Preliminary Principal Components Analysis

One might expect that the 77 characteristics identified in Study 1 would exhibit considerable shared variance. The ratings assembled from Study 2 provide an opportunity to conduct Principle Components Analysis (PCA). PCA is typically used to find the least number of dimensions that nevertheless account for most of the variance. In the current study, our aim is not to reduce the characterization of timbre to a small number of dimensions. Indeed, we expect timbre qualia to entail a large number of dimensions. Instead, our aim is to identify all of the dimensions that can reliably account for any of the unshared variance. By way of illustration, consider the possible timbre qualia "cute." Only a handful of sounds or instruments (such as the sopranino recorder or the ocarina) might be reasonably described as "cute." Nevertheless, the fact that several sounds can be reliably characterized as "cute" suggests that cuteness may warrant inclusion as a *bona fide* timbre qualia, although it may account for only a very small amount of variance when a large number of instruments are considered.

Accordingly, in employing PCA, our aim was to maximize rather than minimize the number of dimensions. Given our current sample size, an 11-dimensional model is the largest that can be validly computed. Note that further data collection may allow us to consider models with additional dimensions. Planned future data collection notwithstanding, our current 11-dimensional PCA solution is in Table 3. The model distinguishes the following 11 factors: (1) low/deep, (2) singing/beautiful, (3) commanding/powerful, (4) buzzy/grainy, (5) focused/precise, (6) funny/cute, (7) metallic/ringing, (8) percussive, (9) hollow, (10) sad/mournful, and (11) brassy.

Table 3. An 11-factor model of timbre qualia. The second column identifies the experimenter-designated label for the corresponding factor. The third column lists pertinent descriptors in order of their given loadings. High negative loadings are also reported for factors 1, 7, 8, and 11.

	Provisional Label	**Descriptors Included**
1	Low/Deep	low, deep, dark, rumbling/booming, heavy, thick/fat, full, muted/veiled, round, serious/solemn, supportive/foundational, unclear/indistinct, warm
		negative loadings bright, brilliant, high, light (in weight), piercing/sharp, shrill/harsh/annoying, sparkling/shimmering, thin/narrow

2	Singing/Beautiful	beautiful, singing, versatile, colorful, dramatic/expressive, gentle/calm, mysterious/ethereal, rich/complex, soaring/floating, soft/smooth, sweet, watery/fluid, wavy/undulating
3	Commanding/ Powerful	commanding/assertive, powerful, aggressive, big, direct/projecting, heroic/noble, loud, noisy, salient/present, visceral
4	Buzzy/Grainy	buzzy, grainy/gravelly, nasal, pinched-constrained, raspy/guttural, reedy, rough, twangy
5	Focused/Precise	focused/compact, precise/clean, clear, pure
6	Funny/Cute	funny/comical, cute/innocent, folk-like/pastoral, happy/joyful, simple
7	Metallic/Ringing	ringing/long, metallic, resonant/vibrant
		negative loadings quick decay, woody
8	Percussive	percussive, ping/ding/ting
		negative loadings airy/breathy, open, sustained/even
9	Hollow	hollow, cool/cold
10	Sad/Mournful	mournful/wailing, sad/melancholy
11	Brassy	brassy
		negative loadings unique/distinct

Discussion

Most of the factors seem to capture perceptual and timbral ideas. Many seem particularly concerned with extrapolating potential features of a sound source that likely has agency (factors 1, 2, and 3) while others demonstrate a preoccupation with the quality of physical material comprising the sound source (factors 4, 5, 7, 8, 9, 11). Factors 6 and 10 stand apart, as they are affective in nature.

Following further analysis, a third study will be run in order to determine how the model from Study 2 may be applied to a more diverse set of instruments. The task will be set up similarly to that of Study 2, but participants will rate instruments only on the set of dimensions as determined by the final model. Rather than rating a set of 20 instruments, participants will be asked to rate a larger set of instruments. This task will produce a set of perceptual, semantic profiles for each of the instruments rated, data which can be used in future studies on timbre perception and orchestration.

References

Arthur, C. (2016). *When the leading tone doesn't lead: Musical qualia in context.* The Ohio State University.

De Munck, V. C. (2009). *Research design and methods for studying cultures*. Rowman Altamira.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT press.

Elliott, T. M., Hamilton, L. S., & Theunissen, F. E. (2013). Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America, 133*(1), 389–404. https://doi.org/10.1121/1.4770244

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *The Journal of the Acoustical Society of America, 61*(5), 1270-1277.

Kendall, R. A., & Carterette, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, 8(4), 369–404.

Kendall, R. A., & Carterette, E. C. (1993). Verbal attributes of simultaneous wind instrument timbres: II. Adjectives induced from Piston's Orchestration. *Music Perception, 10*(4), 469–501.

Kendall, R.A., Carterette, E.C., & Hajda, J.M. (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Perception*, *16*(3), 327-363.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *The Journal of the Acoustical Society of America, 114*(5), 2946-2957.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes. *Psychological Research*, *58*, 177-192.

Pratt, R. L., & Doak, P. E. (1976). A subjective rating scale for timbre. *Journal of Sound and Vibration*, *45*(3), 317–328.

Siedenburg, K., & McAdams, S. (2017). Four distinctions for the auditory "wastebasket" of timbre. *Frontiers in Psychology, 8.* https://doi.org/10.3389/fpsyg.2017.01747

Wallmark, Z. (2014). *Appraising Timbre: Embodiment and Affect at the Threshold of Music and Noise.* University of California Los Angeles, Los Angeles. Retrieved from https://cloudfront.escholarship.org/dist/prd/content/qt99t2t939/qt99t2t939.pdf

Wallmark, Z. (2018). A corpus analysis of timbre semantics in orchestration treatises. *Psychology of Music, 21*, 1-21.

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2014). An interlanguage study of musical timbre semantic dimensions and their acoustic correlates. *Music Perception: An Interdisciplinary Journal, 31*(4), 339–358. https://doi.org/10.1525/mp.2014.31.4.339

Zacharakis, A., Pastiadis, K., & Reiss, J. D. (2015). An interlanguage unification of musical timbre: Bridging semantic, perceptual, and acoustic dimensions. *Music Perception: An Interdisciplinary Journal,* *32*(4), 394–412. https://doi.org/10.1525/mp.2015.32.4.394

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Musical Affect and Embodiment: Fear, Threat, and Danger in the Music of *The Lord of the Rings*

Lindsey Reymore

School of Music, The Ohio State University, The United States

reymore.1@osu.edu

Abstract

Research in music perception suggests various ways in which music might portray, express, or evoke fear and threat (e.g. Huron, 2015, Juslin & Laukka, 2003), but how closely do these findings reflect musical practice? Do composers actually use these techniques when aiming to express fear and threat?

In J.R.R. Tolkien's *The Lord of the Rings*, the protagonists have much to fear along their quest, including the Nazgûl, or wraiths, who are determined to reclaim the Ring for the evil Sauron. The Nazgûl are the narrative's central symbol of fear; Tolkien writes that "their chief weapon was terror" (*Unfinished Tales*). Peter Jackson's three-film adaptation brings the threat of the Nazgûl to life not only visually, but also musically, through composer Howard Shore's soundtrack. The music accompanying the evil Nazgûl is tasked with the illustration of fear and threat and thus provides a rich opportunity to explore in musical practice principles observed by empirical literature.

First, recent research in music perception, speech prosody, and animal ethology was reviewed to create a list of musical techniques that might communicate fear and threat. Studies in music performance and speech prosody suggest that acoustic features which communicate fear include fast tempo, variability in volume, high pitch, rising pitch contour, narrow pitch range, and microstructural irregularities (e.g. Juslin & Laukka, 2003). Techniques theorized to express threat or aggression include scream-like and non-linear sounds, low pitch, loud volume, falling pitch contour, and approaching sounds (e.g. Arnal et al., 2015, Bach et al., 2009).

Musical analyses of the soundtrack accompanying the Nazgûl demonstrate abundant use of these and other factors. This music is compared to music accompanying other, non-fear-centered scenes; far more of the proposed cues are employed in fear scenes than their length-matched non-fear counterparts, supporting a probabilistic model for fear-related affective musical cues. The analyses demonstrate high consistency between the ways in which recent perceptual research suggests that fear and threat are expressed through music and fear-related music in the context of the soundtrack to *The Fellowship of the Ring*.

Introduction

Four small hobbits bearing the nefarious Ring of Power have plenty to fear along their journey, from mythological monsters and hideous Orcs to betrayal and treachery. In J.R.R. Tolkien's trilogy of books, *The Lord of the Rings*, one omnipresent and continuous threat pursues the hobbits as they seek to destroy the Ring of power: the Nazgûl, or Ringwraiths.

The Nazgûl pose one of the most direct threats to the protagonists' central mission, a cause enough for fear, but they are also perhaps the most central and clear symbol of fear in the narrative. Tolkien, the creator of Middle Earth and the author of the original novels, describes how the Nazgûl instill fear in *The Return of the King*:

> "The Nazgûl came again…like vultures that expect their fill of doomed men's flesh. Out of sight and shot they flew, and yet were ever present, and their deadly voices rent the air. More unbearable they became, not less, at each new cry. At length even the stout-hearted would fling themselves to the ground as the hidden menace passed over them, or they would stand, letting their weapons fall from nerveless hands while into their minds a blackness came, and they thought no more of war, but only of hiding and crawling and death."

Tolkien's language emphasizes a direct relationship between the fear fomented by the Nazgûl and sound: the "deadly voices" becoming "more unbearable…at each new cry."

In Peter Jackson's film adaptation of *The Lord of the Rings*, the addition of a soundtrack to the narrative experience provided an artistic opportunity to explore the Nazgûl's sounds, the sounds of fear and threat, through the musical score. In crafting the music of the Nazgûl, composer Howard Shore employed various innovative techniques to represent terror and threat. Considered in light of empirical work on the sounds of fear, many of these techniques may garner their effectiveness from human, physiological, and embodied experiences of fear.

Literature Overview

In reviewing the empirical literature on the acoustic features of fear-related sounds, two broad categories emerge. The first encompasses ways in which sound—and by extension, music—might express or communicate fear, primarily by mimicking the human voice as affected by physiological changes associated with fear; when adapted to music, these could be considered to express or illustrate fear. The second category includes the sounds of threat and aggression, which might be achieved by mimicking the acoustic properties of potential threats in nature, the musical version of which could be considered to express or illustrate threat. Screams and other non-linear sounds in particular inhabit both categories, able to indicate fear as well as threat. More generally, if music can evoke fear, anxiety, or tension, techniques from either category might be employed to do so, directly by mimicking threat—the listener hears a threat and feels afraid—or indirectly, by mimicking vocal expression of fear—the listener hears a human expression of fear, which acts as emotional contagion; if someone near us is afraid, their fear alerts us to a potential threat, and thus we have good adaptive reason to feel fear.

Fear

Despite disagreement about whether fear as such is evoked by music (Zentner, et al., 2008), fear has been one of the most commonly studied music-related emotions; in a literature review on musical expression of emotion, Juslin and Laukka (2003) identified five categories of emotion that have been researched most often in both musical and vocal expression: anger, fear, happiness, sadness, and tenderness.

Acoustic cues may give us information about the emotional state of the individual producing a sound. Spencer's Law posits a relationship among physiology, voice, and emotional expression: emotions influence physiology, and these consequent physiological changes are evident in speech, singing, and by extension, instrumental music. For example, fear is associated with increased adrenaline and peripheral acetylcholine. The flood of these hormones increases muscle reactivity so that muscles can react extra-quickly in response to danger—this causes physical trembling, which can be heard in the voice (Huron, 2015).

The general trends Juslin and Laukka observe in studies of fear expression in speech and music offer several possibilities for consideration in analysis. They conclude that studies in vocal expression and music performance suggest that fear is expressed and communicated by fast tempo, low sound level (except in panic fear), much sound level variability, little high-frequency energy, high pitch level, little pitch variability, rising pitch contour, and microstructural irregularity. Many of these factors align with what might be expected of the results of adrenaline on the voice, most clearly faster tempo and higher/rising pitch (resulting from increased tension in the vocal cords).

Screaming and Other Non-Linear Sounds

Another clear-cut expression of fear is the scream, from which we might also extrapolate expressive features to music. The scream may have several different survival benefits and may communicate to the attacker, outside observers, or both. Most directly, it may serve as a means of scaring or intimidating an attacker and could possibly increase the screamer's own readiness for action through arousal. A scream may warn kin that danger is near, giving near relations a better chance of survival—an indirect evolutionary advantage. Finally, in some species as in humans, the scream may catch the attention of sympathetic others, who may come to help fend off the attacker or rescued the attacked.

The scream is a sound that expresses fear; however, it is also a sound that is frequently referred to as evoking fear—think of the "blood-curdling" scream. The scream can be used as an aggressive tactic, to instill fear, which is the effect of the screams of the Ringwraiths in *The Lord of the Rings*. If you prefer a less fictional example, consider fighting scenarios in which the aggressors scream or yell, as when charging into battle.

The scream is one example of a broader category of sounds that may contribute to creating a musical atmosphere of fear: non-linear sounds. Sounds are considered non-linear when output surpasses a system's limits—the sound is too loud for the system's normal range, like a scream. Blumstein, Davitian, and Kaye (2010) observe that these types of sounds are commonly produced when animals are under stress, and they found evidence supporting the hypothesis that film soundtracks make use of nonlinear sounds to manipulate affective response.

The authors suggest several ways in which composers can make use of musical techniques that produce or mimic nonlinear sounds. Natural possibilities include overblowing brass and woodwind instruments, stopped French horn, and certain combinations of bow location and strength, as well as the use of inharmonic percussion instruments including some gongs and cymbals. Additionally, they suggest several composition techniques that mimic nonlinear sounds, including "frequency-based effects…such as the use of harmonic dissonance, trills, vibrato and sudden pitch change, and amplitude-based effects, such as tremolo string bowing, flutter-tonguing wind instruments, or sudden amplitude change…"

Threat and Aggression

The acoustic features considered thus far have exemplified ways to express or communicate the emotion of fear directly, a situation in which music may ape the physiological effects of fear on speech for expressive purposes. It is possible that music may also evoke fear or anxiety (or at the least, communicate that characters have reason to feel fear or anxiety) specifically by expressing threat or aggression. That is, music might create a sense of fear, or at least, tension, by employing acoustic landmarks of threatening or dangerous sounds.

The sound-size model plays an integral role in this approach: in evolutionary history, we would expect to find threat coming particularly from larger animals. Thus, relative appearance of size for an animal may communicate information. In the animal kingdom, appearing larger is associated with threat, while appearing smaller is associated with submission. These indications of threat and submission extend beyond the visual modality, as observed by Eugene Morton; larger masses vibrate at lower volume. Thus, relatively low-pitched calls communicate threat or aggression while relatively high-pitched vocalizations communicate affinity or submission (as cited in Huron, 2015).

Other sound features that have been suggested to communicate threat and aggression include falling contour, the clarity of a sound's pitch (likely related to the class of non-linear sounds), and loudness (Huron, Kinney, and Precoda, 2006), as well as drone tones (Trevor, 2017, 2018). Huron (2012) has also described the contribution of extensity of sound sources, infrasound, and surprising sounds to the communication of threat. Some evidence has been found suggesting that music can create a sense of safety (Schafer, et al., 2015); the withdrawal of music may consequently undermine a sense of safety.

Another set of features of sound that would seem to be important to attend to for detecting threat and increasing survival are features that tell us an object is approaching. Dominik Bach, John Neuhoff, Walter Perrig, and Erich Seifritz (2009) found that compared to receding aural cues, approaching aural cues were rated as more unpleasant, potent, arousing, and intense, and participants rated as higher the probability that these sounds signified a threat. Extrapolating these results to musical contexts suggests the hypothesis that the use of crescendos can contribute to a sense of threat and danger and may contribute to evoking fear; I offer based on this research that crescendos with certain features would be perceived as more threatening than others, including faster rates of crescendo, larger dynamic range spanned, and overall loudness, though this has not been tested.

Table 1. Acoustic features of fear- and threat-related sounds and their potential musical analogues.

Affect	Acoustic Feature	Corresponding Musical Cue	Citations
Fear	Fast tempo/speaking rate	Fast tempo	Juslin & Laukka, 2003; also see Schafer et al., 2015
	Low sound level (except in panic fear)	Quiet dynamics…or loud dynamics (panic fear)	Juslin & Laukka, 2003
	Lots of sound level variability	Lots of change in dynamics	Juslin & Laukka, 2003
	High pitch	High tessitura	Juslin & Laukka, 2003
	Little pitch variability	Narrow pitch range	Juslin & Laukka, 2003
	Rising pitch contour	Rising pitch/melodic contour	Juslin & Laukka, 2003
	Microstructural irregularity	Vibrato; potential overlap with the category of non-linear sounds	Juslin & Laukka, 2003; Huron, 2015
Fear and/or Threat	Screaming	Screaming/scream-like sounds	Arnal et al. 2015; Belin & Zatorre, 2015; Trevor, 2017
	Non-linear sounds	Overblowing, stopped horn, bowing techniques, inharmonic percussion, dissonance, trills, vibrato, sudden pitch change, tremolo string bowing, flutter-tonguing, sudden dynamic change	Blumstein, Bryant & Kaye, 2012; Blumstein, Davitian & Kaye, 2010
Threat	Falling pitch contour	Falling pitch/melodic contour	Juslin & Laukka, 2003
	Low pitch (and infrasound)	Low tessitura	Shanahan & Huron, 2014; Huron, Kinney & Precoda, 2006, Huron, 2012
	Loudness	Loud dynamics	Huron, Kinney & Precoda, 2006
	"Volume," or extensity	Relatively many sound sources	Huron, 2012
	Approaching sounds	Crescendos, especially quick crescendos and those with wide dynamic range	Bach et al., 2009
	(Lack of predictability?)	Drone tones	Trevor, 2017, 2018
	Surprise	Rapid increase in loudness, abrupt change in tempo, new/unprepared harmony, unexpected modulation, sudden change of texture	Huron, 2012
	Silence	Silence, or withdrawal of music	Schafer et al., 2015

Probabilistic Cues

Table 1 is not intended to suggest that any music portraying fear will use all of these cues, or that use of any one of these cues is a necessary cue for fear. For example, I will point to the use of large choir in the Nazgûl music as contributing to the threatening effect of the music—many voices present high extensity, or number of sound sources, one of the signifiers addressed above as a cue for threat.

However, composer Shore also employs choir frequently in the music of the elves, who are not semantically connected to terror or threat at all, but in contrast, to peace and beauty. So how can I reasonably claim extensity via choir as a cue for fear? Furthermore, many of the cues listed in Table 1 might seem conflicting or just too broad to be useful. For example, rising pitch contour is associated with fear expression, but descending pitch contour is associated with threat! A similar problem arises when considering tessitura, where high pitch is expressive of fear while low pitch is expressive of threat, and when considering loudness, where low sound level can be fearful, high sound level can be aggressive, and sound level variability is also claimed as a cue.

Juslin and Laukka (2003) suggest using Brunswick's lens model, an early model of visual perception, to consider the communicative process in vocal and musical expression and address these types of concerns. Essential to this analogy is that relevant cues for vocal and musical communication are encoded probabilistically. Encoders use a large set of these probabilistic and partially redundant cues, and decoders recognize and use these same cues.

> "The cues are probabilistic in that they are not perfectly reliable indicators of the expressed emotion. Therefore, decoders have to combine many cues for successful communication to occur…each cue is neither necessary nor sufficient, but the larger the number of cues used, the more reliable the communication" (Juslin and Laukka, 2003).

Many different cues can convey the same information; cues are often redundant. Though this is good for the robustness of the system's communicative power, it limits the amount of information that can be shared. Juslin and Laukka suggest that this may be why broader categories of emotion can be efficiently communicated while the finer nuances are not.

The following analyses of the Nazgûl music from *The Lord of the Rings* assume the analogy provided by this model. I do not expect that musical techniques emulating cues of fear and threat in the Nazgûl music will be exclusive to that musical material. However, I do expect I will easily be able to identify a large number of cues in fear-related music. Furthermore, I expect that the music accompanying the scenes of the Nazgûl will contain relatively far more fear cues than that of scenes centered on different, especially positive, affects.

Musical Analysis

The scope of this paper is restricted to scenes from *The Fellowship of the Ring*, the first movie in the three-part trilogy, in order to be able to focus more deeply and capture more detail about the film and soundtrack. Of the three movies, the Ringwraiths are most involved in the first and have relatively less screen time in both the second and third movies. For each of the scenes in *The Fellowship of the Ring* involving the Nazgûl, I noted the use and context of any of the musical techniques described in Table 1. Below, I describe one such scene; cues from Table 1 are marked in bold.

Weathertop

Led by Aragorn away from the town of Bree, the hobbits enjoy some time in relative safety—until the Nazgûl catch up to them once again at Weathertop. A long, low crescendo (**approaching sounds**) accompanies a wide shot from above of the Riders approaching the camp. The crescendo grows underneath a **fast, rising** sixteenth-note figure as the hobbits find the highest ground and prepare to defend themselves. A **descending** third motive enters in the **low range**, building tension, as the violins begin a slow, unsettling *portamento* (**rising pitch contour**) in open fifths up a twelfth, throughout the course of two and a half measures. Tremolo bowing is added to the *portamento*, creating another disturbing extended technique probably exemplifying **microstructural irregularity and non-linearity**. The **descending** 'Threat of Mordor' motive enters in the bass, overlayed with a quieter but very intense choral orchestration (**sound extensity**) of the Nazgûl theme as the wraiths surround the hobbits.

At the moment swords begin to clash, the violins enter on a shrill, high Bb, which develops into a highly dissonant A-Bb-C cluster **drone**. There is a longer prominent crescendo as the wraiths throw Merry and Pippin aside and Frodo stumbles back, falling to the ground. As the Nazgûl approach, the various threatening Mordor motives combine with trembling aleatoric string and wind figures (**microstructural irregularity**). The violin tone cluster enters again, now composed of five or more chromatically-neighboring pitches, creating a striking **high-pitched** dissonance that sustains throughout the approach of the Nazgûl, cutting off at the moment Frodo, cornered with no other defense, yields to the power of the Ring and puts it on his finger, which draws him into the warped, distorted wraith world. Frodo's experience of the wraith world is accompanied only by **harsh, non-linear sound effects**. One of the Ringwraiths stabs Frodo in the shoulder.

Just in time, the clear horns return the music to the scene (**music signifying safety**), proclaiming Strider's heroic theme as he swoops down wielding a flaming torch to save the hobbits. Once the wraiths are driven off, low string and brass punctuate the scene with brief, ominous crescendos (**approaching sounds**) as the party recognizes the severity of Frodo's stab wound.

Table 2. Weathertop musical analysis. The second column notes the particular manifestation of the cue.

Musical Cue	Use in Excerpt
Fast tempo/speaking rate	half=88
Low sound level (except in panic fear)	low sound level while anticipating threat
Lots of sound level variability	lots of crescendo and subito piano; dynamic range of excerpt is wide
High pitch	high strings/choir
Little pitch variability	X
Rising pitch contour	portamento
Microstructural irregularity/non-linear sounds	tremolo, portamento
Screaming	Ringwraith screams, Frodo screams
Falling pitch contour	"Descending Third" and "Threat of Mordor" motives
Low pitch (and infrasound)	primary orchestration low strings and brass
Loudness	loud
"Volume," or extensity	choir
Approaching sounds	frequent short, intense crescendos and several longer ones
Drone tones	high string drones
Surprise	(scream)
Silence	X

Scene Comparisons

In the final section of this paper, I return to the leading premise: music is able to communicate fear, threat, and danger by employing acoustical cues that emulate vocal expression of fear and/or mimicking the acoustical properties of potentially threatening entities. I have suggested quite a few of these cues based on previous research in music, speech prosody, animal ethology, and evolutionary psychology, but this list of cues is most likely not exhaustive. Because these acoustical cues can signify multiple affects, I am suggesting that our brains use these cues in a probabilistic way, meaning that successful communication involves the use of multiple cues that are partially redundant in the information they communicate.

A priori, I predicted I would be able to identify many of the cues derived from the literature review in the Nazgûl music. Based on the premise that affect cues in music are probabilistic, I predicted that music accompanying the scenes of the Nazgûl would contain relatively more types of fear cues than scenes with central affects other than fear. In the following comparison analysis, I used the list of cues presented in Table 1 to probe this last prediction. I counted the number of cue types used in three scenes involving the Nazgûl and three other non-scary scenes, which were randomly selected and matched to the Nazgûl scenes for length. This strategy allows a quantification of the use of these cues in the movie music.

To choose scenes for comparison, I used a random number generator to select scenes from *The Fellowship of the Ring*. As I was looking to compare the Nazgûl scenes to scenes without narrative threat, a priori I decided to eliminate scenes in which physical threat was present and scenes in which the Ring played a prominent role. Excerpts started with the beginning of a scene as indicated on the DVD version, and I determined an appropriate place to end the excerpt based on matching lengths with the Nazgûl clips.

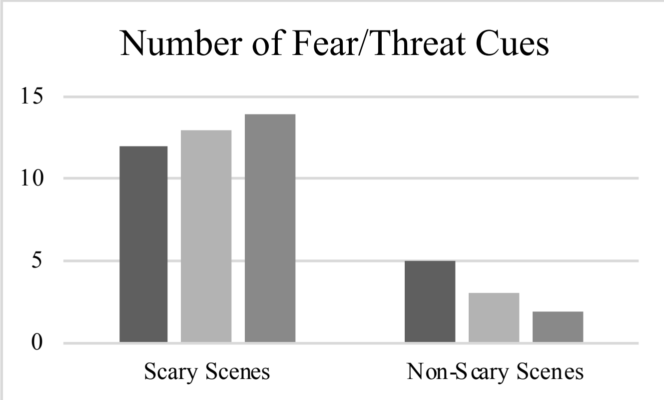
The table below lists the number of fear-related musical techniques in 'scary' and 'non-scary' scenes. The possible cues are taken from Table 1; the only change is that I combined "non-linear sounds" and "microstructural irregularities," into a single category for the purpose of this analysis, as I was not able to tell definitively from listening alone which techniques belong to these categories.

Tables 3a. and 3b. Number of fear/threat cues present in scary and non-scary scenes. Length of the scenes (minutes: seconds) is indicated after the scene name.

	'Scary' Scenes		
Scene Name	Weathertop (3:00)	Flight to the Ford (2:30)	Buckleberry Ferry (2:17)
# of cues identified	13	12	14

	'Non-scary' scenes		
Scene Name	Many Meetings (3:00)	The Fellowship Departs (2:30)	The Shire (2:17)
# of cues identified	3	5	2

Figure 1. Fear/threat cues. Matched bar-colorings indicate scenes matched for length.



The increased frequency with which fear/threat cues are used in scary relative to non-scary scenes supports the theory that fear and danger cues are used probabilistically in encoding and decoding music. Nazgûl scenes, or scary scenes, employ many of the musical features suggested by previous research to signify fear and danger, while non-threatening scenes use far fewer of these types of cues. When music that accompanies non-scary scenes employs the cues suggested to signify fear and danger, the music may be foreshadowing a future threat, or the cue may be employed to signify a different affect; for example, while "fast tempo" can be a cue for fear, it can also be a cue for happiness.

Conclusion

I began by reviewing the literature on acoustic cues for fear and threat and considering how these might be employed in music. Next, I considered specific ways some of these cues might be manifested as compositional techniques by analyzing the Nazgûl music from several scenes in *The Fellowship of the Ring*. This analysis identified many specific musical techniques in the Nazgûl music that recruited the acoustic cues suggested by the empirical literature. Lastly, I compared the Nazgûl music to non-scary music in other portions of the film and concluded that relative to non-scary music, the scary (Nazgûl) music employs far more of the acoustic cues suggested by empirical literature. These findings are consistent with the "lens model" theory of musical communication as submitted by Juslin and Laukka (2003), which suggests that successful communication is a product of multiple, probabilistic cues that are partially redundant.

References

Adams, Doug. (2002). *The Music of The Lord of the Rings.* Alfred Music Publishing: California.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A. L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology*, *25*(15), 2051-2056.

Bach, D. R., Neuhoff, J. G., Perrig, W., & Seifritz, E. (2009). Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology, 74*(1), 28-33.

Bartsch, A., Appel, M., & Storch, D. (2010). Predicting emotions and meta-emotions at the movies: The role of the need for affect in audiences' experience of horror and drama. *Communication Research*, *37*(2), 167-190.

Belin, P., & Zatorre, R. J. (2015). Neurobiology: Sounding the Alarm. *Current Biology*, *25*(18), R805-R806.

Blumstein, D. T., Bryant, G. A., & Kaye, P. (2012). The sound of arousal in music is context-dependent. *Biology letters*, rsbl20120374.

Blumstein, D. T., Davitian, R., & Kaye, P. D. (2010). Do film soundtracks contain nonlinear analogues to influence emotion?. *Biology letters*, *6*(6), 751-754.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18-49.

Eskine, K. J., Kacinik, N. A., & Prinz, J. J. (2012). Stirring images: fear, not happiness or arousal, makes art more sublime. *Emotion*, *12*(5), 1071.

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., ... & Newen, A. (2017, July). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. In *Proc. R. Soc. B*, *284*(1859), 20170990. The Royal Society.

Huron, D. (2015). Affect induction through musical sounds: an ethological perspective. *Phil. Trans. R. Soc. B*, *370*(1664), 20140098.

Huron, D. (2012). Music-Induced Frisson. <https://vimeo.com/34876042>. Accessed 25 November 2017.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review, 1*(3), 170-177.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological bulletin*, *129*(5), 770.

LeDoux, J. (2015). *Anxiety*. New York: Penguin Books.

M.W. (Transcriber) & Shore, Howard (Composer, orchestrator). (2014). *The Fellowship of the Ring*. <http://alcaeru.weebly.com/uploads/7/8/6/0/786082/fotr_complete_transcription.pdf>.

Osbourne, B., Jackson, P., Walsh, F., & Sanders, T. (Producers), & Jackson, P. (Director). (2001). *The Fellowship of the Ring*. New Zealand & The United States: WingNut Films & The Saul Zaentz Company.

Rosenblum, L. D., Wuestefeld, A. P., & Saldaña, H. M. (1993). Auditory looming perception: Influences on anticipatory judgments. *Perception*, *22*(12), 1467-1482.

Schäfer, Thomas, David Huron, Daniel Shanahan, and Peter Sedlmeier. The sounds of safety: stress and danger in music perception. *Frontiers in psychology,* 6 (2015).

Seifritz, E., Neuhoff, J. G., Bilecen, D., Scheffler, K., Mustovic, H., Schächinger, H., ... & Di Salle, F. (2002). Neural processing of auditory looming in the human brain. *Current Biology, 12*(24), 2147-2151.

Shanahan, D., & Huron, D. (2014). Heroes and villains: the relationship between pitch tessitura and sociability of operatic characters. *Empirical Musicology Review*, *9*(2), 141-153.

Tolkien, J.R.R. (2012). *The Lord of the Rings: One Volume*. Houghton Mifflin Harcourt.

Tolkien, J. R. R. (2012). *Unfinished Tales of Númenor and Middle-earth*. Houghton Mifflin Harcourt.

Trevor, C. "Eye Tracking Horror Music: Studying the affective information of drone tones." RMC Interdisciplinary Methods Festival, The Ohio State University, Columbus, OH (July 2017).

Trevor, C. "Signaling Fear in Horror Movie Soundtracks: The Use of Acoustic Features of Human Screams." The Society of Music Theory Annual Meeting, Music Informatics Group Poster Session, Arlington, VA (November 2017).

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, *8*(4), 494.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Absolute Memory for Pitch as a Hypothetical Cognitive Component for Tonic Retention

Arthur Rinaldi,[1] and Nayana Di Giuseppe Germano[2]

[1] *EMBAP/UNESPAR, Brazil*
[2] *Department of Music, UNESP, Brazil*

[1]art.rinaldi@gmail.com, [2]nayanager@hotmail.com

Abstract

In tonal music practice, a composer is able to increase and decrease harmonic tension locally by choosing a particular sequence of chords. Concurrently, harmonic tension may also be globally increased by changing which chord exerts the Tonic function (a process called modulation), and when original Tonic chord reassumes the Tonic function, it generates a large-scale tonal closure effect. If these principles are true, a listener should be able to recognize both the local Tonic and the main Tonic simultaneously after any modulatory passage. Many experiments were designed to provide empirical evidences regarding this theoretical proposition. Research reports are divergent, some corroborating the existence of this cognitive ability (Lerdahl & Krumhasl, 2007), some questioning it (Cook, 1987; Bigand & Parncutt, 1999; Marvin & Brinkman, 1999; Farbood, 2016). Based on a literature review, we argument in favor of the hypothesis that main tonic retention over modulatory passages in tonal music is only possible for subjects capable of retaining in memory some form of absolute pitch information, and that this ability is employed conjointly with relative memory for pitch, which provides a local hierarchical representation mapping for pitches. We also highlight that research on the abilities to encode and retrieve relative and absolute pitch information still requires quantification, and that it could provide an important background for more advanced experiments in musical contexts. We conclude this paper with a set of questions that should guide future empirical researches in scrutinizing our proposed hypothesis.

Introduction

In Western tonal music practice, harmonic progression (sequences of chords) constitutes a background for melodic expressiveness. This harmonic background relies on the principle that there are two main harmonic functions exerted by chords: Dominant (tension) and Tonic (rest). While choosing a particular sequence of chords, a composer is able to increase harmonic tension by moving to a Dominant chord and then decrease it by moving to a Tonic chord.

According to any textbook on tonal harmony, the Tonic not only operates as a point of harmonic rest, but also as a tonal center for the entire composition, providing a reference point whereof all notes/chords are perceived according to a hierarchically organized harmonic mapping. Due to this principle, in tonal music, harmonic tension can also be increased over the course of an entire composition by changing which chord exerts the Tonic function (a process called modulation). Consequently, when the original Tonic chord reassumes Tonic function, it generates a large-scale effect of tonal closure, an essential feature of tonal practice to finish a musical composition.

If the above principles are true, then, in order to comprehend adequately a tonal composition, listeners should be able to recognize both the local Tonic and the main Tonic simultaneously after any modulatory passage. Over the last decades, many experimental researches were designed to provide evidences regarding this theoretical proposition. Research reports diverge, some corroborating the existence of this cognitive ability (Lerdahl & Krumhasl, 2007), some questioning it (Cook, 1987; Bigand & Parncutt, 1999; Marvin & Brinkman, 1999; Farbood, 2016).

In this theoretical paper, we propose the hypothesis that main tonic retention over modulatory passages in tonal music is only possible for subjects capable of retaining in memory some absolute pitch information, an ability directly related to what cognitive literature defines as Absolute Pitch (Bachem, 1937; Parncutt & Levitin, 2013; Germano, 2015). In the following pages, we will argument in favor of this hypothesis and its relevance for future researches. We begin by presenting a literature review on tonic retention on modulatory passages, followed by a literature review on relative and absolute memory for pitch, followed by a discussion section dedicated to the correlations between collected information. In the end, we provide a set of questions that should guide future empirical researches in scrutinizing the proposed hypothesis.

Tonal Hierarchies and Music Cognition

The matter regarding cognitive capacity for main tonic retention over modulatory passages is important because this principle is not only used to explain how tonal music is usually composed, but it is also used to analyze tonal compositions in order to explain how they create a sense of coherence. Albeit not all music theorists explicitly claim that the ability to recognize aurally both the local tonic and the global tonic is mandatory for anyone to enjoy listening to tonal music, it is assumed that the real beauty of a tonal composition may only be truly appreciated if a person is able to aurally understand the rich web of harmonic relations woven by the composer throughout its course in time. This harmonic web is created around the Tonic, which provides a reference point whereof all notes/chords are positioned according to a hierarchically organized harmonic mapping.

Many empirical studies have tried to provide psychological evidences regarding the relevance of tonal hierarchies for a listener's aural experience. Reports show that listeners' evaluations regarding the distances between pitches, chords, and regions (or keys) from a given tonic form consistent patterns (Bharucha & Krumhansl, 1983; Krumhansl & Kessler, 1982, Krumhansl, 1990; Cuddy & Thompson, 1992; Cuddy & Smith, 2000; Krumhansl & Cuddy, 2010). However, such reports have limited relevance regarding our main concern. The majority did not include musical passages with modulations in their experiments, while the remainder were primarily concerned with the psychological process of recognizing changes of tonal center and/or with measurement of the distance (degree of tension) between the original tonic and the new tonic after a modulatory passage. Due to this, all experiments used short musical passages, limiting moreover any evaluation as per the influence of local and global Tonic in a listener's aural experience from their findings.

The main psychological evidence in favor of the influence of the main Tonic after a modulatory passage assumption comes from Lerdhal & Krumhansl (2007). Their aim was to test a theoretical formal model (Tonal Pitch Space) developed by Fred Lerdahl, which generates quantitative predictions regarding the psychological degree of tension and attraction for events in any passage of tonal music (Lerdahl, 2001). Their findings are particularly relevant because their experiments included longer musical segments, with full modulatory processes and a return to the main tonal center in the end.

From the four components included in Lerdahl's model, we will focus exclusively on prolongational structure and pitch-space model. Prolongational structure is a representation of the hierarchical event structure in a musical passage. It is based on a rule system which generates a hierarchically stratified representation of connections among key events in a musical passage (be it a musical phrase or an entire composition), pointing out changes in tension levels between them. On the other hand, pitch-space model is a representation of a listener's cognitive schema for mapping distances between pitches, chords, and tonal regions, based on his tacit long-term knowledge regarding tonal practice in general. It quantifies the distance between any chord in one tonal region and any other chord in the same or any other tonal region. Due to prolongational structure, events inherit pitch-space distances from events superordinate to them.

Lerdahl's prolongational structure is directly influenced by musical theories of pitch reduction, such as Schenker (1935), and so considers that the main Tonic exerts harmonic influence on the entire course of a composition. As the hierarchical representation of prolongational structure is used for quantifying distances in pitch-space model, it can be observed that Lerdahl's formal model considers both the effect of the local Tonic and the main Tonic while measuring an event's level of tension and relaxation. Since the theoretical quantifications provided by Tonal Pitch Space model were strongly supported by test results (Lerdahl & Krumhansl, 2007), their findings corroborate the assumption that the main Tonic still exerts influence after a modulatory passage, and that the large-scale effect of tonal closure is an essential psychological feature of tonal music (and not only a compositional convention).

Counterevidences are provided by several researchers that have adopted a different methodological perspective, conducting aural experiments focused specifically on the comparison of musical passages with and without modulation. Research reports suggest that the effects of memory retention of the main Tonic are restricted to a very short time span, having little to no influence on a subject's sense of harmonic large-scale tension and relaxation (Cook, 1987; Marvin & Brinkman, 1999; Farbood, 2016). Bigand and Parncutt (1999) report similar findings but include the observation that in one of their experiments the return to the main Tonic was experienced as increasing tension, instead of the theoretically predicted sense of relaxation. Conjointly, these findings suggest that listeners used local musical features, including stylistic features of tonal practice and harmonic progression conventions, while judging changes in tension and relaxation over the course of a composition, instead of relying on a global harmonic pitch schema.

In face of such divergent reports, it is necessary to question the exact role exerted by the main Tonic in Lerdahl's report. It may exert a smaller role regarding experienced changes in tension level than advocated by his own theory. On the other hand, it may exert a variable role, being more or less important according to the cognitive capacities of different listeners. If a tonal center provides a hierarchically organized harmonic mapping which guides a subject's perception of pitches in a musical passage, then it is essential to investigate if it is cognitively possible to overlap two distinctive harmonic mappings, one locally (a variable tonal center due to modulatory passages) and one globally (centered on the main Tonic). We believe this is only possible if a listener is able to retain in memory not only pitch information in relation to a tonal center, but also some form of absolute pitch information.

Absolute and Relative Memory for Pitch

Cognitive literature provides two basic concepts regarding pitch perception: Relative Pitch and Absolute Pitch (for a recent critical literature review on both abilities, see Germano, 2015). Relative Pitch is the ability to perceive pitches based on the intervallic distance between them. It is considered widespread, as the general population is capable of recognizing a song well-known, regardless of transposition, based on the sequence of melodic intervals that characterizes that specific song. Among musicians, Relative Pitch is presented under a stricter perspective. It is defined as the ability to consciously recognize and name different types of musical elements, such as intervals, chords, scales and harmonic progressions. Developed through years of ear training practice, it is a more refined version of general Relative Pitch. As shown in the previous section, the principle of tonal hierarchies is directly related to Relative Pitch ability, as it assumes that all pitches in a musical passage are perceived hierarchically in relation to a tonal center.

Absolute Pitch is a distinct ability, usually presented by cognitive literature in opposition to Relative Pitch. It is commonly described as a very rare cognitive ability which allows a subject to identify any note (i.e., a specific pitch-class, such as G/Sol) by name without external reference. However, contrary to common assumption, most possessors of this ability face difficulties under certain conditions (e.g., in certain timbres and registers). They are also prone to errors, mostly by a semitone. These variations among possessors still demand further evaluation in terms of quantification and prevalence. As seen in Germano (2015), this lack of referential values causes each researcher to adopt a different cut-off point for Absolute Pitch classification, making it difficult to compare their findings and conclusions as they could be measuring cognitively distinct populations under the same label.

Regardless of the issues raised, it is consensual that Absolute Pitch requires a stable (possibly long-term) representation for pitches in memory. As absolute representation is reported by animal learning investigators, it is reasonable to hypothesize that humans are also capable of maintaining in memory some kind of absolute pitch information (Levitin, 1994). There are evidences that absolute memory for pitch can be found (to some extent) on non-possessors of Absolute Pitch: not only on musically trained subjects (Terhardt & Ward, 1982; Terhardt & Seewan, 1983; Hsieh & Saberi, 2008; Schlemmer, 2009) but also on subjects with little to no musical training (Levitin, 1994; Schellenberg & Trehub, 2003; Frieler et al., 2013; Ben-Haim et al., 2014). A similar finding is also reported by Germano et al. (2018), who conducted a large experiment on pitch perception. According to the authors, results suggest that the ability to recognize isolated pitches fits better to a dimensional model (continuous) than to a categorical one, leading them to question if the traditional perspective of a division between Absolute Pitch possessors and non-possessors is entirely correct. Conjointly, such reports suggest that many individuals are able to maintain some kind of stable representation for absolute pitch information in memory, although with variable degrees of precision.

Some authors also report that certain pitches can be more accurately and rapidly recognized than others (Miyazaki, 1989; Miyazaki, 1990; Takeuchi & Hulse, 1991; Marvin & Brinkman, 2000; Ben-Haim et al., 2014). Researchers believe this could be an exposure effect caused by a relatively high frequency of recurrence of certain pitches in musical practice (such as C/Do or A/La). Contrariwise, Hsieh & Saberi (2008) report no evidences for this difference on pitch recognition.

Research on the abilities to encode and retrieve relative and absolute pitch information still needs development, mostly in terms of quantification: what is the mean value for each ability in the general population? And among musicians? What is the probability distribution for different levels of each ability on both populations? Which factors can reduce the efficacy of each ability? To what extent?

In the previous section, we asked if it is cognitively possible to overlap two distinctive harmonic mappings, one locally and one globally. We believe this is possible because listeners can access two kinds of pitch memory. For local harmonic relations, it is used relational memory for pitch, which under cultural constraints related to tonal practice produces a hierarchically organized harmonic mapping around the tonal center of a musical passage. For large-scale harmonic relations, we hypothesize that absolute memory for pitch must be used, which would allow the recognition of the same pitch in distinct musical segments with different tonal centers (i.e., regardless of tonal hierarchies).

Discussion

As seen in the previous section, there are enough reports to consider the possibility that listeners could retain in memory both relative and absolute pitch information. Although it is considered a rare ability, absolute memory for pitch may only be less developed in the general population, possibly due to a greater importance of relative pitch information for humans in general. As pointed before, exposure effect seems to ease the encoding of an absolute representation for certain pitches. If we consider that most research reports conducted experiments on music students (i.e., subjects who studied a musical instrument for years and were familiar with basic musical concepts), it is wise to question the role that previous experiences may have exerted on developing pitch representations in comparison with the general population.

Hsieh & Saberi (2008) provide an interesting report. In a comparison between Absolute Pitch possessors and non-possessors, they emphasize that both groups were not accurate on a frequency match task using an oscillator, but were extremely accurate in voicing a target pitch. The authors attributed this difference to the presence of a procedural memory, allowing a precise access to internal absolute pitch representations through the vocal-motor system. This is another example of how exposure effects could reinforce the retention of absolute pitch information.

Returning to our main issue, the capacity to retain in memory the main Tonic after a modulatory musical passage, it is important to remind that research reports are divergent. It is also relevant to point out that most experiments used subjective questions (sense of pleasure, coherence, completion and tension level), whose answers could be given due to several variables (such as level and focus of attention, level of musical knowledge, familiarity with tonal music, individual memory capacity, personal musical preferences). In order to achieve a more precise evaluation regarding this ability, future researches need to limit the number of variables and to adopt objective questions.

From a theoretical perspective, tonic retention ability seems to be problematic if a listener relies exclusively on relative pitch information. The overlapping of two harmonic mappings, one for the local Tonic and another for the main Tonic, would cause the same pitch to possess two distinct tonal hierarchies (and possibly two opposing functions, Tonic/rest and Dominant/tension) simultaneously. This a broader issue if we consider that: 1) the accuracy of relative memory for pitch is not yet quantified, so it is not reasonable to assume that all subjects are able to recognize aurally all pitches in relation to a tonal center; 2) many tonal compositions overlap more than two harmonic mappings. Faced with such a cognitive conflict, it is not surprising that listeners focus on local musical features when judging changes in tension values.

The use of absolute memory for pitch could provide a solution for this problem. As it allows the recognition of pitches in musical segments with different tonal centers (i.e., regardless of tonal hierarchies), it provides a hypothetical explanation for tonal closure. However, as this ability presents such high variability in terms of precision amongst studied subjects, it is also expected that many subjects would face difficulties while recognizing the return of the main Tonic. In fact, it could be a psychological reality for a minor percentage of the general population.

From a musical perspective, it is essential to emphasize that composers use several resources in order to induce a listener's sense of tension level, which include the tonal closure effect. All choices regarding dispositions of themes, textures, dynamics, tempo, orchestration and register are made in conjunction with the harmonic structure. Therefore, it is common practice that, when the main Tonic returns, other musical elements are also presented similarly as they were presented at the beginning of a composition. Through the repetition of a music previously heard, composers try to ease the recognition of the return to the initial tonal center. This is a very important procedure, especially if we consider that most listeners do not possess a stable long-term representation for all pitches, like traditional Absolute Pitch Possessors.

Absolute memory for pitch could also prove to be relevant for non-tonal music. Many contemporary composers (such as Paul Hindemith and Luciano Berio) repeat certain pitches to establish them as structural reference points in their musical works. If subjects are capable of retaining such pitches in memory, even after periods of time where they are not prominent, then absolute memory for pitch could also be an important cognitive resource in creating a sense of coherence on non-tonal compositions.

Lastly, it should be noted that there is a lack of studies regarding the processing and recognition of harmonic information on Absolute Pitch possessors. A better evaluation of their capacity and precision for mapping changes in tonal center (i.e., what pitch functions locally as Tonic after each modulatory passage) could provide a reference point for experiments with the general population.

Closing Arguments and Future Implications

In this paper, we propose the hypothesis that main tonic retention over modulatory passages in tonal music is only possible for subjects capable of retaining in memory some form of absolute pitch information, an ability called by cognitive literature absolute memory for pitch. We believe it allows the recognition of the same pitch in distinct musical segments with different tonal centers (i.e., regardless of tonal hierarchies), and that it is used conjointly with relative memory for pitch, which allows the representation of a hierarchically organized harmonic mapping around the tonal center of a musical passage.

In the previous sections, we provided theoretical background for our hypothesis, including a discussion section regarding its pertinence. We also highlighted that research on the abilities to encode and retrieve relative and absolute pitch information still needs quantification in terms of mean value and probability distribution. It should focus on testing subjects using basic musical elements, such as isolated pitches, music intervals and chords, and it should also try to chart which factors are able to enhance or reduce performance (e.g., age, gender, timbre, register, response time) and to what extent. This information could provide an important background for more advanced experiments in musical contexts, such as those necessary to evaluate our hypothesis.

We conclude this paper by proposing a set of objective questions that should guide future empirical researches in scrutinizing our proposed hypothesis. Some questions were designed specifically for subjects with musical knowledge, but they were mostly designed to be applicable to the general population.

- After subjects listen to a musical composition presenting modulation: 1) Are they able to sing the main Tonic? 2) Are they able to play the main Tonic in a piano keyboard? 3) Are they able to adjust an oscillator to match the main Tonic? 4) Can they recognize the main Tonic among alternatives? 5) Do subjects perform better in the singing task? If so, are they accessing a procedural memory?
- If subjects listen to two versions of the same composition presenting modulation (one is the original and returns to the main Tonic and the other is altered in order to end in the new tonal center): can subjects identify which is the original version and which is the modified one?
- Regarding the above questions: 1) Is the memorization of the main tonic easier in some compositions? If so, what are the reasons? Is it due to reinforcement by repetition of other musical elements (e.g., if it presents a similar or equal theme in comparison to the beginning of the composition)? Are the tonal centers closely related? Is it due to high pitch recurrence rate? 2) Is there any difference in performance if subjects know the question before listening to the stimuli? If they actively focus on memorizing the main Tonic, do they perform better? 3) Are some pitches easier to be recognized than others? If so, which pitches? Are they the same for most subjects? 4) Does musical training affect performance for main Tonic retention?
- Do subjects encode absolute pitch information as abstract pitch-classes, or holistically (as a combination of pitch, timbre and height)? Is there a difference among musicians and non-musicians?

Acknowledgements. We thank FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) for the provided funding.

References

Bachem, A. (1937). Various types of absolute pitch. *Journal of the Acoustical Society of America*, 9, 146-151.

Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13(1), 63-102.

Ben-Haim, M. S., et al. (2014). Pitch Memory and Exposure Effects. *Journal of Experimental Psychology*, 40(1), 24-32.

Bigand, E., & Parncutt, R. (1999). Perceiving Musical Tension in Long Chord Sequences. *Psychological Research*, 62, 237-254.

Cook, N. (1987). The Perception of Large-Scale Tonal Closure. *Music Perception*, 5(2), 197-205.

Cuddy, L. L., & Smith, N. A. (2000). The perception of tonal pitch space and tonal tension. In D. Greer (Ed.), *Musicology and sister disciplines: Past, present and future* (pp. 47-59). London: Oxford.

Cuddy, L. L., & Thompson, W. F. (1992). Asymmetry of perceived key movement in chorale sequences: Converging evidence from a probe-tone analysis. *Psychological Research*, 54(2), 51-59.

Farbood, M. M. (2016). Memory of a Tonal Center after Modulation. *Music Perception*, 34(1), 71-93.

Frieler, K., et al. (2013). Absolute Memory for Pitch: A Comparative Replication of Levitin's 1994 Study in Six European Labs. *Musicae Scientiae*, 17(3), 334-349.

Germano, N. G. (2015). Em busca de uma definição para o Fenômeno do Ouvido Absoluto [Seeking a definition for Absolute Pitch phenomenon]. (unpublished master's dissertation). Instituto de Artes da UNESP. São Paulo, Brazil. http://www.academia.edu/15426596/Em_busca_de_uma_defini%C3%A7%C3%A3o_para_o_fen%C3%B4meno_do_Ouvido_Absoluto

Germano, N. G., et al. (2016). Absolute Pitch: In Search of a Testable Model. *Proceedings of the International Conference on Music Perception and Cognition 14*. San Francisco, USA, 710-713.

Germano, N. G., et al. (2018). Psychometric Features to Assess Absolute Pitch: Looking for Construct Validity Evidences Regarding Isolated Pitch Tasks in Undergraduate Brazilian Music Students. *Proceedings of the 15th International Conference on Music Perception and Cognition (ICMPC15)*. Graaz, Austria.

Hsieh, I. H., & Saberi, K. (2008). Dissociation of procedural and semantic memory in absolute-pitch processing. *Hearing Research*, 240, 73-79.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial map of musical keys. *Psychological Review*, *89*, 334-368.

Lerdahl, F. (2001). *Tonal pitch space*. New York: Oxford University Press.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling Tonal Tension. *Music Perception*, 24(4), 329-366.

Levitin, D. J. (1994). Absolute Memory for Musical Pitch: Evidence from the Production of Learned Melodies. *Perception & Psychophysics*, 56(4), 414-423.

Marvin, E. W., & Brinkman, A. (1999). The Effect of Modulation and Formal Manipulation on Perception of Tonic Closure by Expert Listeners. *Music Perception*, 16(4), 389-407.

Marvin, E. W., & Brinkman, A. (2000). The Effect of Key Color and Timbre on Absolute Pitch Recognition in Musical Contexts. *Music Perception*, 18(2), 111-137.

Miyazaki, K. (1989). Absolute Pitch Identification: Effects of Timbre and Pitch Region. *Music Perception*, 7(1), 1-14.

Miyazaki, K. (1990). The Speed of Musical Pitch Identification by Absolute-Pitch Possessors. *Music Perception*, 8(2), 177-188.

Parncutt, R., & Levitin, D. J. (20018, May 4). Absolute pitch. *Grove Music Online*. Retrieved from http://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000000070

Schlemmer, K. B. (2009). Do Absolute Pitch Possessors Have a Field Independent Cognitive Style? *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music (ESCOM 2009)*. Finland: University of Jyväskylä, 480-486.

Schellenberg, E. G., & Trehub, S. E. (2003). Good Pitch Memory Is Widespread. *Psychological Science*, 14(3), 262-266.

Schenker, H. (1935). *Der freie Satz [Free Composition]*. Wien: Universal.

Smith, N. A., & Cuddy, L. L. (2003). Perceptions of musical dimensions in Beethoven's Waldstein sonata: An application of Tonal Pitch Space theory. *Musicae Scientiae*, 7(1), 7-34.

Smith, N. A., & Schmuclker, M. A. (2008). Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors. The Journal of the Acoustical Music Society of America, 123(4), EL77-EL84.

Takeuchi, A. H., & Hulse, S. H. (1991). Music Perception, 9(1), 27-46.

Terhardt, E., & Ward, W. D. (1982). Recognition of Musical Key: Exploratory Study. The Journal of the Acoustical Music Society of America, 72(1), 26-33.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Behavioral and Neurophysiological Effects of Singing and Accompaniment on the Perception and Cognition of Song

Yke P. Schotanus[1], Lynn S. Eekhof[2], Roel M. Willems[2]

[1] *Institute for Cultural Inquiry, ICON, Utrecht University, The Netherlands*

[2,3] *Centre for Language Studies, Radboud University Nijmegen, Nijmegen, The Netherlands*

[1]schotschrift.teksten@planet.nl

Abstract

In order to further investigate the effects of singing and accompaniment on the processing of language, a classroom experiment reported earlier was followed by an EEG experiment, using the same materials. 24 participants listened to four songs, each in one of four versions: spoken, sung a cappella, complete (sung with accompaniment), or vocalized (sung a cappella on 'lala'). During listening, EEG was measured, and after each song, a questionnaire was filled out. Behavioral results suggest that singing supports cued word recall, even after just one exposure, and focus on the lyrics. Furthermore, an accompaniment supports positive affect and appreciation of voice quality, and decreases seriousness. A preliminary EEG data analysis reveals that out-of-key notes elicit a slightly larger ERAN and N400 than in-key notes, a smaller P2 and N5, and a larger P600, followed by a larger late negativity ('N1400'). However, the larger ERAN and P2 are not visible in all conditions; the larger N400 is only significant in the condition complete, the larger P600 predominantly in the condition vocalized, and the late negativity only in the condition a cappella. These differences lead to the conclusion that the processing of in-key and out-of-key notes interacts with the presence of interpretable lyrics, indicating that music might affect the meaning of words or vice versa. The interaction between the processing of these notes and the presence of an accompaniment is more difficult to interpret.

Introduction

The processing of song lyrics is thought to be affected by the song's music. On the one hand, music is thought to increase arousal and attention, which could support language processing. On the other hand, violations of both rhythmic and musical expectancies within the melody can hamper language processing (Kunert, 2017; Gordon et al., 2010), although they might also accentuate linguistic events (Schotanus, 2015).

To date, interactions between language and music have been tested with deviant notes or chords in short standardized stimuli only, not in complete songs in which notes or chords can have different functions, can be repeated, and followed by other unexpected notes. Furthermore, the role of accompaniment is seldom taken into account.

In a classroom study, in which 274 pupils reacted to different versions of four different songs, Schotanus (2016) has found that pupils rate sung lyrics as more intelligible and comprehensible than spoken lyrics, and that they rate accompanied song versions (either sung or spoken) both emotionally and aesthetically more positive, and easier to focus on than unaccompanied ones. Furthermore, their ratings of the emotional meaning of the lyrics were more in line with the author's intentions if they were sung and accompanied. Finally, the voice quality was rated less positive, and the singing less pure in song versions without accompaniment.

These results were interpreted to be the result of Musical Foregrounding, and thus to support the Musical Foregrounding Hypothesis (MFH; Schotanus, 2015), according to which music functions as a foregrounding device (Miall & Kuiken, 1984), obstructing the processing of sung words, and thereby accentuating them as well, if the listener is able to overcome the obstruction. In this study, the processing of difficulties in the music, especially on-beat silences or loud rests (London, 1993; Honing et al., 2009) and of out-of-key notes (ooks), was hypothesized to be more demanding in versions without accompaniment. In accompanied versions, loud rests would be absent, and harmonies would not just be implied but actually sound. As a result, foregrounding would be maximum in a-cappella versions, possibly too much so for the listener. Hence the enhanced concentration, tonal perception, and valence in accompanied versions, but also the enhanced processing of the lyrics in sung versions in general.

The aim of the current study is to replicate the results of the classroom experiment and to test whether there is neurophysiological evidence that a cappella versions are more demanding than accompanied ones, and that music affects the meaning of words.

Earlier research on the processing of ooks (and other unexpected tones or chords) has shown somewhat confusing results. In several studies (see Kim, 2013, for a review) both in-key notes (iks) and ooks elicit an Early right anterior negativity (ERAN; associated with a preconscious detection of an unexpected event) and an N5 (associated with the meaningful integration of this event). The amplitude of both ERAN and the N5 are usually larger for unexpected notes and chords, and can also be modulated by position. Furthermore, the ERAN for single notes tends to be earlier than the ERAN for chords. However, other researchers found a larger P2 instead of an ERAN (associated with the violation of memory-based expectations) for ooks (Choi et al, 2014), a bipolar ERAN (Sammler et al., 2012), or a right frontal N350 combined with a large P600 (Patel et al., 1998) (associated with the integration of syntactical anomalies). The P600 might sometimes mask an N5 (Koelsch et al., 2005).

Importantly, these ERPs have shown to be largely independent of musical experience, attention, and repetition, although these factors do influence the amplitudes. For example, both P2 and ERAN are usually larger in experienced musicians, and both ERAN and N5 are somewhat smaller if listening to the music is not the participant's main task, or if a particular note or note sequence is repeated during an experimental trial (Choi et al., 2014; Kim, 2013).

In line with these findings, we expected to find ERANS or P2s and N5s for iks, and the same ERPs with larger amplitudes, plus a P600 or an N350 for ooks. We expected the smallest amplitudes for complete (accompaniment effect) and the largest for vocalized conditions (attention effect).

Finally, the N400, associated with the processing of linguistic meaning, but sometimes affected by music as well (Koelsch et al., 2004) might be larger for ooks, but absent in vocalized versions, due to the absence of lyrics.

Method

Participants

24 participants (18 women) were recruited from the subject pool of the Radboud University Nijmegen to take part in this experiment in exchange for money (*n* = 22) or course credit (n = 2). Participants were 19 to 37 years old (*M* = 24;4, *SD* = 4.8) native speakers of Dutch without a history of neurological problems or hearing deficits. All participants except one were right-handed. They were either still in university (*n* = 21), or had finished university a few years prior to the experiment (*n* = 3). Self-reported musical sophistication, measured by the Gold MSI (Bouwer et. al, in preparation (translation); Müllensiefen et al., 2014) varied from 36 to 111 (*M* = 67.7; *SD* = 19.1 (scale maximum = 136)), musical experience (an MSI subscale) from 7 to 41 (*M* = 22,0; *SD* = 11.9 (scale maximum = 49)).

Materials

Ecologically valid stimuli, namely four complete cabaret songs in Dutch, were used for the experiment (all stimuli and questionnaires available online (Schotanus, 2017)). Each song contained several occurrences of ooks and (different kinds of) loud rests in the melody part. The frequencies of these events per song can be found in Table 1 below. For the EEG analysis of the ooks, an equal amount of iks of similar length was selected in similar metric positions (see Figure 1). Thus, possible confounding effects of metrical position, note length, musical context, and repetition are eliminated. All EEG markers were time-locked using Praat (Boersma & Weenink, 2015).

All songs were pre-existing but seldom-performed cabaret songs composed and sung by the first author, a male baritone. Using these songs ensured that there were enough ooks and loud rests in each song for EEG-measures and practically excluded the possibility that participants had prior knowledge of the songs. A disadvantage might be that the songs are written in the popular tradition which is not strictly tonal and in which several out-of-key notes are might not be perceived as out-of-key (Temperley, 2001). However, the musical expectations of Western people (musician or not) have shown to be largely in line with the rules of tonal music (see Huron, 2006, for a review). Furthermore, a classically-schooled musician has judged all the target notes as 'feeling out-of-key within the context'**.**

The recordings were made using a Neumann TLM 103 microphone and an Avalon VT 737 SM amplifier. In order to ascertain that all the singing was in tune, the voice-treatment software Waves Tune, Renaissance Vox compression and Oxford Eq were used to perform vocal pitch correction.

Table 1: distribution of musical foregrounding events over the four songs.

Song: / Event:	1	2	3	4
Out-of-key notes	38	29	36	36
Loud rest	94	79	66	55

Figure 1. Out-of-key notes and in-key-control notes in the verse part of one of the songs (WM).

Accompaniments were improvised according to the chord scheme of each song on a keyboard and recorded using ProTools 10 by Christan Grotenbreg, a professional musician.

All songs are of approximately the same complexity. In a separate online survey, 42 participants recruited via Amazon Mechanical Turk between 24 and 66 years of age (Mean 36.6; SD 10.5; 48.7 % male) rated the complexity of all accompaniments and vocalized voice parts. Mean complexity ratings ranged between 4.1 and 4.9 on a 7-point scale. A regression on these ratings with song, song version, and their interaction showed no significant effects. The effect of song was just trending to significance (*p* = .066). Lyrics were found to be of the B2-level according to the CEFR tool (Council of Europe, 2011; Velleman & Van der Geest, 2014; Stichting Accessibility, 2014). Nevertheless, there are clear differences in tempo and content, which might affect appreciation. Two songs are relatively upbeat (about 110 BPM) and have love-related lyrics, while the other two are somewhat slower (approximately 80 BPM), and have lyrics about dealing with life's threats. Hence, in our statistical analyses, we used random intercepts for songs.

All songs were recorded in four different conditions: spoken, a cappella, vocalized (sung a cappella on 'lalala'), and complete. In the complete versions, the sung lyrics are presented together with a harmonic accompaniment, with a musical event at every beat. The isolated voice part of this recording was used for the a-cappella version. Furthermore, both the vocalized and the spoken version are sung or recited in accordance with the timing of the complete version. That is, they were recorded while the 'singer' listened to the accompaniment, and started each phrase at approximately the same time as the sung lyrics would have started in the song. As a result, all tracks in all conditions had approximately the same duration (± 2 seconds), excluding a possible confound of rate of presentation when comparing the spoken and the musical conditions. However, the speech in the spoken version is quite slow and riddled with silences, which might create foregrounding effects in the spoken version as well.

In order to assess appreciation, perception, comprehension, and recall, questionnaires were designed. For each condition, there was a questionnaire with 7-point Likert scales assessing participants' judgments of the lyrics, music, performance, listening experience and voice of the singer. Furthermore, there were song specific questionnaires containing three very simple comprehension questions to assess whether participants had listened to the lyrics. In addition, there were several questions to assess comprehension of the lyrics (to be reported elsewhere), and a fill-in-the-blank test to measure verbatim recall. Six sentences were selected, evenly distributed over the full length of the song. In half of these sentences, a word that was supported by rhyme (alliteration, assonance or end rhyme) was left out, whereas in the other half neutral words were left out. The questionnaires specific to lyrics were not filled out after the vocalized condition.

Procedure

The experiment took place in a small, sound-proof booth with a desk and a computer screen in front of the participant from which the stimuli were played. Another screen was present in the booth but invisible to the participant, to conduct the recording of the EEG data. Stimuli were pseudo-randomly assigned to participants in such a way that every participant heard every song and every condition exactly once.

Before the experiment started, participants were asked to sign informed consent and fill out the Gold MSI and another questionnaire (not discussed here). During the experiment, songs were presented over Sennheiser HD 215 headphones using Presentation software (v. 17.1, Neurobehavioral Systems) at a sound level that was judged to be optimal by each individual participant after the presentation of several beeps. The start of each song was indicated with a beep and a fixation cross accompanied each song until the end in order to make sure participants would not erroneously think the song had finished during silences in the versions without an accompaniment.

After each song, participants manually filled out the questionnaires about the track they had just heard. The complete experiment, including the set-up of the EEG, took about 100 minutes.

Data recording and analysis

Behavioral data were analyzed in RStudio (version 1.1.383; R Core Team, 2017). For dimension reductions of the questionnaires, we used the *psych* and *GPArotation* packages (Revelle, 2017; Bernaards & Jennrich, 2005), for linear mixed models the *lme4* package (Bates et al., 2012).

EEG data were recorded in Brain Vision Recorder with a 64 channel system with active electrodes. Four electrodes were used for the vEOG and hEOG, and two electrodes were placed on the mastoids. An additional electrode was placed on the forehead as a ground. Data was recorded at a sampling rate of 500 Hz with a low cutoff filter of 0.016 Hz and a high cutoff, anti-aliasing filter of 1000 Hz.

All EEG data were analyzed using the Fieldtrip MATLAB toolbox (Oostenveld, Fries, Maris & Schoffelen, 2011). For analysis, the continuous EEG signal was cut into epochs starting 500 ms before and ending 2000 ms after the onset of the target note. The data were re-referenced to the mean of the two mastoids. In addition, detrending and a low-pass filter of 70 Hz were applied. After the manual rejection of any large and obvious disturbances of the signal, an Independent Component Analysis was run on the data to remove artifacts such as eye blinks and noise. We focused on time windows related to the components of interest. We compared iks and ooks in each condition between 100-175 (early ERAN), 175-225 (later ERAN and P2); 300-400 (N350 and onset N400); 400-500 (N5); and 600-900 (P600). Furthermore, we compared iks between conditions a cappella and the other conditions between 100-225; 300-400 and 400-500. Post hoc, after a visual inspection of the waveforms we tested the significance of a few unexpected but salient ERPs.

Results

We will present a preliminary analysis of a part of our data.

Questionnaires

After a Principal Axis Factoring analysis with oblique rotation (direct oblimin) on the Likert-scale items occurring in all questionnaires concerning song versions with text, four factors with eigenvalues over 1 were retained: Positive affect (representing joyfulness, sensitivity, energeticness, intelligibility, comprehensibility, and the opposite of boringness and tiringness); Seriousness (representing sadness, sensitiveness, heaviness, strikingness of formulations, and the ability to make you think); Voice quality (representing naturalness, relaxedness and pleasantness of the 'singer's' voice); and Strikingness of textual features (representing funniness and, again, strikingness of formulations). Eigenvalues were 3.32, 3.01, 2.57 and 1.80 respectively.

For each component separately, a linear mixed model was constructed with condition as a fixed effect and random intercepts for participants and songs. Likelihood Ratio Tests were used to compare the models with and without the fixed effect of condition. This revealed that condition had a significant effect on all components: Positive affect ($\chi^2(2) = 14.58$, $p = 0.0007$), Seriousness ($\chi^2(2) = 14.90$, $p = 0.0006$), Voice quality ($\chi^2(2) = 7.08$, $p = 0.03$) and Textual features ($\chi^2(2) = 15.80$, $p = 0.0004$).

Post-hoc Tukey tests were performed to compare the means of the three conditions. For Positive affect, the complete versions were scored significantly higher than the spoken ($p = 0.001$) and a-cappella versions ($p < 0.001$). Spoken and a cappella did not differ significantly ($p = 0.99$). For Seriousness, the score for the complete versions was significantly lower than the score for the spoken one ($p < 0.001$). The score for the a-cappella versions was higher, but not significantly higher, than the score for the complete ones ($p = 0.096$). Similarly, the score for the a-cappella versions was lower, but not significantly lower, than the score for the spoken ones ($p = 0.082$). For Voice quality, the score for the complete versions was significantly higher than the scores for the spoken ($p = 0.041$) and a-cappella versions ($p = 0.047$). Again, the difference between the spoken and a cappella versions was not significant ($p = 0.999$). Finally, for Strikingness of textual features, the complete versions were scored significantly higher than the spoken ones ($p < 0.001$).

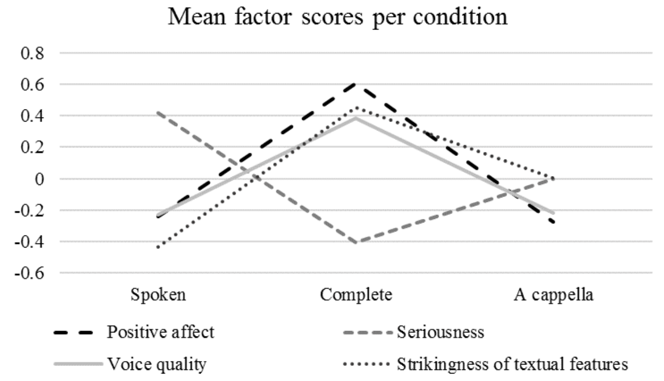


Figure 2. Mean factor scores per condition.

Compared to the complete versions the a-cappella versions again scored lower but this difference only approached significance (p = 0.07). The same was observed for the difference between the spoken and the a-cappella versions (p = 0.08).

Recall

Figure 3 shows the mean scores per condition for the fill-in-the-blank recall test. As there were six recall trials per song, the maximum total score was 6. Of these six trials, half were supported by rhyme, and half were not. One song trial was excluded because the participant scored below chance level on the comprehension check questions (score < 2).

A linear mixed model was fitted to the recall data to assess the effect of condition on total recall. Condition was added as a fixed effect, with random intercepts for participants and songs. A Likelihood Ratio Test showed that condition had a significant effect on total recall ($\chi^2(2) = 7.20$, $p = 0.027$). A post-hoc Tukey test revealed that only the difference between the spoken and the a-cappella versions was significant (p = 0.015), with superior recall in the a cappella condition. The differences between the spoken and the complete versions (p = 0.41) and between the complete and the a-cappella versions (p = 0.28) were not significant, neither was the effect of rhyme ($\chi^2(1) = 0.14$, p = 0.71).

ERPs

Iks. The ERPs elicited by iks do not differ significantly. They all show an ERAN somewhat earlier than expected (around 100 ms after target onset (to), see figure 4), followed by a P2 and an N5 peaking relatively late in the conditions complete and a cappella. We did not expect this difference and hence we did not measure the significance of it yet. In our target time windows, 300-400 and 400-500 ms after to, there were no significant differences between a cappella and complete or a cappella and vocalized. However, in the time window 100-225 ms after to four electrodes show a significant negativity for vocalized compared to a cappella, representing a larger and later ERAN for vocalized ($p < 0.05$, see figure 4).

Ooks. The ERPs elicited by ooks show significant differences compared to those elicited by iks in all conditions. However, these differences vary per condition and are often visible in another time window than expected. Furthermore, most differences are significant only on a 0.05 level which might cause problems with multiple comparisons. Nevertheless, it is interesting to report them.

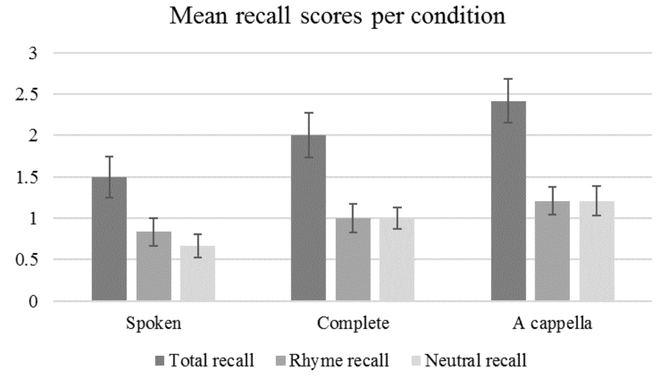


Figure 3. Mean recall scores per condition.

Time window 100-175; 175-225; 100-300. In the two time windows associated with the ERAN, the only difference which reaches significance is a small one between a cappella iks and a cappella ooks in the time-window 100-175. Three electrodes show a significant negativity for ooks ($p < 0.05$, see figure 4). However, visual inspection of the waveforms revealed that the ERANS often peak earlier and the P2s later. Hence, a post hoc analysis of the time windows 100-300 showed a significant negativity for all ooks compared to all iks ($p < 0.05$), which seems mainly due to a cappella ooks, where the P2 is almost absent.

Time window 300-400; 400-500. In the time-window 300-400, the waveforms of both a cappella and complete show a negative tendency for ooks. This difference is significant only for 5 centro-parietal electrodes in the condition complete ($p <$ 0.05, see Figure 4, bottom). In the condition vocalized on the other hand, ooks show a positive tendency in this time window, which becomes significant between 400 and 500 on eleven centro-parietal electrodes ($p < 0.05$, see figure 4, bottom). In the conditions a cappella and complete on the other hand ooks show a positive shift in this time window.

Time windows 600-900; 500-1000. Within the time window 600-900 all conditions together ($p < 0.05$), and the condition vocalized ($p < 0.05$) show a significant positivity for ooks across the whole scalp. In the complete condition 3 electrodes show such a significant positivity ($p < 0.05$). As the difference curves indicated that the P600 peaks either earlier or later (partly due to a late N5 peak) we also tested the time window 500-1000. In this time window, the positivities for both vocalized and all conditions together are significant at a higher level ($p < 0.005$). A cappella shows both positivities and negativities at various electrodes, hence on average there were no significant effects.

Time windows 1350-1450, 1350-1800. In all conditions, the ook-curve shows a late negative shift. In the a-cappella condition, this leads to a significant negativity peaking around 1400. This negativity is significant over the whole scalp between 1350 and 1450 ($p < 0.05$); and at 8 electrodes between 1350 and 1850 ($p < 0.05$), see figure 4 upper half.

Discussion

The effect of musical context and musical complexity on the processing of song lyrics was tested in an experiment with both offline behavioral measures and online EEG measures. The results of important parts of the data confirm and extend the results of an earlier classroom study, and might be interpreted as support for the MFH.

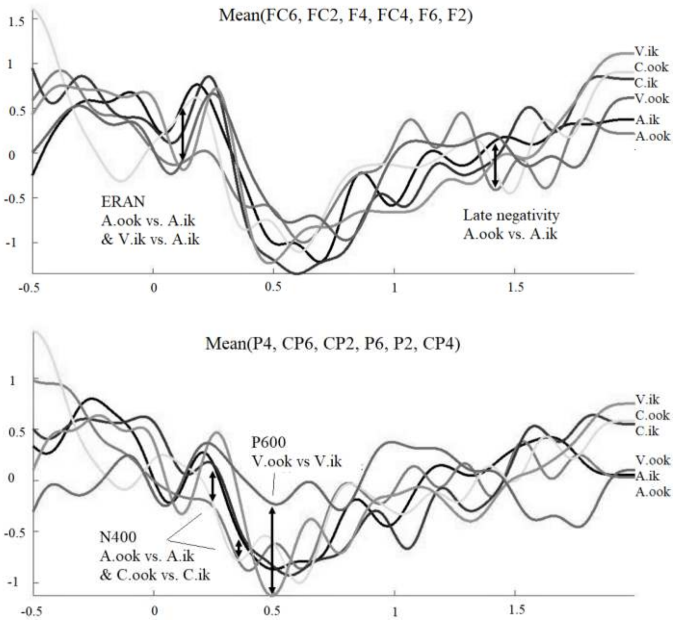


Figure 4. Grand average waveforms of ERPs elicited by iks and ooks at right-anterior (above) and right-centro-parietal electrodes (below) in the conditions: Vocalized (V), A cappella (A) and complete (C). Arrows show salient early and late right-anterior negativities (above), N400s and P600s (below).

Behavioral results

Complete versions were scored significantly higher on Positive affect, Appreciation of voice quality and Strikingness of textual aspects, and significantly lower on Seriousness than their spoken counterparts. Furthermore, recall was significantly better after hearing a cappella version than a spoken version, which proves that music can support recall without repetition. These results directly support the MFH's claim that music, via musical foregrounding, can have beneficial effects on the listener by increasing appreciation for and perception and cognition of song lyrics.

The results also seem to confirm the complex predictions concerning a cappella singing. We hypothesized that the foregrounding is maximum in a cappella music which might make hard to overcome. And indeed in the condition a cappella the scores for Appreciation of voice quality and Positive affect are lower than in the condition complete. At the same time, however, cued word-recall is enhanced, which indeed indicates a maximum effect of foregrounding. Furthermore, the scores for the text centered factors Seriousness and Strikingness of textual features indicate a slightly enhanced focus on the lyrics in this condition.

One important difference with the classroom study is that in the current study intelligibility and comprehensibility (worked up in Positive affect) do not seem to be enhanced in a cappella singing compared to speech. The fact that it was enhanced in the classroom study because a classroom is much more demanding for intelligibility and concentration than a soundproof booth.

ERPs

In spite of differences in musical-syntactic saliency, length and metric position of ooks, and of the repetitive structure of the music, the ERPs elicited by ooks differ on average from those elicited by iks. This indicates that even in a musical context which is not strictly tonal, the schematic expectations of Western listeners concerning tonality are stronger than veridical expectations. However, while the ERPs elicited by iks are roughly similar across conditions, the differences between the ERPs elicited by iks and ooks are not. These results are just partly in line with our predictions.

Iks. We hypothesized the occurrence of an ERAN and an N5 in all conditions. Furthermore, we expected the ERAN in the condition complete to be later and smaller than in the unaccompanied conditions, and largest in the condition vocalized. At the same time, the N400 was hypothesized to be smaller in this condition. Indeed, we did find an ERAN and an N5 in all conditions, however, the only significant difference we found, was a larger ERAN in vocalized compared to a cappella, which is in line with the expectation that the presence of lyrics reduces attention for the music.

Ooks versus iks. We expected larger ERANs, P2s, N400s, N5s, and P600s for ooks than for iks in all conditions, and possibly an N350. Furthermore, we expected the amplitudes of these ERPs to be smallest in the condition complete, and largest in the condition vocalized, although the N400 was expected to be the lowest in that condition.

ERAN. Just a cappella ook shows a significantly larger ERAN. Furthermore, unexpectedly, the P2 is reduced in the conditions a cappella and complete, resulting in a significant negativity for ooks in general between 100-300ms after to, which might be interpreted as part of a build-up towards the N400.

N350, N400, N5. In the time window 300-400 both a cappella and complete show negativities, while vocalized does not. However, only the negativity in the condition complete is significant. Nevertheless, the waves for both a cappella ook and complete ook show a clear peak around 375 ms after to., As this peak is more salient in centro-parietal electrodes, instead of right anterior electrodes (where Patel (1998) found the N350), we think it might rather be interpreted as a larger N400 than as an N350. In the time window 400-500, which is the onset of the N5 we found an unexpected positivity for ooks in all conditions. Although this positivity is significant only in the condition vocalize, it is unmistakable that the ooks curve in a cappella and complete has taken a positive turn in this time window as well. This is a puzzling result. It seems to indicate that ooks are interpreted as linguistically but not musically meaningful and that the onset of the P600 reduces the N5 immediately.

N5, P600. All conditions together, and the conditions complete and vocalized separately, show a significant P600 for ooks though it is just marginally significant in the condition complete. What is puzzling, however, is that at several electrodes this P600 does affect the onset but not the amplitude of the N5 at its ultimate peak (around 600ms after to in the condition complete, and around 750ms in the condition vocalized ook). As a result the P600 peaks at unexpected moments. Possibly, if both P600 and N5 are relatively strong. This is the result.

Late negativity. It is unclear what kind of neural activity the late negativity visible predominantly in the condition a cappella, and possibly somewhat late in the condition vocalized represents.

Overall. Obviously, the absence of interpretable lyrics plays a part in the deviant processing of vocalized ooks. However, it is striking that it did not affect the processing of vocalized ik in such a salient way as well. Possibly, as expected, the focus on the language has alleviated the effect of the ooks in the conditions a cappella and complete. However, the effect of ooks is not just smaller in these conditions, above all it is different. Notably, the ERAN for ooks was largest in the condition a cappella, while the N400 for ooks was larger in the condition complete. Because of the N400 for ooks in a cappella and complete, it is more reasonable to think that the ooks are interpreted as more meaningful in the presence of language.

Conclusion

Both behavioral and EEG data show significant differences between musical conditions. The behavioral data support our hypothesis that singing would enhance lyric processing. It even enhances cued word-recall after a single exposure of a song. Furthermore, as expected, an accompaniment enhances the appreciation of voice quality and positive valence and suppresses seriousness. In line with our expectation that music might affect the meaning of words, the processing of iks and ooks seem interacts with the presence of lyrics. Ooks seem to be interpreted as being more meaningful in conditions with lyrics. Evidence for our expectation that an accompaniment would ease the processing of lyrics is more ambiguous, although the ERAN for ooks in the condition a cappella is larger than in the condition complete. Finally, it is remarkable that several components (ERAN, N400, P600) seemed to be earlier than expected, while others (P2, N5) were later.

Acknowledgments. The first author would like to thank his supervisors for their comments; his colleagues from UiL OTS for their help with Lime Survey; Christan Grotenbreg for recording the stimuli and digitally manipulating the recordings; and NWO, the Dutch Government, and SG het Rhedens for granting him the opportunity to avail himself of a PhD scholarship for teachers. The current paper is partly based on an internship project of the second author.

References

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Bernaards, C. A. & Jennrich, R. J. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. *Educational and Psychological Measurement, 65*, 676-696.

Boersma, P. & Weenink, D. (2015). *Praat: doing phonetics by computer* [Computer program]. Retrieved, 28 October 2015 from http://www.praat.org/

Bouwer, F., Schotanus, Y. P., Sadakata, M., Mullensiefen, D & Schaefer, R. (in preparation) *Measuring musical sophistication in the low countries; validation of a Gold MSI translation in Dutch*.

Choi, I., Bharadwaj, H. I., Bressler, S., Loui, P., Lee, K. & Shinn-Cunningham, B. G. (2014) Automatic processing of abstract musical tonality. *Frontiers in Human Neuroscience*, 8, 988.

Gordon, R. L., Magne, C. L. & Large, E. W. (2011). EEG correlates of song prosody: a new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology,* 2, 352.

Honing, H. J., Ladinig, O. Winkler I. & Háden G. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials (ERP). *Annals of the New York Academy of Sciences*, nr. 1169: *The Neurosciences and Music: Disorders and Plasticity*, p. 93-96.

Huron, D. (2006) *Sweet anticipation*. Cambridge, MA: MIT Press.

Kim, J. N. (2013) *Online processing of tonal melodies: effects of harmonic expectations* (doctoral dissertation). Evanston, IL.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T. C., Friederici, A. D. (2004). Music, language, and meaning: Brain signatures of semantic processing. *Nature Neuroscience 7,* 302-307.

Koelsch, S., Gunter, T. C., Wittfoth, M. & Sammler, D. (2005). Interaction between syntax processing in language and in Music: an ERP study. *Journal of Cognitive Neuroscience*, 17(10), 1565-1577.

London, J. M. (1993) Loud rests and other strange metric phenomena (or, meter as heard). *Music Theory online*. Http://mto.societymusictheory.org/issues/mto.93.0.2/mto.93.0.2london.art.

Miall, D. & Kuiken, D. (1994). Foregrounding, Defamiliarization, and Affect Response to Literary Stories. *Poetics*, 22, 389-407.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE,* 9(2). Questionnaire retrieved from https://www.gold.ac.uk/music-mind-brain/gold-msi/download/.

Oostenveld, Oostenveld, R., Fries, P., Maris E. & Schoffelen J-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience, 11*.

Patel, A., Gibson, E., Ratner, J., Besson, M. & Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717-733.

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Revelle, W. (2017). Psych: Procedures for Psychological, Psychometric, and Personality Research (version 1.7.8) [Computer software]. Evanston, IL: Northwestern University.

Sammler D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H.-J., Knösche, T. R., Wellner, J., Widman, G., Elger, C., Friederici, A. D., Schulze-Bonhage, A. (2012). Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage,* 64, 134-146.

Schotanus, Y. P. (2015) The musical foregrounding hypothesis: How music influences the perception of sung language, Ginsborg, J., Lamont, A., Philips, M. & Bramley, S. (Editors) *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, 17-22 August 2015, Manchester, UK.

Schotanus, Y. P. (2016/submitted) *Both singing and accompaniment support the processing of song lyrics and change their meaning.* (Based on) Y.P. Schotanus, Music supports the processing of song lyrics and changes their contents: Effects of melody, silences, and accompaniment (Abstract), In: *Proceedings of the 14th International Conference on Music Perception & Cognition (ICMPC14)*, July 2016, San Francisco. Zanto, T.P. (Ed.). Adelaide: Causal Productions, p. 361. (Full text available on Researchegate and AcademiaEdu, and in Schotanus, 2017).

Schotanus, Y.P. (2017). Supplemental materials for publications concerning three experiments with four songs, hdl:10411/BZOEEA, DataverseNL Dataverse, V1

Temperley, D. (2001) *The cognition of basic musical structures*. Cambridge, MA/London, UK: MIT.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Out-of-Key Notes and On-Beat Silences as Prosodic Cues in Sung Sentences

Yke Schotanus[1]

[1] *Institute for Cultural Inquiry (ICON), Utrecht University, The Netherlands*

[1]schotschrift.teksten@planet.nl

Abstract

Violations of musical syntactic expectancies such as out-of-key notes are known to interact with linguistic processing, due to shared syntactic integration resources, located in Broca's area. As these are syntactic integration resource, researchers have assumed that such events negatively affect the processing of language, and that they do not affect semantics. However, the results of this study challenge both assumptions. An online listen-experiment shows that out-of-key notes sometimes do affect semantics. Thirty participants listened to thirty sung sentences in three conditions and rated the plausibility of literal and colored (emotional, ironic or metaphoric) interpretations . Out-of-key notes significantly affected these ratings. Loud rests (on beat silences) did not yield a similar effect.

Introduction

Language and music share several characteristics. Moreover, recent findings indicate that processing language and processing music even rely on shared neural resources (Sammler et al., 2010; Lidji et al., 2009; LaCroix et al., 2015). For example, in line with Patel's Shared Syntactic Integration Resources Hypothesis (SSIRH, Patel, 2003), evidence is found that the processing of musical and linguistic syntax interact in Broca's area (Kunert, 2017). Presumably as a result of this interaction, both neurophysiological and behavioral studies show that violations of musical-syntactic expectancies (VMSEs), such as out-of-key notes, affect the processing violations of linguistic syntactic expectations (VLSEs) (see Kunert, 2017, 18-21, for a review) and that reading grammatically complex sentences while listening to music negatively affects music processing (Kunert, Willems & Hagoort, 2016). As the interaction in Broca's area is purely syntactic (Kunert, 2017) out-of-key notes and unexpected harmonies are claimed not to affect semantics, although in musical analyses it is quite usual to interpret such musical events as meaningful, for example ironic (Burns, 2000), and several studies show that out-of-key notes can evoke specific qualia (Huron, 2006; Arthur, 2018).

The discussion on the possible interaction between VSMEs and semantics is blurred by three problems which occur in several SSIRH-related studies. Firstly they show a biased focus on negative effects, secondly they do not take into consideration a possible effect of prosody, and thirdly they often base their conclusions on edited data in which the original data are incorporated. Examples will follow.

Please note that there is no question about the evidence for the SSIRH. What is questioned here, is the hypothesized negative effect of VMSEs on language processing as a whole. The competing hypothesis here is, that VMSEs might be difficult to process, and thus might pace the processing of simultaneously presented words (Slevc et al., 2009), but ultimately will support language processing, if the VMSEs can be interpreted as meaningful prosodic accents. This might sound paradoxical, but a comparable paradoxical effect, known as foregrounding, is shown in linguistics, literature and film (Miall & Kuiken, 1994; Hakemulder, 2004; Hakemulder, 2008). According to the Musical Foregrounding Hypothesis (MFH, Schotanus, 2015) such an effect is also present in song. The MFH might shed a new light on the debate about VSMEs and semantics, which might lead to the conclusion that indeed this debate is blurred by the three problems mentioned above.

Both a focus on negative effects and the use of edited data can be found in a study by Poulin Charronat et al. (2005). In this study participants heard sung sentences accompanied with eight chords. The last chord was either an tonic chord or a less expected subdominant chord, ant the last word was either a word or a non-word, and if it was a word it was either semantically related or unrelated to the linguistic context. Afterwards, participants had to decide whether the last word was a non-word or not. The researchers found that tonic chords supported the detection of semantically related words as words. Furthermore, their images indicate that unrelated chords support the detection of semantically unrelated words, but they did not report that. Instead they reported that the difference between semantically related and unrelated conditions was larger for targets sung on tonic chords than those sung on subdominant chords. Reporting a difference is reporting edited data, and in this case it masks the positive effect of semantic unrelated chords on the detection of semantically unrelated words. Such a positive effect of unexpected chords would be in line with the results of a study by Curtis et al. (2003) in which unexpected chords support the recognition of unexpected words (and vice versa). Possibly, participants interpret unexpected chords as prosodic accents signaling wrongness, or peculiarity.

Such a mechanism, might also be the key to a study by Koelsch and Steinbeis (2008). In this study participants listened to five-word sentences, presented along with five-chord sequences. A surprising last chord presented along with a low-probe but correct word (such as 'beer' in 'He saw the cool BEER') elicited a so called N5, a brain potential associated with musical meaning. Three years later Carrus, Koelsch and Bhattacharya (2011) found a similar interaction which was close to significance, but furthermore no one has been able to replicate these results. However, studies that tried but failed to replicate them (Carrus, Pearce and Bhattacharya, 2013, among others) do not use stimuli in which the unexpected chord can be interpreted as a reflection of the meaningful unexpectedness of the word combined with it, only Steinbeis & Koelsch (2008) and Carrus et al. (2011) do. Therefore, it is important that the possibility of a prosodic effect of VMSEs on language processing is taken into account.

The third study reporting an interaction between VSMEs and violations of semantic expectancies is a study by Perruchet and Poulin Charronat (2013) who presented participant-paced combinations of chord-sequences with either garden-path or unambiguous sentences. They found that the difference in reading time between the two types of sentences was significantly lower when the chord sequence was completely in key, than when the chord sequence contained an out-of-key chord. The suggestion was, that the out-of-key chords interact with the semantic garden path, making them more difficult to process. However, a closer investigation of the data reveals that the out-of-key notes did not cause increased reading times for the garden-path sentences but decreased reading times for their unambiguous equivalents. Apparently, the out-of-key chords have somehow supported the processing of the unambiguous sentence. Possibly because it made sense as a prosodic accent. A detailed investigation of the stimuli has to ensure that such an explanation makes sense or not. However, it is no wonder that other studies, designed to assess the claim that VMSEs interact with semantics (such as Kunert, 2017), have failed to replicate the results of this study. These studies used different kinds of sentences as stimuli and did not reckon with a possible prosodic effect of the VSMEs.

To support the claim that VMSEs indeed can function as prosodic cues, and subsequently that WMSEs are sometimes able to affect semantics, an online listen experiment was conducted. The main aim of this experiment was to test whether VMSEs can change the interpretation of ambiguous sung sentences. Apart from out-of-key notes also loud rests (i.e. on-beat silences, London, 1993) will be used as violations of musical syntactic expectancies. Although Honing (2009, 119) suggests that a loud rest does not accentuate notes preceding or following it, it is likely that it does affect the processing of words preceding and following it, at least if it occurs at positions where in speech a silence would be interpreted as a 'pause for effect'. Loud rests cause substantial brain activity, more specifically, a so-called mismatch negativity (Ladinig et al., 2009), so apart from interrupting the linguistic phrase a loud rest might also distract attention. Furthermore, as the mismatch negativity is a preconscious phenomenon, the listener might even misattribute it to the language. Moreover, rhythmic manipulations are known to affect both musical and linguistic syntax (Gordon et al, 2015), and language comprehension (Quené and Port, 2005; Gordon et al, 2011), and they interact with the effect of simultaneously presented out-of-key notes and linguistic syntactic expectancies (Jung et al, 2013). Finally, in speech pauses can be interpreted as prosodical cues (Tyler, 2013).

Method

Participants and Procedure

31 Participants (30 of which completed the whole survey) were recruited via e-mail, Facebook, one of the websites Neerlandistiek.nl or Proefbunny.nl, or by live recruiting in pop-up laboratories in three book shops in Rheden area, the Netherlands. They were between 19 and 78 of age (M = 41,7; SD = 19,2)) and 74% was female. They completed the survey online, and received 5€ if they left their adress. Musical experience is measured by using a translation in Dutch of the Musical Training scale of the Goldsmith Musical Sophistication Index (Bouwer et al., forthcoming; Müllensiefen et al., 2014). Furthermore five questions were asked concerning literary experience. A Factor analysis on these questions resulted in two Factors, one indicating writing experience, and another one indicating disinterest in wording.

Stimuli

Three sets of stimuli were created, all of them consisting of different versions of the same thirty sung sentences. In each condition ten sentences were presented in their original form, i.e. sung fluently and in-key. In another ten sentences the original recording was edited such that two target words were delayed, creating an on-beat silence. In the remainder ten the same words were edited such that they were on beat but out of key. The pitches were changed by one semitone without harming the melodic contour. All sentences were sung a cappella, but preceded by a short piano-intro establishing a rhythm and a key.

There were sixteen target sentences and fourteen fillers sung to ten different melodies. In the target sentences the musically manipulated words all were thought to be ambiguous to a certain extent. Either because they could be interpreted in an ironic, or in a metaphoric, metonymic or very emotional way (with disgust, for example). The fillers were thought to be unambiguous. Furthermore, some of the fillers did not have the two part structure.

After each sentence the participants read three interpretations of the sentence, one of them literal, two of them more or less 'colored' (i.e. ironic, metaphoric, metonymic or very emotional). For example, the interpretations proposed for the sentence 'The shirt I bought last week is pink, did you happen to was it?' (Mijn nieuwe overhemd is rose; heb jij het soms gewassen?, see Figure 1) are:

A. Did you happen to wash my new shirt? I cannot find it. It is pink.
B. I have bought a pink shirt. Did you happen to wash it?
C. My new shirt is suddenly pink. Did you spoil it by washing it?

For each interpretation the participants were asked to rate on a seven point scale to what extent they thought it was plausible or not.

Figure 1. Three versions of one target sentence (The shirt I bought last week is pink; did you happen to wash it) : Fluent in key (top), Out of key (middle) and Loud rest (bottom).

Before recording, all sentences and interpretations were read and rated by an independent colleague unaware of the design of the experiment. When colored interpretations of fillers were rated as somehow reliable, these interpretations were skipped or changed, or the filler was further disambiguated. If neither of the colored interpretations of targets seemed to be somehow plausible, at least one interpretation or the target itself was changed. Furthermore, after recording, the author changed the wording of some interpretations, dependent on his own interpretation of the out-of-key and loud-rest versions.

All sentences were sung by the author (a male baritone), and recorded by Christan Grotenbreg, a professional musician, in his studio. The piano intros were improvised by Christan Grotenbreg on a keyboard connected to ProTools 10 (Desktop recording). The voice was recorded using a Neumann TLM 103 microphone, and an Avalon VT 737 SM amplifier. Digital conversions were conducted using Apogee Rosetta. To avoid confounds concerning purity and timing, voice-treatment software was used: Waves Tune, Renaissance Vox compression, and Oxford Eq.

Analysis

The results were analyzed using both linear and generalized linear Mixed models in SPSS. In line with Quené & Van den Bergh (2004) crossed classified analyses were run with random intercepts for both participant and sentence.

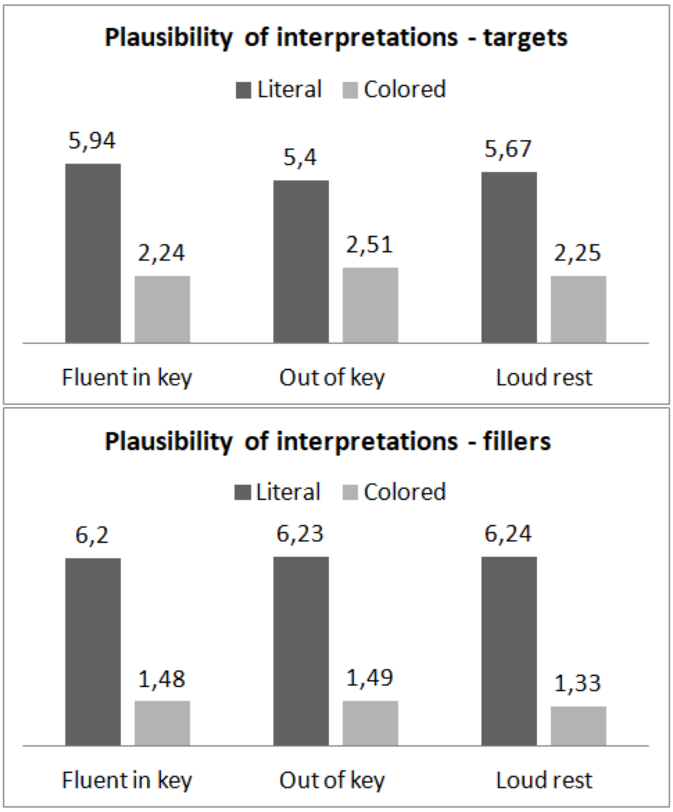


Figure 2. Mean plausibility ratings per condition for literal and colored interpretations of targets (top) and fillers (bottom). SDs fluent-in-key; out of key, and loud rest respectively for targets literal: 1.88; 2.00; 1.79; targets colored: 1,51; 1.92; 1.67; fillers literal: 1.34; 1.29; 1.33; fillers colored: 1.12; 1.08; 0.95.

Results

As Figure 2 shows, literal interpretations have been rated as much more plausible than colored ones in each condition. However, in the conditions out-of-key and loud-rest the difference is smaller. Literal interpretations are rated as slightly less plausible, while colored interpretations are rated as slightly more plausible. Conversely, the plausibility ratings for the fillers are constant across conditions.

In order to investigate whether this effect is significant several Mixed models linear regressions were conducted on the plausibility ratings for interpretations of targets. First, an intercept only model was run with two random intercept factors: participant and sentence. However, as sentence did not show a significant effect and an intercept only model without sentence was slightly more powerful (see Table 1), in the full model just one random effect was defined: participant. Apart from that the main effects of condition (fluent in key, out of key or loud rest), literal or not and condition*literal or not wer tested, plus the interaction between these factors. Musical training, writing experience and disinterest in wording did not show any significant effect and were deleted from the model. As Table 1 shows, there is a main effect for literal or not, but not for condition, although the difference between 'out-of-key' and 'loud rest' is close to significant. However, there is a significant interaction between literal or not and condition, especially for literal*out of key.

Table 1. Mixed models linear regression on Plausibility of interpretation

	df	-2 ll[a]	BIC[b]
Intercept only			
With sentence	4	6684.75	6713.88
Without sentence	3	6684.75	6709.60
Full Model	8	5854.61	5869.16
Type III & Estimates		*beta*	*F / t*
Fixed			
Intercept			1874.50***
Intercept		2.25 (0.13)	17.89***
Condition			0.87
Fluent in key		-0.01 (0.14)	-0.06
Out of key		0.26 (0.14)	1.87[+]
Literal or not			1134.81***
Literal		3.41 (0.17)	19.94***
Cond.*Lit.ornot[c]			5.83**
F.i.k.*literal[c]		0,28 (0.24)	1.19
O.o.k.*literal[c]		-0.53 (0.23)	-2.18*
Random			*Wald Z*
Residual		3.17 (0.12)	26.64***
Subject		0.19 (0.07)	2.86**

[a] -2 Restricted log likelihood

[b] Bayesian information criterion

[c] Condition times literal or not; F.i.k. = fluent in key; o.o.k. = out of key; other interactions are redundant

[+] *$p = 0.06$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$*

Table 2. Glmm binomial regression on Colored > literal

Model	Df	-2 ll[a]	BIC[b]
Intercept only			
With sentence	4	2380.48	2392.85
Without sentence	3	2333.11	2339.29
Full model	5	2351.11	2357.29
Type III, estimates		*b (SD)*	*F / t*
Fixed effect condition			2.81[+]
Intercept		1.81 (0.27)	6.61***
Loud rest		-0.06 (0.32)	-0.06
Out of key		-0.62 (0.30)	-2.09*
Fluent in key		0	
Random			*Z*
Participant		0.70 (0.29)	2.44*

[a] -2 log pseudo likelihood
[b] Bayesian information criterion
[+] $p = .061$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

An investigation of the occasions in which one of the colored interpretations was rated more plausible than the literal one revealed that this was more often the case in the out-of-key condition than in one of the other conditions (see Figure 3). However, a binomial regression on colored > literal revealed that the effect of condition is just close to significant, although the parameter estimate for out of key is significant compared to the one for fluent in key (see Table 2).

Because the predicted effect on 'at least one' colored interpretation might be attenuated by including all colored interpretations in the first regression, and because all ratings are the result of weighing up colored and literal interpretations, a third regression was run on a variable 'colored minus literal' representing the difference between ratings for the most plausible colored interpretation and the ratings for the literal interpretation. This regression showed a significant main effect of condition, indicating that the selective use of out-of-key notes significantly decreases the difference between the plausibility of literal and of colored interpretations.

Table 3. Mixed models regression on 'colored minus literal'.

Model	df	-2 l.l.[b]	BIC[c]
Intercept only	4	2496.49	2451.25
Full model	6	2418.13	2436.68
Type III, Estimates		***Estimate (SE)***	***F / Z / t***
Intercept			32.49***
Residual		6.95 (0.49)	14.81***
Participant		2.61 (0.79)	3.31**
Sentence		1.85 (0.76)	2.43*
Fixed effect condition			4.83**
Intercept		-3.01 (0.49)	-6.14***
Loud rest		0.26 (0.29)	0.87
Out of key		0.89 (0.29)	3.02**
Fluent in key			

[a] Levels/Parameters
[b] -2 Restricted log likelihood
[c] Bayesian information criterion
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

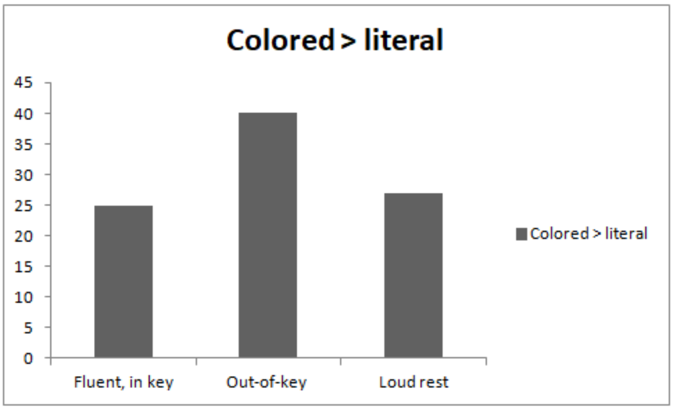


Figure 3. Occasions per condition in which a colored interpretation was rated more plausible than the literal one.

Discussion

The results of this Experiment show that out-of-key notes which are well aligned with ambiguous words can affect the interpretation of a sung sentence, concerning the effect of loud-rest the results are inconclusive. As the manipulation of the target notes was executed digitally in order to be sure that the singer's tone of voice was the same in all conditions, especially the differences between the out-of-key condition and the condition fluent-in-key were quite small. Nevertheless, the results show that out-of-key notes make literal interpretations less plausible, while supporting colored interpretations.

A regression analysis on the plausibility ratings only showed an interaction effect for out of key*literal which indicates that literal interpretations are less plausible in out the out-of-key condition, although the results seem to show an increase of the plausibility of colored interpretations. However, in this analysis all colored interpretations were included, although the hypothesis is that at least one of them would be more plausible in a manipulated version. Furthermore, in at least 15 cases the out-of-key condition has caused a complete shift in interpretation. An additional regression on colored > literal showed that this effect was close to significance. Finally, a regression on the difference between the plausibility of the most plausible colored interpretation and the literal one showed a significant main effect of condition, which was larger than the interaction-effect of out-of-key*condition in the first regression, indicating that the increased plausibility of colored interpretations plays a substantial part.

Furthermore, a comparison with the plausibility ratings for the fillers, which are practically equal in all conditions, with standard deviations half as high as those for the plausibility ratings for targets, shows that the deviant plausibility ratings for literal interpretations in out-of-key versions are not due to an effect of mere difficulty of processing the out-of-key notes but must be related to the ambiguity of the manipulated words and the plausibility of the alternative interpretations.

Unexpectedly, loud rests did not affect the interpretation of target sentences significantly, although the plausibility-ratings for literal interpretations of these target sentences do show a decrease in the loud-rest condition which is not visible in the ratings for fillers. Probably, the lack of significance is due to stimulus creation. Because of unnatural divisions between phonemes, and the absence of other prosodic cues, the breaks between words and word parts caused by the digitally-created loud rests might have sounded too artificial. Subsequently, the obstruction might be attributed to stimulus editing and not to the intention of the singer. Future research should probably make use of sentences sung with naturally performed loud rests, intended as pause for effect.

Given the fact that the results of this experiment show an interaction of musically syntactic surprises and linguistic semantics, it would be interesting to investigate whether this kind of prosodic interaction would elicit deviant brain potentials. Possibly, the N400 or the N5 will be affected. Concerning loud rests the form of the sentences used here would allow for an investigation of the difference between a within-phrasal loud rest (such as the last one in each sentence) and a between-phrasal loud rest (such as the first one). Possibly, a larger mismatch negativity would be found in the former, which would indicate that indeed there was a language affect, although in this case it did not affect meaning.

Concerning the out-of-key note further research is required as to whether its effect is just a matter of accentuation, or whether there is an intrinsic meaning to it, either related to specific pitch-related qualia (Arthur, 2018), or to the 'wrongness' or peculiarity of the note. Please note that a melodic accent through 'highest pitch' is a result of backward priming, while the out-of-keyness of a note is immediately perceivable.

Finally, as the MFH explains the effect of out-of-key notes in this experiment as a result of accentuation by obstruction, it would be very interesting to test whether simultaneously presented violations of both musical syntactical and linguistic syntactical expectancies are also able to affect language processing in a positive way, or whether in these cases the obstruction is too large to overcome and to make sense of it.

Conclusion

The results of this experiment show that out-of-key notes well-aligned with ambiguous words enhance the plausibility of colored interpretations of sung sentences containing such ambiguous words. Hence, these results support the hypotheses that VMSEs can work as prosodic accents, and that VMSEs working as prosodic accents can interact with semantics. Possible prosodic effects should therefore always be taken into account in research concerning the interaction between VMSEs.and the processing of language. The results of this experiment do not support the hypothesis that loud-rests affect the plausibility of colored interpretations as well. Possibly due to stimulus creation.

Acknowledgements. I would like to thank my supervisors prof. dr. E. Wennekes, dr. F. Hakemulder and dr. R. Willems for their comments; my colleagues from UiL OTS for their help with Lime Survey and statistics and giving me the opportunity to present and discuss part of these results in an ELiTu Talk; Lynn Eekhof for pretesting the sentences; Christan Grotenbreg for recording the stimuli and digitally manipulating the recordings: several bookshop owners and web hosts for helping me finding participants; my participants for completing the survey; and NWO, the Dutch Government, and SG het Rhedens for granting me the opportunity to avail myself of a PhD scholarship for teachers.

References

Arthur, C. (2018) A perceptual study of scale-degree qualia in context. *Music Perception*, 35(3), 295-314.

Bouwer, F., Schotanus, Y. P., Sadakata, M., Mullensiefen, D & Schaefer, R. (in preparation) Measuring musical sophistication in the low countries; validation of a Gold MSI translation in Dutch.

Burns, L. (2000) Analytic methodologies for rock music: harmony and voice leading strategies in Tori Amos's 'Crucify'. Everett, W. (Ed) *Expression in pop-rock music: a collection of critical and analytical essays*. New York/London: Garland.

Carrus, E., Koelsch, S., & Bhattacharya, J. (2011). Shadows of music–language interaction on low frequency brain oscillatory patterns. *Brain and Language*, *119*(1), 50–57.

Carrus, E., Pearce, M. T., & Bhattacharya, J. (2013). Melodic pitch expectation interacts with neural responses to syntactic but not semantic violations. *Cortex*, *49*(8), 2186–2200. http://doi.org/10.1016/j.cortex.2012.08.024

Curtis, M.E. & Bharucha, J.J. (2003) *Tonal Violations Interact with Lexical Processing: Evidence from Cross-modal Priming*. Paper presented at the annual conference of the Society for Music Perception and Cognition, Las Vegas, Nevada.

Gordon, R. L., Jacobs, M. S., Schuele, C. M., McAuley, J.D. (2015) Perspectives on the rhythm-grammar link and its implications for typical and atypical language development. *Annuals of the New York Academy of Science*, 1337, 16-25.

Gordon, R. L., Magne, C. L. & Large, E. W. (2011). EEG correlates of song prosody: a new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology,* 2, 352.

Hakemulder, J.F. (2004). Foregrounding and its effect on readers' perception. *Discourse Processes, 38*(2), 193-218.

Hakemulder, J. (2008). The more you see, the more you get: How spectators' use their limited capacity for attention in responses to formal aspects of film. In J. Auraher & W. van Peer (Eds), *New beginnings in literary studies* (pp. 332-352). Newcastle: Cambridge Scholars Publishing.

Honing, H. J., Ladinig, O. Winkler I. & Háden G. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials (ERP). *Annals of the New York Academy of Sciences*, nr. 1169: *The Neurosciences and Music: Disorders and Plasticity*, p. 93-96.

Honing, H.J. (2009) *Iedereen is muzikaal: wat we weten over het luisteren naar muziek.* Amsterdam: Nieuw Amsterdam.

Huron, D. (2006). *Sweet Anticipation*, Cambridge, MA: MIT Press.

Jung, H., Sontag, S., Park, Y.S. & Loui, P. (2015) Rhythmic Effects of Syntax Processing in Music and Language. *Frontiers in Psychology, 6,* 1762.

Kunert, R., Willems, R. M., & Hagoort, P. (2016). Language influences music harmony perception: effects of shared syntactic integration resources beyond attention. *Royal Society Open Science*, *3*(2), 150685.

Kunert, R. (2017) Music and language comprehension in the brain. PhD Thesis, Radboud University Nijmegen, Nijmegen. Retreived, oktober 2017 from: http://www.mpi.nl/publications/escidoc-2353161.

LaCroix, A. N., Diaz, A. F. & Rogalsky, C. (2015) The relationship between the neural computations for speech and music perception is context-dependent: an activation likelihood estimate study. *Frontiers in Psychology,* 6, 1138.

Lidji, P., Jolicoeur, P., Moreau, P., Kolinsky, R. & Peretz, I. (2009) Integrated Preattentive Processing of Vowel and Pitch: A Mismatch Negativity Study. *Annals of the New York Academy of Sciences* , 1169, 481-484.

London, 1993. Loud rests and other strange metric phenomena (or, meter as heard). *Music Theory online*. Http://mto.societymusictheory.org/issues/mto.93.0.2/mto.93.0.2london.art.

Miall, D. & Kuiken, D. (1994). Foregrounding, Defamiliarization, and Affect Response to Literary Stories. *Poetics*, 22, 389-407.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE,* 9(2). Questionnaire retrieved from: https://www.gold.ac.uk/music-mind-brain/gold-msi/download/.

Patel, A.D. (2003). Language, music, syntax and the brain. *Nature Neuroscience* 6(7):674-681.

Perruchet, P. & Poulin-Charronnat, B., (2013) Challenging prior evidence for a shared syntactic processor for language and music, *Psychonomic bulletin and review*, Austin TX, (online 2012)

Poulin-Charronnat, B., Bigand, E., Madurell F. & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition,* 94(3), 67-78.

Quené, H. & Port (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62 (1), 1-13.

Quené, H. & Van den Bergh, H. (2004) On Multi-Level Modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43 (1-2), 103-121.

Raman, R., Herndon, K., & Dowling, W. J. (2016) Effects of familiarity, Key Membership and Interval Size on Perceiving Wrong Notes in Melodies, In: Zanto, T.P. (Ed.), *Proceedings of the 14th International Conference on Music Perception & Cognition (ICMPC14*), July 2016, San Francisco. Adelaide: Causal Productions, p. 361. 756-759

Sammler, D., Baird, A., Valabrègue, R., Clément, S., Dupont, S., Belin, P. & Samson, S. (2010) The Relationship of Lyrics and Tunes in the Processing of Unfamiliar Songs: A Functional Magnetic Resonance Adaptation Study. *Journal of Neuroscience*, 30(10), 3572–3578.

Schotanus, Y. P. (2015) The musical foregrounding hypothesis: How music influences the perception of sung language, Ginsborg, J., Lamont, A., Philips, M. & Bramley, S. (Editors) *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, 17-22 August 2015, Manchester, UK.

Slevc L., Rosenberg, J., Patel, A. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic bulletin & review*, 16 (2), 374-381.

Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*(5), 1169–1178. http://doi.org/10.1093/cercor/bhm149 Winkler, I., Haden, G., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. Proceedings of the National Academy of Sciences, 106, 2468-2471.

Tyler, J. (2014) Prosody and the interpretation of hierarchically ambiguous discourse. Discourse Processes, 51: 8, 656-687.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Off-Beat Phrasing and the Interpretation of the Singer's Tone of Voice

Yke Schotanus[1]

[1] Institute for Cultural Inquiry (ICON), Utrecht University, The Netherlands

[1]schotschrift.teksten@planet.nl

Abstract

Both music and lyrics are thought to affect the emotional meaning of a song, but to date it is not exactly clear how and to what extent music does so. Possibly, timing is an important factor. Both singers and composers often create off-beat onsets of important linguistic events, such as the first stressed syllable in a phrase (henceforward: phrase onset). However, off-beat events are more difficult to process, which is hypothesized to cause a foregrounding effect, which would affect the interpretation of the singer's state of mind, his or her intentions, and the meaning of the words. An online listening experiment was created to test this hypothesis. Thirty participants listened to 27 piano-accompanied sung sentences, consisting of five or six syllables, some of them statements, some of them questions, imperatives, or incomplete sentences. In nine of them the phrase-onset was on-beat, in 9 it was early and in the remaining 9 it was late. After each sentence participants rated 10 items concerning the way words, music and singer are perceived. Three factors emerged from a factor analysis on these ratings. Regressions on these factors show that they are hardly affected by timing. Surprisingly, also sentence type did just marginally affect one factor. This indicates that music is more important in communicating aspects of meaning such as sincerity, self-security or compellingness.

Introduction

Music is thought to affect the meaning of a song, or at least the emotional meaning. Ali and Peynircioğly (2006) even conclude that music has more impact than lyrics when it comes to perceived happiness. However, to date it is not exactly clear in which ways music does so, although a wealth of studies have addressed the question as to whether pure emotions can intentionally be expressed in such a way that the listener perceives them (for example: Gabrielsson & Juslin, 1994; Scherer et al., 2017). To my knowledge there are just a few studies regarding the way music influences perceived affect, for example perceived sincerity, sociability or submissiveness (Huron, Kinney & Precoda, 2006; and Shanahan & Huron, 2014). Nevertheless, perceived affect, is very important feature in popular music especially when it comes to perceived authenticity (Bracket, 1995). According to Pattison (2015) off-beat phrasing (starting the first stressed syllable of a linguistic phrase on a weak beat) is a useful technique to manipulate affect. And indeed, both singers and composers often create such phrase onsets (Temperley, 2001). Pattison argues that off-beat phrases sound less stable, which makes the listener feel that there might be some subtext to the lyrics, for example because the singer is upset.

These presumptions are in line with several theories. In the first place Dynamic Attending Theory (DAT) (Jones, 1976; Large & Jones, 1999), assumes that our attention oscillates in accordance with a given speech rhythm or musical rhythm. Thus, in Western listeners attention is optimal at strong beats, and in a stress-timed language like Dutch it is optimal at stressed syllables. In line with that, several studies have shown that stressed syllables presented off-beat were more difficult to process than stressed syllables presented on-beat (Gordon, 2011; Quené & Port, 2005). Moreover, off-beat phrasings often cause loud rests, or on-beat silences (London, 1993). Such loud rests are known to cause substantial brain activity (Honing et al., 2009), which might distract from language processing. So, both Dynamic Attending and loud-rest processing might obstruct the processing of the words.

However, there is more to it. In literary research, obstructions of the processing of language through stylistic features is known to enhance the perceived salience of either the message or the wording if and only if the obstruction can be interpreted as meaningful (Miall & Kuiken, 1994; Hakemulder, 2004). In line with that The Musical Foregrounding Hypothesis (Schotanus, 2015) states that obstructions of language processing caused by musical events such as off-beat phrasings are, if possible, interpreted as meaningful prosodic cues. The timing of the stimulus, or the confusing of the listener, might be attributed to the intentions or the emotional state of the singer, who might be perceived to be hasty, compelling, lingering, or uncertain, or something like that. Furthermore, the salience of either words or melody might be increased and positively valued. These assumptions were tested in an online listen experiment, with on-beat phrases, early phrases (one eighth before the down beat), and late phrases (on the second beat). On-beat phrases were hypothesized to sound more convincing. The singer will be perceived as more sincere, less insecure, less emotional, and less compelling. On the other hand, on beat phrases might be more predictable and stable and therefore less interesting, and less emotionally loaded (Menon & Levitin, 2005).

Method

Participants and Procedure

Thirty participants (between 18 and 87 years of age; M 55.4; SD = 17.8) were recruited through social media and websites such as proefbunny.nl. They listened at a self-chosen time and place to 27 sung sentences, preceded and accompanied by a piano accompaniment (total track duration about 12 seconds). After each sentence they rated on a seven point likert scale whether they agreed with the statements that: the singer was sincere, insecure, and sounded compelling; whether the lyrics were emotional, superficial, and sounded natural; whether the fragment sounded loaded and energetic; and whether the melody was interesting, and music and lyrics were a good match. Although the hypotheses mainly concerned the interpretation of the singers tone of voice, there were just a few questions that directly addressed the singer, in order to mask the aim of the experiment. A factor analysis was planned to unveil the connections between the ratings. As participants had to rate 27 stimuli, the number of questions had to be limited.

Half way through the experiment participants answered a few questions on their musical and literary training. Musical training was measured with a translation in Dutch of the Musical training subscale of the Gold MSI (ranging from 7 to 49, Bouwer et al., in preparation; Müllensiefen et al., 2015). Participants score ranged between 7 and 45 (Mean 24,5; SD 9.5). From a factor analysis on the literary-training items, two factors emerged: literary training and disinterest in wording. Musical training, literary training and disinterest in wording were used as covariates in the regressions. The whole experiment lasted about half an hour, participants who completed the whole questionnaire and left behind their address at the end of it received a 5€ book gift voucher.

Stimuli

27 sentences were sung to 9 melodies. There were 13 directives, 5 questions, 6 statements, and 3 elliptic sentences (for example an address, or: Door red, shutters green. All sentences consisted of 3 metrical feet; 18 of Trochees, 9 of Iambs. Hence, the former consisted of 5 syllables and the latter of 6. The melodies were created by the author to express feelings appropriate to at least one of the sentences sung to it, and to vary in contour, key and harmony, as off-beat phrasing was thought to have a general effect independently of text and music. They also varied in measure: 6 melodies were in three-four time, three in four-four time. The sung melodies were preceded and accompanied by piano music improvised by Christan Grotenbreg, a professional musician. He was asked to create different kinds of accompaniments, whether or not using the harmonies suggested by the author, but always establishing a beat. Hence, all sentences would have a clear and similar rhythmic structure, aligned to a well established beat, but would sound relatively interesting and ecologically valid, as far as possible given the atomic design of the study.

In most cases neither the melody nor the accompaniment ended on a tonic. In order to measure harmonic closure for all of the melodies forty Amazon Mechanical Turk workers who did not speak Dutch and could not address the contents of the sentences were presented an example of each melody and were asked to rate whether the fragment sounded like it was finished, whether there was some remaining musical tension after the last note, and whether in their minds they heard some final notes they would expect to follow that last one. A Principal Components Factor analysis resulted in one factor representing non-closure (see Appendix).

All sentences were sung once, but were digitally edited in three different ways: early, on beat and late (see Figure 1.). In the on-beat version the three stressed syllables in the sentence were aligned with the first beat, in the early versions the onsets of the stressed syllables were timed one-eighth note before the first beat, and in the late versions they were aligned with the second beat. There were no fillers created, as it is impossible not to time either off beat or on beat. Afterwards, the 81 musical fragments, were distributed over three sets of stimuli, such that each participant heard each sentence once, and each melody once in each version.

Figure 1. Stimulus example. One sentence 'Liefste, liefste, blijf' [Darlin', Darlin', stay.] In three versions: Early (A100); On beat (A100a) and Late (A100b).

All sentences were sung by the author (a male baritone), and recorded by Christan Grotenbreg in his studio. The piano intros were improvised by Christan Grotenbreg on a keyboard connected to ProTools 10 (Desktop recording). The voice was recorded using a Neumann TLM 103 microphone, and an Avalon VT 737 SM amplifier. Digital conversions were conducted using Apogee Rosetta. To avoid confounds concerning purity and timing, voice-treatment software was used: Waves Tune, Renaissance Vox compression, and Oxford Eq.

Analysis

The ratings were analyzed using Principal Axis Factoring with rotation (direct oblimin). Subsequently, crossed classified regression analyses were conducted on the factors using Mixed models.

Results

Factor analysis. After a Principal Axis factor analyses on the ten ratings for each sentence, three factors emerged with an eigenvalue higher than 1 (see Table 1.): Rightness, a combination of naturalness, sincerity, aesthetic valence, and, to a lesser extent, energy, emotionality, and self-security; Upsetness, a combination of insecurity, emotionality and a lack of energy; and Compellingness, a combination of emotional load, and compellingness.

Table 1. Factor analysis on the ten ratings per sentence[a].

	Rn[b]	Un[c]	Cn[d]
Initial Eigenvalue	3.51	1.85	1.04
% of variance predicted	35.12	18.47	10.38
Rotation sums square loadings	2.91	1.29	1.75
Factor loadings			
Lyrics sound natural	**.77**	-.19	.37
Singer seems to be sincere	**.72**	-.12	.38
Music and lyrics are good match	**.72**	-.19	.20
Melody is interesting	**.70**	-.19	.21
Singer is insecure	-.24	**.58**	
Fragment sounds energetic	.43	**-.57**	.22
Lyrics are emotional	.46	**.52**	.49
Fragment sounds loaded	.40	.48	**.77**
Singer sounds compelling	.26		**.67**
Lyrics are superficial	-.34		-.22

[a] KMO .78; $p < .001$; df 45; determinant .04
[b] Rightness
[c] Upsetness
[d] Compellingness

Descriptives and Regressions

Depending on Figure 2. Rightness is rated relatively high and Compellingness relatively low for on-beat versions. Upsetness ratings are relatively high for early versions, but the differences are smaller. Whether any of these results are significant or not, was tested in a series of regression analyses. First an intercept only model was tested on each factor in which random intercepts were estimated for participant, sentence and melody. After that models were created using the following factors and covariates: condition (early, on beat or late), meter (Iamb or Trochee), measure (binary or ternary), sentence type (directive, statement, ellipse, question), tempo (slow, mixed, fast), non-closure, musicianship, writing experience, disinterest in wording and interactions between these factors and covariates. Predictors that did not show a significant effect were left out, except for condition. Alternative regressions with melody (df = 8) as a predictor instead of a random factor are slightly less powerful. These models are similar, except that aspects of melody such as meter, measure, tempo and non-closure were redundant and had to be deleted, from these regressions, just as sentence type.

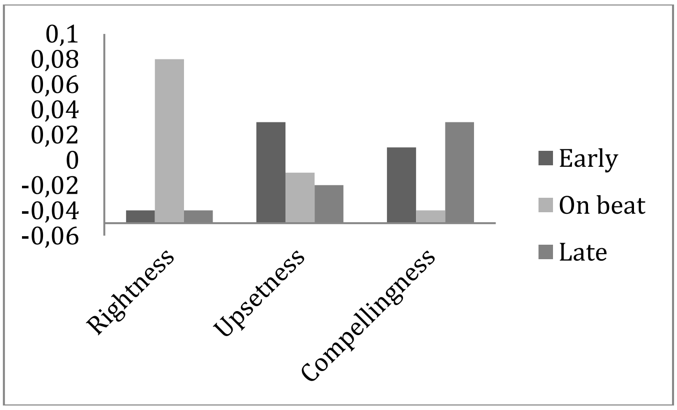


Figure 2. Factor Means per condition. Especially on-beat ratings for Rightness seem deviant. Corresponding SDs Rightness: 1,05; 0.99; 0.91; Upsetness: 1.02; 1.02; 0.96; and Compellingness: 1.0; 1.02; 1.02.

Table 2. Crossed classified linear regression on rightness.

Models	df	-2LL[a]	AIC[a]
Intercept only	5	2013.64	2023.64
Full model	14	1950.72	1978.72
Details		Estimate (SE)	F/Z
Fixed			
Intercept	1		2.43
Condition[b]	2		3.01[+]
Meter[b]	1		12.56**
Measure[b]	1		10.69*
Cond*Meter[b]	2		1.92
Writingexp.[b]	1		8.22**
Upsetness[b]	1		6.73*
Compellingness[b]	1		4.61*
Random			
Participant		0.19 (0.06)	3.48**
Sentence		0.08 (0.03)	2.40*
Melody		0.01 (0.02)	0.31

[a] Model fit indicators: -2 Log likelihood and Akaikes information criterion
[b] Estimates (SE), if not redundant, for condition: Early: 0.00 (0.11); On beat: 0.26* (0.11); for Meter: Trochee: -0.42* (0.16); for Measure: three-four time: -0.45* (0.14); Late*Trochee: 0.03 (0.14); On beat*Trochee: -0.25 (0.14); Writingexperience -0.25** (0.09); Upsetness 0.10* (0.04), Compellingness 0.07* (0.03).
[+] $p = 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Rightness. The effect of timing on rightness does not show a significant main effect unless an interaction with syllable count is included in the model, although even then the effect is only marginally significant (p = .05). Furthermore, sentence type, tempo, non-closure, musicianship nor disinterest in wording significantly affected Rightness.

Table 3. Crossed classified linear regression on upsetness.

Models	df	-2LL[a]	AIC[a]
Intercept only		1791. 39	1801.39
Full Model[b]	14	1717.57	1743.57
Details full model		Estimate	F / Z
Fixed			
Condition[c]	2		0.39
Measure[c]	1		11.43**
Tempo[c]	2		92.31***
Non-closure[c]	5		4.75*
Musicianship[c]	1		9.11**
Writing experience[c]	1		8.49**
Disinterest wording[c]	1		6.92**
Random			
Participant		0.06 (0.19)	2.98**
Sentence		0.05 (0.17)	2.73**

[a] Information criteria: -2 Log likelihood, and Akaike's information criterion (-2ll adjusted for model complexity)
[b] Model without random intercept for Melody; model with Melody, but without nonclosure was slightly weaker: df 13, -2ll: 1720.31; AIC: 1746.31
[c] Estimates (Standard Error), if not redundant: Late 0.03 (0.06); On beat -0.02 (0.06); Measure=3: -0.35** (0.10); Slow: 1.55*** (0.13); Mixed tempo 0.35* (0.13); Non-closure: -0.30* (0.14); Musicianship -0.02** (0.05); Writing experience -0.15** (0.05); Disinterest wording 0.08** (0.03).

Upsetness. Timing does not affect Upsetness significantly, nor do sentence type, meter, Rightness and Compellingness. Tempo is by far the main predictor, indicating that slower melodies sound more upset than faster ones. On the other hand, three-four time, and non-closure make the fragments sound more upset. Finally, people who are either trained as writer or as musician tend to give slightly lower upsetness ratings, while people who are not interested in wording give higher ones.

Compellingness. In both Model A (p = .064) and B (see Table 4) the main effect of timing on compellingness ratings approaches significance. However the interaction between timing and musicianship is significant in both models, indicating that musicians tend to rate on-beat sentences as less compelling than late ones. Musicians also tend to rate sentences in general as less compelling. Furthermore, iambs seem to be more compelling than trochees, and quick melodies more compelling than slow ones. Writing experience and measure show no significant effect. Finally, open ended melodies tend to increase compellingness ratings, an effect that approaches significance.

Discussion

The main aim of this study was to test whether off-beat phrasing, as an example of musical instability, affects the emotional meaning of a sung sentence, especially considering the singer's tone of voice. The results indicate that indeed there are differences between early, on-beat and late sentences but they are only marginally significant. However, the results do show that music does affect the interpretation of the singer's tone of voice, and indicate that musical stability might play an important part.

Table 4. Crossed classified linear regression on compellingness.

Models	df	-2LL[a]	AIC[a]	
Intercept only	5	1979.58	1989.58	
Model	14	1910.08	1946.08	
Details Model B		Estimate	F / Z	*p*
Fixed				
Intercept	1		1.25	.271
Condition[b]	2		2.65	.071
Meter[b]	1		10.79	.003
Tempo[b]	2		4.14	.027
Nonclosure[b]	1		3.85	.060
Musicianship[b]	1		4,73	.038
Musician.*cond.[b]	2		4.21	.015
Rightness	1		5.56	.019
Sentence type[b]	3		3.42	.031
Random				
Participant		0.15 (0.04)	3.34	.001
Sentence		0.08 (0.03)	2.95	.003
Melody		redundant		

[a] AIC: 1946.12; variables in model: Melody, Condition, Musicianship, Musicianship*condition, Rightness

[b] Estimates (Standard Error) if not redundant: Late -0.01 (0.18); On beat 0.34[+] (0.18); Trochee: 0.84** (0.25); Slow: -0.73* (0.33); Mixed tempo 0.21 (0.24); Non-closure: 0.39[+] (0.20); Musicianship: -0.01 (0.01); Late*Musicianship: 0.00 (0.01); On beat*Musicianship: -0.02* (0.01); Sentence type: Directive: 0.38* (0.17); Statement: 0.27 (0.19); Incomplete -0.26 (0.26); Rightness: -0.08* (0.03).

[+] $p \leq 0.06$; * $p < 0.05$; ** $p < 0.01$; $p < 0.001$

After each sentence participants rated to what extent they agreed with ten short statements. A factor analysis on these ratings resulted in three factors. The first factor was called rightness, as it seems to indicate that the stimulus was both aesthetically and morally right. Rightness (or 'just rightness') is known as a factor in aesthetics (Aaftink, 2014), but without the moral implication of sincerity. The second factor was called upsetness, as it seems to indicate that the singer was insecure because he was emotional and lacked energy. Finally, the third factor was called compellingness, as compellingness was the main contributor.

It is not easy to connect these factors to the original predictions, but given the part that naturalness and sincerity play in rightness, and the part that insecurity and compellingness play in upsetness and compellingness, one would expect rightness ratings to be higher on beat, and upsetness and compellingness to be higher off beat, which is indeed the pattern shown in Figure 2. However, aesthetic valence was hypothesized to be related to off-beat phrasings which is not in accordance with this pattern. Furthermore, the pattern is not convincingly significant.

The regression on rightness does show a close to significant (p = .050) main effect of condition, but only if an insignificant interaction with meter is involved indicating that trochaic sentences, i.e. melodies without a pick-up note, are rated less right in on-beat versions. Hence, the pick-up note in Iambic sentences might play a substantial part in the effect of condition. One explanation might be that the pick-up note accentuates the beat in the melody and thus the misalignment with the accompaniment. Another explanation might be that in late versions the pick-up notes are dissonant with the accompaniment. This would be an extra dissonant on top of the one often caused by the stressed syllable in early sentences. However, these dissonants do not occur in each sentence, and if they do they occur mostly either in early or in late sentences, while there is no difference in rightness ratings between early and late ones. Furthermore, in popular music people it is normal that melodies are syncopated while the accompaniment is not, and that syncopations are perceived in a different way than in so-called 'classical music' (Temperley, 2001, 239-247; Burns, 2000). Therefore, it is more likely that the main effect of condition is due to Dynamic attending and/or loud rests.

The relatively small effect of condition in all regressions might be due to the fact that the complete melodic line was shifted in relation to the accompaniment. Gordon (2011) found that beat tracking shifts to the rhythm of stressed syllables instead of that of strong beats if stressed syllables consistently occur on weak beats. The effect size could also have been attenuated because the perceived instability might be interpreted and valued in very different ways, dependent on the specific sentence and the melody. Probably, the items to be rated should either be sentence specific, or more clearly aimed at measuring a difference between musical and mental stability, and at clearer factors (for example: convincingness, credibility, sincerity, straightness (i.e. absence of subtext), naturalness, stability, self-security, unrest, upsetness, compellingness, annoyingness, etc.).

Although the regression analyses presented here, were slightly more powerful than regressions with melody as a fixed factor, they have to be interpreted with caution as neither the aspects of melody nor the categories of sentence type are counterbalanced sufficiently, while melody is. However, in an explorative way, these regressions are more informative than the regression with melody, as they give indications why certain melodies make a singer sound more sincere, more upset, or more compelling. Tempo (or note rate), seems to affect compellingness and upsetness in opposite directions, probably as an indicator of energy. Meter seems to affect compellingness and rightness in opposite directions, possibly because the pick-up note in Iambs enhances predictability and stability. Either the familiarity with four-four time beats, or a general preference for binary structures (Temperley, 2001, 39) might explain why a three-four time decreases both rightness and upsetness ratings. Finally, as dominant-endings are perceived as relatively open it might come as no surprise that non closure increases compellingness, and increases upsetness, although one would expect upsetness not to be associated with complete closure.

Furthermore, the fact that sentence type had to be deleted from all models except one is rather surprising. Admittedly, just as several aspects of melody, sentence type was not properly counterbalanced, but nevertheless the results of this experiment indicate that melody affects the emotional meaning of a song and the interpretation of the tone of voice of the singer more strongly than sentence type. Even the compellingness of directives compared to questions or statements, seems to be in no way comparable to the effect of meter.

Please note that this is just about sentence 'type', the effect of the text of each sentence specifically is integrated in the random effect of sentence. Nevertheless, given the relatively strong effect of melody and accompaniment on all ratings, it is clear that music can affect the interpretation of the singer's intentions, his tone of voice, and even his sincerity.

Conclusion

The results of the current experiment only show small effects of timing. Moreover, these effects cannot unambiguously be related to timing as such. The effect of dissonance as a result of shifting the melody in relation to the accompaniment might also play a part. On the other hand, several aspects of the design might have attenuated the effect of timing. Therefore, future research might search for other ways to create somehow ecologically valid, varied, and attractive combinations of melody and accompaniment, that would not cause dissonances (presumably a simple vamp, or a djembe beat); to create a more consistent set of statements to be rated; and/or to create stimuli in which not all

An important secondary finding of this study is that music affects the interpretation of the singer's tone of voice, his state of mind, his intentions and his sincerity, and subsequently will affect the emotional meaning of a song. Music even seems to overrule the effect of an important linguistic factor such as sentence type. As the various aspects of music and language affecting the ratings were just meant to be able to generalize the hypothesized effect of timing, they were not properly counterbalanced and their effects have to be interpreted with caution. Nevertheless, they show the need of research about the way music affects the interpretation of song, and hopefully inspire new experiments. Creating authenticity is very important in popular music, and tone of voice is crucial in music therapy, advertisement and games.

Acknowledgements. I would like to thank my supervisors Emile Wennekes, Frank Hakemulder and Roel Willems, for their comments; my colleagues from UIL OTS for their help with Lime Survey and statistics and giving me the opportunity to present and discuss part of these results in an ELiTu Talk; my colleagues from musicology for their critical comments on the stimuli; Christan Grotenbreg for recording the stimuli and digitally manipulating the recordings; several bookshop owners and web hosts for helping me finding participants; my participants for completing the survey; and NWO, the Dutch Government, and SG het Rhedens for granting me the opportunity to avail myself of a PhD scholarship for teachers.

References

Aaftink, C. (2014) *Kaleidoscope: A phenomenological-empirical study of beauty. University of Alberta* (Doctoral dissertation). Retrieved from: era.library.ualberta/files/1c18dg58c/Aaftink_Cathelein_201409_PhD.pdf.

Ali, S. O. & Z. F. Peynircioğly (2006) Songs and emotions: are lyrics and melodies equal partners. *Psychology of music, 34(4),* 511-534.

Bouwer, F., Schotanus, Y. P., Sadakata, M., Mullensiefen, D & Schaefer, R. (in preparation) *Measuring musical sophistication in the low countries; validation of a Gold MSI translation in Dutch*.

Brackett, D. (1995) *Interpreting popular music*. Cambridge/New York/Melbourne: Cambridge University Press.

Burns, L. (2000) Analytic methodologies for rock music: harmony and voice leading strategies in Tori Amos's 'Crucify'. Everett, W. (Ed) *Expression in pop-rock music: a collection of critical and analytical essays*. New York/London: Garland.

Gabrielsson, A. & Juslin, P. N. (1996) Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, 24, 68-91.

Gordon, R. L., Magne, C. L. & Large, E. W. (2011). EEG correlates of song prosody: a new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology,* 2, 352.

Hakemulder, J.F. (2004). Foregrounding and its effect on readers' perception. *Discourse Processes, 38*(2), 193-218.

Honing, H. J., Ladinig, O. Winkler I. & Háden G. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials (ERP). *Annals of the New York Academy of Sciences*, nr. 1169: *The Neurosciences and Music: Disorders and Plasticity*, p. 93-96.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review,* 1, 170-177.

Jones, M.R. (1976). Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological Review*, 83, 323-3.

Jung, H., Sontag, S., Park, Y.S. & Loui, P. (2015) Rhythmic Effects of Syntax Processing in Music and Language. *Frontiers in Psychology, 6,* 1762.

Large, E.W. & Jones, M.R. (1999). The dynamics of attending: how people track time varying events. *Psychological Review*, 106, 119–159.

London, J. M. (1993) Loud rests and other strange metric phenomena (or, meter as heard). *Music Theory online*. Http://mto.societymusictheory.org/issues/mto.93.0.2/mto.93.0.2london.art.

Menon, V. & Levitin, D. (2005). The rewards of music listening: Response and physiological connectivity of the mesolymbic system. *NeuroImage*, 28(1), 175-184.

Miall, D. & Kuiken, D. (1994). Foregrounding, Defamiliarization, and Affect Response to Literary Stories. *Poetics*, 22, 389-407.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLoS ONE,* 9(2). Questionnaire retrieved from: https://www.gold.ac.uk/music-mind-brain/gold-msi/download/.

Pattison, P. (2015) Lesson 46: Phrasing. *Songwriting: writing the lyrics.* Berkelee: Coursera. Retreived, March, 2015. https://www.coursera.org/learn/songwriting-lyrics/lecture/3BtUo/phrasing.

Quené, H. & Port (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62 (1), 1-13.

Scherer, K., Trznadel, S., Fantini, B. & Sundberg, J. (2017). Recognizing emotions in the singing voice. *Pschychomusicology: Music, Mind and Brain*, 27(4), 244-255.

Shanahan, D., & Huron, D. (2014). Heroes and villains: The relationship between pitch tessitura and sociability of operatic characters. *Empirical Musicology Review,* 9, 46-59.

Schotanus, Y. P. (2015) The musical foregrounding hypothesis: How music influences the perception of sung language, Ginsborg, J., Lamont, A., Philips, M. & Bramley, S. (Editors) *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music*, 17-22 August 2015, Manchester, UK.

Temperley, D. (2001) *The cognition of basic musical structures*. Camebridge, MA/London, UK: MIT.

Appendix

Examples of the remaining eight melodies (Figure 3, for the first one see Figure 1), and the non-closure ratings for each of them, plus the distribution of sentence-type, tempo, meter and measure across melodies (Table 5.)

Non-closure ratings were largely in line with musical theory. A fragment ending on the dominant (220) received high non-closure ratings and a fragment in which both melody and accompaniment ended on a tonic received the lowest (210). However, surprisingly, fragment 190, with the chord progression Em-Em-Em(add2)-Em(add2)-Em-A-F#/A#, also received a very low non-closure rating, even by musically trained listeners, although it ends with an out-of-key chord.

Table 5. Mean non-closure rating, meter, measure, tempo and sentencetype per melody. Features are not counterbalanced.

Melody	Non-clos.	Meter	Tempo	Measure	Sent.[1]
100	0.13	Trochee	Low	4/4	d/s/s
130	0.22	Trochee	Moderate	3/4	d/d/d
150	0.48	Trochee	Low	3/4	d/s/q
170	0.10	Trochee	Moderate	4/4	d/d/q
190	-0.28	Trochee	Low	3/4	d/d/q
210	-0.69	Trochee	Low	3/4	e/e/s
220	0.44	Iamb	High	3/4	d/q/s
250	-0.28	Iamb	High	4/4	d/d/d
280	-0.13	Iamb	Moderate	3/4	q/s/e

[1] Sentence type: d = directive; s= statement; q = question, e = ellipse

Figure 3. Examples of melodies: 130, 150, 170, 190, 210, 220,250 and 280 in on-beat position. For melody 100, see Figure 1.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Dissecting the Effects of Working Memory, General Fluid Intelligence, and Socioeconomic Status on Musical Sophistication

Daniel Shanahan,[1] Emily M. Elliott,[2] David John Baker,[3]

Juan Ventura,[4] Elizabeth Monzingo,[5] Hailey Holt,[6] and Hannah Keller[7]

[1] *Ohio State University, United States*
[2,3,4,6] *Louisiana State University, United States*
[5] *Seton Hall University, United States*

[1]shanahan.37@osu.edu, [2]eelliott@lsu.edu, [3]dbake29@lsu.edu,

[4]javentu2014@gmail.com, [5]emonzi1@lsu.edu, [6]haileyholt27@gmail.com, [7]hkelle9@lsu.edu

Abstract

An increasing amount of literature has sought to clarify many earlier findings between music and intelligence, such as the so-called "Mozart effect". A large portion of this literature has focused on the relationship between musical training and various cognitive measures. Both correlational (Slevc et al., 2016; Talamini et al., 2016; Talamini et al. 2017; Schellenberg, 2006; Degé et al., 2011; Ruthsatz et al., 2008) and experimental studies (Schellenberg, 2004; Moreno et al., 2011) suggest that some sort of relationship exists between music and cognitive abilities. Recent studies, such as Swaminathan et al., 2017, have argued that differences in cognitive ability later in life may be remnants of a selection bias early on where intelligent children who have a high aptitude for music self select into studying music, which is then further confounded by factors such as socioeconomic status. To further investigate this, they found evidence suggesting that musical aptitude is likely a better explanation than musical training for explaining differences in general fluid intelligence. This study seeks to replicate and extend these findings, suggesting that differences in musical aptitude and socioeconomic status can serve as better predictors of individual differences in general fluid intelligence. Whereas the previous study examined how the Musical Ear Test (Wallentin *et al.,* 2010) can predict general fluid intelligence, this study takes a more comprehensive approach, in which the dependent variable is a composite score of general fluid intelligence, and with a measurement of musicality that is arguably more ecologically valid and updated (the melodic memory and beat perception tasks from Goldsmiths Musical Sophistication Index; Müllensiefen *et al.,* 2014).Using these measurements, we were not able to replicate previous findings. General fluid intelligence and working memory capacity were significantly correlated, but neither were predicted by measurements of socioeconomic status. Years of formal training and the measurement of musical sophistication, however, were significantly predictive of both working memory and general fluid intelligence.

Introduction

A great deal of research in recent decades has focused on the relationship between music and intelligence. Most of these studies are correlational (Slevc et al., 2016; Talamini et al., 2016; Talamini et al., 2017; Schellenberg, 2006; Degé et al., 2011; Ruthsatz et al., 2008), but some have involved experimental studies studies as well (Schellenberg, 2004; Moreno et al., 2011). Recently, Swaminathan *et al.* (2017) argued that differences in cognitive ability later in life may be remnants of a selection bias early on where intelligent children who have a high aptitude for music self select into studying music, which is then further confounded by factors such as socioeconomic status. To further investigate this, the authors found evidence suggesting that musical aptitude is likely a better explanation than musical training for explaining differences in general fluid intelligence. Here, we attempt to replicate these findings, and also suggest that differences in musical aptitude and socioeconomic status can serve as better predictors of individual differences in general fluid intelligence. This study uses a number of metrics for musicality, such as the Goldsmiths Musical Sophistication test (specifically, the melodic memory and beat perception tasks)

Methodology

Participants. Two hundred fifty-four students enrolled at Louisiana State University completed the study. We recruited students, mainly in the Department of Psychology and the School of Music. A total of 8 participants were not eligible due to reporting hearing loss, and one individual was greater than 5 SD from the mean age and was removed. Finally, 6 participants were identified as univariate outliers in their performance on the recall portion of at least one of the working memory capacity measures. Thus, 239 participants met the criteria for inclusion. The eligible participants were between the ages of 17 and 38 (M = 20.64, SD = 3.23; 76 males; 1 person did not identify gender). Participants' formal years of musical training was rated on a 9-point likert scale (M = 4.18; SD=2.16) Participants volunteered, received course credit, or were paid $20, and the study was approved by the LSU Institutional Review Board.

Procedure. Participants were asked to perform the Gold-MSI self-report inventory, Tonal span, Symmetry span, Operation span, Gold-MSI beat perception test, Gold-MSI melodic memory test, Gold-MSI sound similarity test, Number Series, and the Raven's Advanced Progressive Matrices (RAPM). The entire experiment lasted roughly 90 minutes in total. Tonal span and sound similarity were not included in the current set of analyses.

Measures

Self-Report Data. In addition to the usual information about demographics, we asked three questions regarding socioeconomic status, similarly to the previous research

(Swaminathan et al., 2017). Specifically, we asked participants to rate their family income, as well as the highest education their father received, and the highest education their mother received, all on a likert scale of 1-9.

Goldsmith's Musical Sophistication Index (Gold-MSI). Participants completed a 38-item self-report inventory and questions consisted of free-response answers or choosing a selection on a Likert scale that ranged from 1-7 (see Müllensiefen et al., 2014).

Symmetry span (SSPAN). Participants completed a two-step symmetry judgment and were prompted to recall a visually-presented red square on a 4 X 4 matrix (Unsworth et al., 2005). In the symmetry judgment, participants were shown an 8 x 8 matrix with random squares filled in black. Participants had to decide if the black squares were symmetrical about the matrix's vertical axis and then click the screen. Next, they were shown a "yes" and "no" box and clicked on the appropriate box. Participants then saw a 4 X 4 matrix for 650 ms with one red square after each symmetry judgment. During square recall, participants recalled the location of each red square by clicking on the appropriate cell in serial order. Participants were provided practice trials to become familiar with the procedure. The test procedure included three trials of each list length (2-5 red squares), totaling 42 squares and 42 symmetry judgments.

Operation span (OSPAN). Participants completed a two-step math operation and then tried to remember a letter (F, H, J, K, L, N, P, Q, R, S, T, or Y) in an alternating sequence (Unsworth et al., 2005). In the math operation, participants saw an arithmetic problem (e.g., (4/4) – 1 = ?) and clicked the screen when they mentally solved the problem. Then, they were presented a digit on the next screen (e.g., 0) and had to click either the "true" or "false" box, depending on whether the presented answer matched the problem on the previous screen. The letter was presented visually for 1000ms after each math operation. During letter recall, participants saw a 4 x 3 matrix of all possible letters, each with its own check box. Letters were recalled in serial order by clicking on each letter's box in the appropriate order. Letter recall was untimed. Participants were provided practice trials, and the test procedure included three trials of each list length (3-7 letters), totaling 75 letters and 75 math operations.

Raven's Advanced Progressive Matrices (RAPM). Participants were presented a 3x3 matrix of geometric patterns with one pattern missing (Raven et al., 1998). Up to eight pattern choices were given at the bottom of the screen. Participants had to click the choice that correctly fit the pattern above. There were three blocks of 12 problems, totaling 36 problems. The items increased in difficulty across each block. A maximum of 5 min was allotted for each block, totaling 15 min. The final score was the total number of correct responses across the three blocks.

Number Series. Participants were shown 15 problems with a maximum time of 4.5 minutes (Thurstone, 1938). They began with four practice problems that were untimed, and were given instructions to solve the next problem by selecting the next number that would complete the series. Following the opportunity to ask any questions, they completed the timed experimental trials. The final score was the number of problems solved within the time limit.

Scores on Raven's Advanced Progressive Matrices and Number Series were converted to z-scores, and were used to assess general fluid intelligence. SSPAN and OSPAN were converted to z-scores as well, and were used as measurements of working memory capacity. The general musical sophistication scores from the Gold-MSI, along with the self-report question from the Gold-MSI about formal years of musical training, were used to assess levels of musical sophistication.

Analysis

Looking only at measurements of general fluid intelligence and musical sophistication, we find a significant – albeit not entirely strong– relationship ($p < .001$; df = 234; $R^2 = .07$). When years of formal musical training is included (with general fluid intelligence being predicted by both musical sophistication and years of formal musical training), however, formal years of musical training was the only significant predictor (see Table 1).

Table 1. Regression model with both general musical sophistication and years of formal musical training as predictors of general fluid intelligence ($p < .001$; df = 234; adjusted $R^2 = .089$).

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.2675	0.3454	-3.67	0.0003
GENERAL	0.0092	0.0053	1.75	0.0817
Formal	0.1237	0.0511	2.42	0.0163

In isolation, neither the family income nor the highest education level of mother or father were significant predictors of general fluid intelligence. When included in the larger model that also incorporated years of formal musical training and musical sophistication, they were similarly non-significant (see Table 2). Similarly, Swaminathan, *et al.* (2017), found that mother's education was predictive of music training, but not general fluid intelligence.

Table 2. Linear regression model including all variables predicting general fluid intelligence. Note that years of formal training is the only significant predictor.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.4866	0.4320	-3.44	0.0007
GENERAL	0.0088	0.0053	1.66	0.0976
Formal	0.1170	0.0514	2.28	0.0238
familyIncome	-0.0276	0.0371	-0.74	0.4580
highestFather	0.0491	0.0426	1.15	0.2507
highestMother	0.0274	0.0451	0.61	0.5439

A comparison of the models also demonstrates that years of formal training provides a significant improvement in the original model using only musical sophistication. The other metrics of socioeconomic status, however, do not improve the model.

Table 3. ANOVA table comparing the additive linear regression models predicting general fluid intelligence. Model 1 = general musical sophistication; Model 2 adds formal years of musical training; Model 3 adds family income; Model 4 adds highest level of education of the father, and Model 5 adds highest level of education of the mother.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	234	394.32				
2	233	384.65	1	9.66	5.85	0.0164
3	232	384.56	1	0.09	0.06	0.8108
4	231	380.69	1	3.87	2.34	0.1272
5	230	380.08	1	0.61	0.37	0.5439

Working Memory. General fluid intelligence and working memory often overlap quite a bit with one another, and our results are no different ($t = .005$ $df = 404$, $p = .99$) The results of the combined model, now predicting working memory rather than general fluid intelligence, is nearly identical. As can be seen in Table 4, both musical sophistication and years of formal training are significant predictors, but measurements of socioeconomic status are not (Adjusted $R^2 = .16$; $p < .001$; $df = 234$).

Table 4. Linear regression model including all variables predicting working memory. Note that both the general score of the Goldsmiths musical sophistication index and years of formal training are significant predictors.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.1503	0.2678	-4.30	0.0000
GENERAL	0.0089	0.0033	2.66	0.0085
Formal	0.0827	0.0324	2.56	0.0113
familyIncome	-0.0366	0.0230	-1.59	0.1131
highestFather	0.0446	0.0268	1.66	0.0973
highestMother	0.0018	0.0281	0.06	0.9499

Musical Sophistication. Although we would predict that family income would be predictive of both formal musical training and musical sophistication, we found no significant relationship with either (see Tables 5 and 6).

Table 5. Measurements of socioeconomic status as predictors of musical sophistication. There were no significant relationships.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	80.1783	4.5118	17.77	0.0000
familyIncome	-1.1088	0.5983	-1.85	0.0651
highestFather	0.6278	0.6932	0.91	0.3661
highestMother	0.8015	0.7294	1.10	0.2729

Table 6. Measurements of socioeconomic status as predictors of formal musical training. Again, there were no significant relationships.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.8211	0.4620	8.27	0.0000
familyIncome	-0.1162	0.0613	-1.90	0.0592
highestFather	0.0465	0.0710	0.66	0.5131
highestMother	0.1286	0.0747	1.72	0.0865

Discussion

Interestingly, formal years of musical training seems to covary quite a bit with the Goldsmiths musical sophistication index. Formal training is more predictive of both general fluid intelligence and working memory. The results suggest that measurements of socioeconomic status are not significant predictors of either general fluid intelligence or formal years of music training. This somewhat contradicts previous work on the subject (Swaminathan et al., 2017; although see Slevc et al., 2016 for a different outcome). Future work is needed with a broader sample of participants from a larger range of socioeconomic status before drawing any conclusions, as this sample was drawn from a population of university students.

Conclusion

Musical sophistication, general fluid intelligence, and working memory are all complex, multi-valenced aspects that can often be difficult to assess cleanly. We attempted to replicate previous work on this topic with a large sample of participants from both psychology and music subject pools, but were unable to replicate some previous previous findings (such as mother's education being significantly correlated with music training). Contrary to what we would expect, we found no significant relationships between socioeconomic status and years of formal musical training or musical sophistication.

Acknowledgements. The authors would like to thank Brian Ritter, Adam Rosado, Connor Davis, Chris Young, and Katherine Vukovics for helping run participants on this study.

References

Degé, F., Kubicek, C., & Schwarzer, G. (2011). Music lessons and intelligence: A relation mediated by executive functions. Music Perception: An Interdisciplinary Journal, 29(2), 195–201.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. PloS One, 9(2), e89642. http://dx.doi.org/10.1371/journal. pone.0089642

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. Psychological Science, 22(11), 1425–1433.

Raven, J., Raven, J. C., & Court, J. H. (1998). Manual for Raven's Progressive Matrices and Vocabulary Scales. Oxford, England: Oxford Psychologists Press.

Ruthsatz, J., Detterman, D., Griscom, W. S., & Cirullo, B. A. (2008). Becoming an expert in the musical domain: It takes more than just practice. Intelligence, 36(4), 330–338.

Schellenberg, E. G. (2004). Music lessons enhance IQ. Psychological Science, 15(8), 511-514. doi:10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. Journal of Educational Psychology, 98(2), 457.

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. Cognition, 152, 199-211. doi:10.1016/j.cognition.2016.03.017

Swaminathan, S., Schellenberg, E. G., & Khalil, S. (2017). Revisiting the association between music lessons and intelligence: Training effects or music aptitude? Intelligence, 62, 119-124. doi:10.1016/j.intell.2017.03.005

Talamini, F., Carretti, B., & Grassi, M. (2016). The Working Memory of Musicians and Nonmusicians. Music Perception: An Interdisciplinary Journal, 34(2), 183–191.

Talamini F., Altoè G., Carretti B., & Grassi M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. PLoS ONE, 12(10), e0186773. doi:10.1371/journal.pone.0186773

Thurstone, L. L. (1938). Primary Mental Abilities, By L.L. Thurstone. University of Chicago Press.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. Behavior Research Methods, 37(3), 498-505. doi:10.3758/BF03192720

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. Learning and Individual Differences, 20(3), 188-196. doi:10.1016/j.lindif.2010.02.004

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Towards the physical correlates of musical timbre(s)

Saleh Siddiq,[1] Christoph Reuter,[1] Isabella Czedik-Eysenberg,[1] Denis Knauf,[2]

[1] *Department of Musicology, University of Vienna, Austria*
[2] *Vienna University of Technology, Austria*

[1]saleh.siddiq@univie.ac.at, [1]christoph.reuter@univie.ac.at, [1]isabella.czedik-eysenberg@univie.ac.at

Abstract

Timbre is traditionally understood as a characteristic to identify or discriminate musical instruments. Many studies about musical instrument timbre ascertained two or three dimensions and concordantly established spectral and temporal features as the correlates of the first two dimensions. A potential third dimension was associated with a heterogeneous variety of features. As most studies reduced instruments to a single tone, the aforementioned timbre dimensions and their correlates were in reality obtained only for single sounds and not musical instruments.

The question then is, what are the physical correlates of musical timbre, once pitch and dynamics are taken into account, and which of these correlates apply to the discrimination or identification of musical instruments?

To investigate this issue, a hearing experiment was conducted in order to obtain the perceived dissimilarities of the sounds of five musical instruments, each presented in three different pitch and dynamic levels. Through multidimensional scaling, a spatial configuration of the data was calculated. In order to identify the physical correlates of the perceptual dimensions, correlations between the coordinate vectors of the spatial dimensions and a number of timbre descriptors obtained via audio signal analysis were calculated.

The results show pitch as the primary timbre factor with the harmonic spectrum and attack transients second and third respectively. Moreover it becomes clear that none of the isolated features applies as reliable discriminator of musical instruments.

Presumably, the core of the problem is the mismatch of research questions, aiming at musical instruments, and methods, based on isolated single sounds. To address this problem, a possible clarification of the terminology concerning timbre is proposed.

Introduction

Timbre researchers frequently deploy methods that project perceived timbre dissimilarities as spatial distances in a virtual, n-dimensional, Euclidean space (Grey, 1975; Krumhansl, 1989; Lakatos, 2000; McAdams, Winsberg, Donnadieu, & de Soete, 1995; Wedin & Goude, 1972; Wessel, 1979). The subjective timbre dissimilarities are obtained empirically through listening tasks. The question to be answered is: Which physical sound properties yield the most senseful explanations of the virtual space's main axes in terms of timbre relations? Traditionally, researchers attempted to answer the question via an educated guess, i.e. a subjective interpretation of the given data by the researcher himself.

Numerous studies gathered plenty of parameters to explain the dimensions of the virtual space. With great accordance, the dimensions 1 and 2 were linked to a scale of the spectral envelope and a scale of the temporal envelope, respectively (Caclin, McAdams, Smith, & Winsberg, 2005). The first one, usually referred to as sharpness or brightness, describes the spectral energy distribution. The latter one consists of underlying parameters that describe the attack transient of the sound. A possible third dimension is not as unanimously connected to any physical properties as the first two dimensions (Caclin et al., 2005). These results are questionable for the following two reasons: 1. As defined by the American National Standards Institute (ANSI), timbre "is that attribute of auditory sensation in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" (1960, p. 45). Although this definition was heavily criticized (e.g. Schouten in Plomp, 1970; Bregman, 1990; Patterson, Walters, Monaghan, & Goudrain, 2010), researchers, as a result, increasingly compared musical instruments at the same pitch and thus effectively reduced instruments to one single tone. 2. Practically every study searches for properties to discriminate or identify the timbre of musical instruments. But a single instrument can be played in various ways and with different forces (articulation and dynamics, respectively) and therefore produce a broad range of different timbres. As a result, reducing instruments to a single tone leads to a comparison of exactly that: single isolated tones or sounds, incidentally produced by one instrument or the other. Thus, research questions and methods do not match.

Research questions

Studies with several tones per instrument, which usually factor in at least pitch and/or dynamics, suggest a shift in the correlation of perceptual timbre dimensions and physical sound properties (Marozeau, de Cheveigne, McAdams, & Winsberg, 2003; Handel & Erickson, 2004). Thus, the main question of this study is: which physical sound properties are best suited to explain the perception-based spatial distribution of instrument sounds in a virtual space, once pitch and dynamics (i.e. a more realistic representation of musical instruments) are taken into account? Additionally it has to be asked, if any of these physical sound properties qualify as discriminator(s) of musical instruments.

Methods

In a hearing task, the subjective dissimilarities of the tested sounds were obtained. By means of a multidimensional scaling (MDS), which translates the perceived dissimilarities into spatial distances, a low-dimensional configuration was calculated. Through computational audio feature extraction, all tested sounds were examined in terms of their physical properties. These analyses returned a numerical value to describe the characteristic of a given property. To reveal underlying coherences, the correlation coefficients (Pearson's

r) between the coordinate vectors of the spatial axes and the column of values for the obtained audio signal features were calculated.

Stimuli

Five non-percussive classical orchestra instruments (bassoon, cello, clarinet, flute, trombone) were tested. The instruments were chosen because of their different excitation mechanisms (bassoon: double reed, cello: bowed string, clarinet: reed, flute: oscillating air jet ("air reed"), trombone: player's lips ("lip reed") in a bowl-shaped mouthpiece). Every instrument was tested at three different pitches taken from its natural range (for a more detailed description see: Siddiq, Reuter, Czedik-Eysenberg, & Knauf, 2017). Each pitch was presented in three different dynamic levels (*pp*, *mf*, *ff*), making it a total of nine tones per instrument and 45 tones altogether. All sounds were taken from the Vienna Symphonic Library (VSL).

Participants

A total of 43 subjects at the age of 13 to 75 years (Ø = 32.86, SD = 14.93; 20 females, 22 males, 1 other) participated in our experiment. Of these people, 32 were currently active musicians and eight formerly active with two to 60 years of experience (Ø = 20.66, SD = 14.08). Two participants were non-musicians.

Listening task

Participants were asked to subjectively rate the dissimilarities of all tested sounds in a pairwise comparison. Ratings were given through a seven-step scale where one is the least and seven the greatest dissimilarity. Participants were instructed orally and on the screen, were then presented all the stimuli beforehand, and had 20 rehearsal-trials before the actual test. Every pair could be repeated as often as necessary. A half-time break was mandatory, further breaks could be taken freely (for a detailed description of the procedure see: Siddiq et al., 2017).

Evaluation

The subjective data were aggregated into a dissimilarity matrix. Based on this matrix, a MDS (Package „smacof", ordinal MDS, see: de Leeuw & Mair, 2009) was calculated. The result is a four-dimensional configuration (stress-I (Kruskal) = 0.0320). Obviously, the MDS distances represent the subjective data (proximities) relatively well ($r = .97$, $p < .00001$). The audio signal analysis was performed via MIRtoolbox (Lartillot, Toiviainen, & Eronen, 2008) in Matlab.

Results

As the 2D-plot of dimension 1 and 2 unveils, all the sounds are sorted by ascending pitch along dimension 1. Along dimension 2, (1.) the location of the cello sounds, (2.) the order by pitch within the woodwinds, and (3.) the order by dynamics within the trombone indicate an overall order by spectral energy distribution.

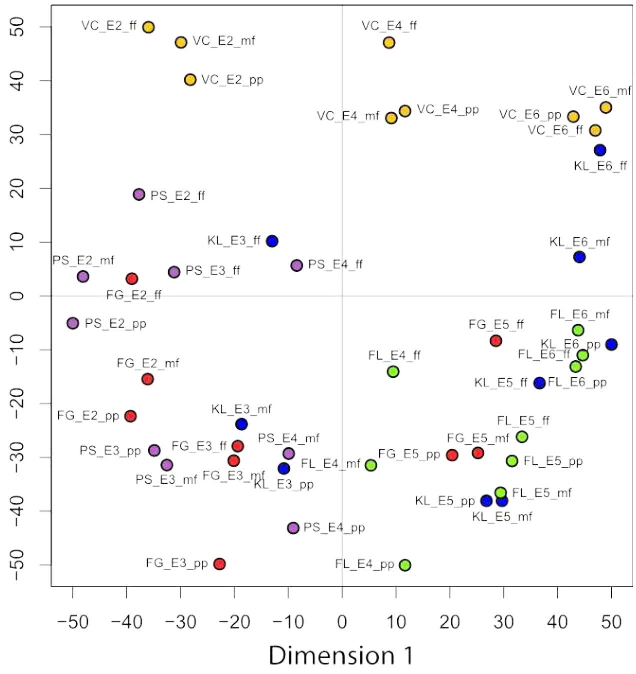


Figure 1. Dimension 1 (x-axis) and dimension 2 (y-axis) of the MDS configuration. FG = bassoon (red), FL = flute (green), KL = clarinet (blue), PS = trombone (lavender), VC = cello (yellow). E2 to E6 = pitch in scientific notation. *pp* (pianissimo), *mf* (mezzoforte), *ff* (fortissimo) = levels of musical dynamics.

According to these observations, dimension 1 correlates strongly with basically every descriptor of pitch. The relation even gets stronger if the correlation is calculated based on the ranks instead of the actual values. This (ranks r > values r) is a general finding and applies to all properties, so correlation coefficients are given for ranks unless specified otherwise. As hinted above, dimension 1 correlates strongly with the fundamental frequency of the stimuli ($f0$; $r = .97$, $p < .000001$), the zero crossing rate (ZCR; $r = .76$, $p < .000001$), and negatively with roughness (Sethares, 1998; $r = –.86$, $p < .000001$).

Dimension 2 indeed correlates with descriptors of the spectral energy distribution. It correlates strongly with brightness ($r = .86$, $p < .000001$), spectral rolloff ($r = .72$, $p < .000001$), and the spectral centroid (SC; $r = .62$, $p < .000001$). Moreover it correlates with the nominal assignment of dynamics ($r = .56$, $p < .0001$). Therefor, we manually assigned numerical values to the three dynamic levels (*pp* = 1, *mf* = 2, *ff* = 3).

As with every type of factor analysis, the first MDS-dimension explains the greatest amount of variance within the data. Then, dimension 2 explains most of the remaining variance and so on. Therefore, it is all but unexpected that correlations with dimensions 3 and 4 are not as clear-cut as with dimensions 1 and 2. Dimension 3 correlates weakly and negatively but significantly with the attack time ($r = –.32$, $p < .05$). Dimension 4 showed no informative correlations.

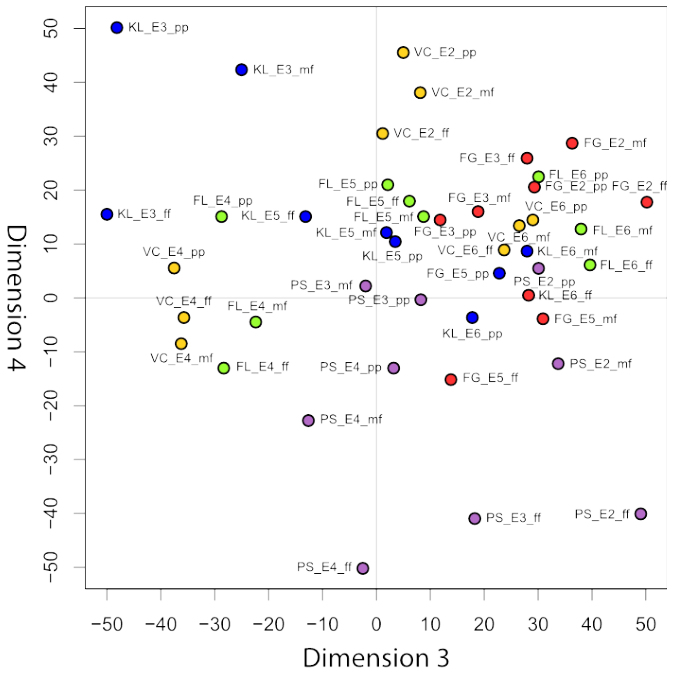


Figure 2. Dimension 3 (x-axis) and dimension 4 (y-axis) of the MDS configuration. For label explanation please see caption of Figure 1.

An examination of the spectrograms of the sounds along the respective dimensions reveals not only what the calculated correlations tell, but in some cases even further relations, that were not described by any of the used physical sound properties (see Figure 4). Again, dimension 1 can be identified as an almost perfect pitch-scale (Figure 4(A)). Dimension 2 can, again, be recognized as axis of the spectral energy distribution (Figure 4(B)). A closer look at the waveforms of the sounds along dimension 3 visually confirms the negative correlation with the attack time (Figure 4(C)). It's visible to the naked eye that sounds with steeper attack slopes tend to crowd on the right hand side of the scale. Please note that the calculated attack time only takes the rise until the first peak into account. Therefore, any irregularities in the attack transient may force the algorithm to return a much shorter attack time than the sound actually takes to reach its steady-state amplitude (see Figure 4(C), e.g. VC_E4 (all dynamics) or KL_E5_ff and pp). With that in mind, it is clearly visible that the correlation should be even higher.

Correlations with dimension 4 are weak and insignificant. A thorough investigation of the spectrograms but especially listening to the sounds leads to the impression that noise components, such as blowing noise or bowing noise, increase and become more dominant along dimension 4.

Discrimination of musical instruments

Generally, the sounds of each instrument strew across every dimension. Only dimension 2 correlates with the annotated category "instrument" ($r = .68$, $p < .001$). This category is analogous to the dynamics category (see above); i.e. every instrument was encoded with a value (bassoon = 1, flute = 2, clarinet = 3, trombone = 4, cello = 5) and subsequently subjected to correlation analysis in this form. If the axes of the plane of dimension 2 and dimension 3 are rotated, so that the x-axis parallels the linear regression line of the instrument category, the assorting of the instruments becomes even clearer ($r = .74$, $p < .001$).

Figure 3 shows an obvious clustering of the cello apart from the rest. Besides one outlier each, the bassoon and the flute are located in rather narrow corridors. The trombone and especially the clarinet are more widely scattered across the plain. Based on these findings, it seems save to state that the found descriptors are insufficient to reliably differentiate entire musical instruments.

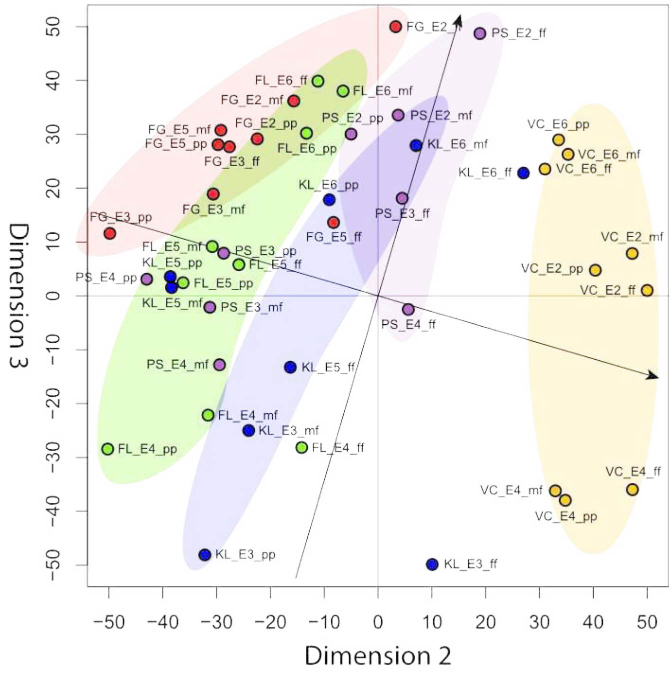


Figure 3. Dimension 2 (x-axis) and dimension 3 (y-axis) of the MDS configuration. The rotated dimension 2 (horizontally descending arrow) parallels the linear regression line of the external scale "instrument", the colored areas indicate the rough locations of the respective instrument clusters. For label explanation please see caption of Figure 1.

Summary

In short, dimension 1 is a dimension of pitch and dimension 2 correlates with descriptors of the harmonic spectrum. This finding might be interpreted as congruent to Helmholtz's unidimensional approach on timbre (von Helmholtz, 1863; von Helmholtz, 1875), which lead to the unfortunate misconception that became the standard definition of timbre. Beyond that, dimension 3 is seemingly connected with the temporal envelope, the attack transients to be precise. Although statistically insignificant, it stands to reason that dimension 4 might be explained through the noise components of the sounds. Carl Stumpf, who adopted and significantly advanced Helmholtz's thoughts on timbre, labeled Helmholtz's timbre as "timbre in a narrow sense" and combined all the sound properties, that explain dimensions 3 and 4, under the term "timbre in a broader sense" (Stumpf, 1890; for a short review of Helmholtz's and Stumpf's timbre concepts see: Reuter & Siddiq, 2017). Thus, these results may be interpreted as an empirical evidence of Stumpf's timbre concept.

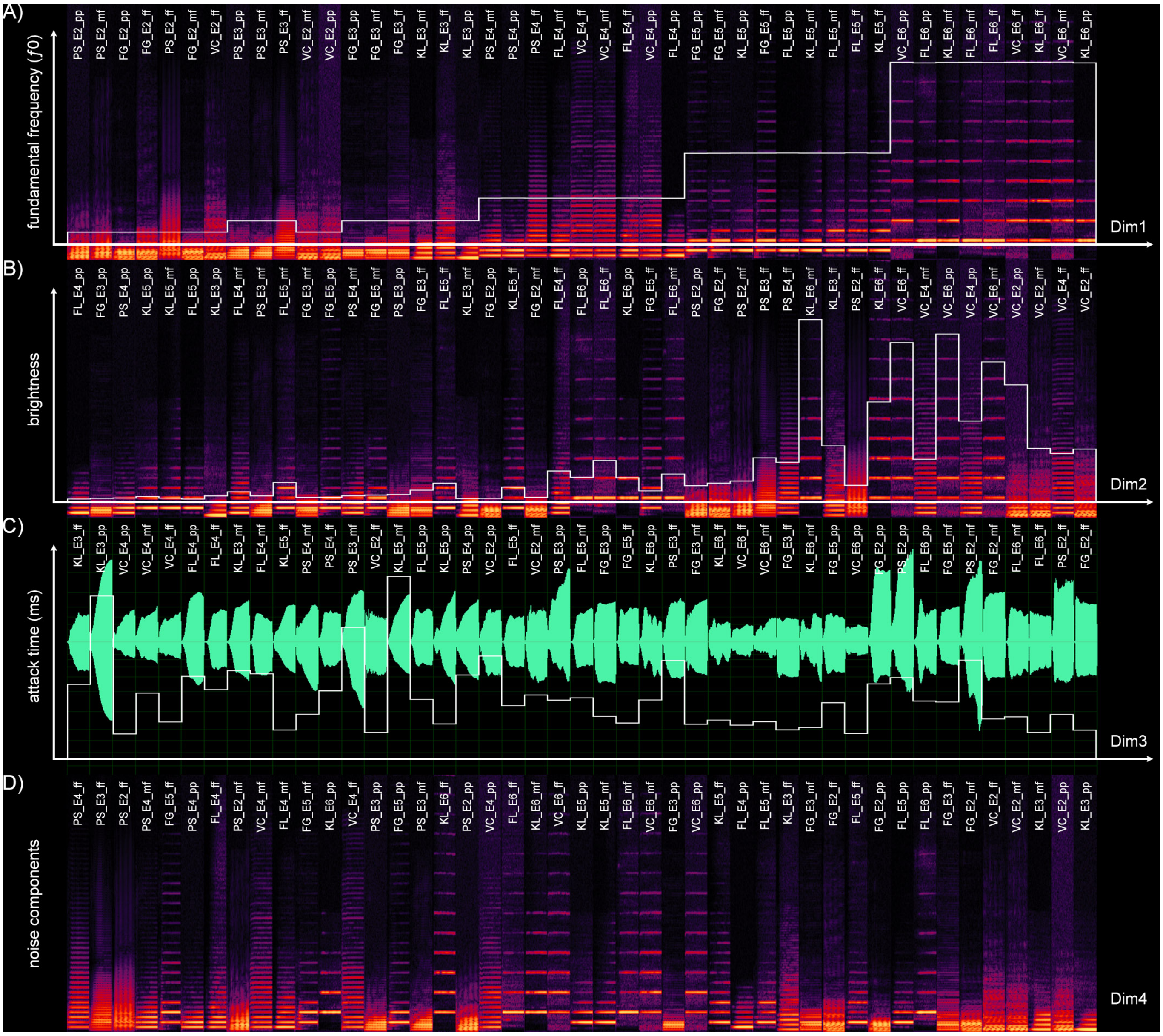


Figure 4. Spectrograms and Waveform of the sounds along the dimensions 1, 2, 4, and 3 respectively (x-axis). Superimposed on that is a histogram of the strongest correlating descriptor for each dimension (none for dimension 4). The x-axis of the histograms fits the spectrograms, the y-axis of the histograms was scaled up for better visibility.

Conclusions

The following assumptions may be taken from the results: First, pitch is a factor of timbre. Of course, as already stated by Stumpf (1890), brightness increases while roughness decreases with ascending pitch (even of sine waves). Moreover, it is generally known, that many instruments operate in different registers which, in turn, can only be accessed in dependence of pitch (e.g. Reuter, 2002; Rossing, Moore, & Wheeler, 2002).

Second, the mental image of musical instrument as entity does not apply. Unsurprisingly, sounds cluster physically and perceptually according to their properties. Obviously, sounds with e.g. similar pitch resemble each other more than sounds from the same instrument but with different pitch heights. If we consider musical instruments the "parents" of their sounds and all the sounds from the same instrument "siblings", we might say: siblings are not necessarily similar and physical or perceived similarity does not necessarily point to the same parent instrument.

Based on these findings, it becomes clear, that musical instruments cannot be appropriately tested on the basis of single sounds. Part of the problem is the ambiguous scope of the term "timbre". To facilitate communication concerning timbre and assist future clarification on research topics, we propose the following lexical resolution of the semantics behind the term timbre: In German language exists a term "Tonfarbe" which translates as "tone color". It was used by Stumpf in order to refer to pure tones (Stumpf, 1890). The term "timbre", or "sound color" (German: "Klangfarbe") would apply to the level of single sounds and/or noises like musical tones (produced by e.g. striking one key on the piano or one string on the guitar). At the level of musical instruments, Stumpf used the term "Instrumentalfarbe" which literally translates as "instrument color" (Stumpf, 1926). If need be, it might be changed to "voice color" in reference to the human voice. The following table gives a brief overview:

Table 1. Terminological concept for different levels of timbre

Level	physical correlate	German	English
Tone	pure tone	Tonfarbe	tone color
Note	complex tone/ isolated sound	Klangfarbe	sound color/ timbre
Instrument	entirety of all sounds from one instrument	Instrumental-farbe	instrument color

The proposed terms would not necessarily supersede the use of the word "timbre" in all circumstances but rather clarify matters where the use of "timbre" is equivocal. Moreover, "timbre" would naturally serve as (what biology knows as) genus type for the new bundle of specific terms within the scope of timbre.

Acknowledgements

Supported by funds of the *Österreichische Nationalbank* (OeNB, Anniversary Fund, project number: 16473). Special thanks to the *Vienna Symphonic Library* for supporting our work with two free "Vienna Super Package" licenses.

References

American National Standards Institute (1960). *American Standard. Acoustical Terminology (Including Mechanical Shock and Vibration)*. New York, NY: ASA

Bregman, A.S. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Caclin, A., McAdams, S., Smith, B. K., Winsberg, S. (2005). Acoustic correlates of timbre space dimensions. A confirmatory study using synthetic tones. *JASA* (118(1), 471–482. http://dx.di.org/10.1121/1.1929229

de Leeuw, J., & Mair, P. (2009). Multidimensional Scaling Using Majorization: SMACOF in R. *Journal of Statistical Software*, 31(3), 1–30.

Grey, J.M. (1975). *An Exploration of Musical Timbre using computer-based Techniques for Analysis, Synthesis and perceptual Scaling*. Stanford University, CCRMA Report No. STAN-M-2.

Handel, S. & Erickson, M. L. (2004). Sound Source Identification: The Possible Role of Timbre Transformations, *Music Perception* 21(4), 587–610. http://dx.doi.org/10.1525/mp.2004.21.4.587

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand? In S. Nielzen & O. Olsson (Eds.), *Structure and Perception of electroacoustic Sound and Music* (pp. 43–53). Amsterdam, NL: Elsevier.

Lakatos, S. (2000). A common Perceptual Space for harmonic and percussive Timbres. *Perception & Psychophysics* 62(7), 1426–1439. http://dx.doi.org/10.3758/BF03212144

Lartillot, O., Toiviainen, P., & Eronen, T. (2008). A Matlab Toolbox for Music Information Retrieval. In: C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications. Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation E.V.* (pp. 261–268) Heidelberg, DE: Springer. http://dx.doi.org/10.1007/978-3-540-78246-9_31

Marozeau, J., de Cheveigne, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency, *JASA* 114(5), 2946–2957. http://dx.doi.org/10.1121/1.1618239

McAdams, S., Winsberg, S., Donnadieu, S., & de Soete, G. (1995). Perceptual scaling of synthesized musical timbres. Common dimensions, specificities, and latent subject classes. *Psychological Research*, 58(3), 177–192. http://dx.doi.org/10.1007/BF00419633

Patterson, R. D., Walters, T. C., Monaghan J. J. M., & Goudrain, E. (2010). Reviewing the Definition of Timbre as it Pertains to the Perception of Speech and Musical Sounds. In E. A. Lopez-Poveda, A. R. Palmer & R. Meddis (Eds.), *The Neurophysiological Bases of Auditory Perception* (pp. 223–233). New York, NY: Springer Verlag. http://dx.doi.org/10.1007/978-1-4419-5686-6

Reuter, C. (2002). *Klangfarbe und Instrumentation. Geschichte – Ursachen – Wirkung* [Timbre and instrumentation. History – causes – effects]. Frankfurt am Main, DE: Lang

Reuter, C., & Siddiq, S. (2017). The colourful life of timbre spaces. Timbre concepts from early ideas to meta-timbre space and beyond. In. C. Wöllner (Ed.), *Body, Sound and Space in Music and Beyond. Multimodal Explorations* (pp. 150–167). Abingdon, UK: Routledge.

Rossing T. D., Moore, R. F., & Wheeler, P. A. (2002). *The Science of Sound* (3rd ed.). San Francisco, CA: Addison Wesley.

Plomp, R. (1970). Timbre as a multidimensional attribute of complex tones. In R. Plomp, & G. F. Smoorenburg (Eds.), *Frequency analysis and peridocity detection in hearing* (pp. 397–414). Leiden, NL: Sijthoff.

Sethares, W. A. (1998). *Tuning, Timbre, Spectrum, Scale*. London, UK: Springer.

Siddiq, S., Reuter, C., Czedik-Eysenberg, I., & Knauf, D. (2017). Die physikalischen Korrelate von Instrumentalklangfarben [The physical correlates of the timbre of musical instruments]. In. *Fortschritte der Akustik – DAGA 2017. 43rd Deutsche Jahrestagung für Akustik*, March 6–9, 2017 (pp. 194–197). Berlin, DE: DEGA.

Stumpf, C. (1890). Tonpsychologie [Tone Psychology] (Vol. 2). Stuttgart, DE: Hirzel.

von Helmholtz, H. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik* [On the sensation of tone as a physiological basis for the theory of music]. Braunschweig, DE: Vieweg.

von Helmholtz, H. (1875). *On the Sensation of Tone as a Physiological Basis for the Theory of Music* (A. J. Ellis, Trans.). London, UK: Longmans, Green, and Co.

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology*, 13(1), 228–240. http://dx.doi.org/10.1111/j.1467-9450.1972.tb00071.x

Wessel, D. L. (1979). Space as a Musical Control Structure. *Computer Music Journal*, 3(2), 45–52.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

PERCEPTUAL DIFFERENCES AND PREFERENCES BETWEEN BINAURAL AND STEREO MIXES OF MUSIC

Adèle Simon,[1] Geoff Luck[2]

[1,2] *Department of Music, Art and Culture Studies, University of Jyväskylä, Finland*

[1]adsimon@student.jyu.fi, [2]geoff.luck@jyu.fi

Abstract

Binaural technology allows the creation of a virtual auditory space, which, when presented over earphones, produces sound that appears to originate from different positions in space. Among other applications, this technology can be used to create three-dimensional (3D) music, with instruments virtually located all around the listener, not only on a stereo left/right axis. Adding a new dimension expands the creative possibilities to music makers and has the potential to offer a novel experience for listeners. Binaural technology has been broadly studied, yet research on the perceptual aspect is minimal, especially as regards musical content

The present study investigated affective differences between a binaural and a stereo music listening experience, especially focusing on preferences between the two mixing methods. The experiment consisted of a comparison between three different mixing technologies for musical excerpts, based on Nicol et al's (2014) recommendation to assess the quality of experience for 3D audio. Preference for each mix was directly rated by participants. Each excerpt was presented to participants in a stereo mix, a generic binaural mix (based on KEMAR Head Related Transfer Functions (HRTF) (Gardner & Martin, 1995)) and also in an individualised binaural mix (based on Head Related Transfer Functions (Iwaya, 2006) pre-selected to fit each participant). The idea was to study the differences between binaural music listening with different types of HRTF.

The outcomes of this study reveal a significant preference for 3D music, especially when the binaural mix is created with individualized HRTF. These results add more understanding to the perception on of 3D audio, and lead to further investigations on the use of binaural technology for musical and entertainment purpose.

INTRODUCTION

When talking about recorded music, especially when played through headphones, stereo mixing, with sources distributed on a left/right axis, is still the norm. However, thanks to other technologies, it is also possible to play sound as if it were emanating from every direction, in a three-dimensional space, thus creating 3D-Music.

Binaural technology, a technique based on a numerical reproduction of the physical human cues used for sound localisation, can create a virtual auditory space (VAS) for the listener and give the impression of having audio sources coming from every possible direction.

The concept of binaural technology relies on physiological cues that influence the perception of the sound : interaural distance that will affect both delay and level of the sound perceived by both ear ; pinna, head, torso shape that affect the spectral content of the sound (Wightman & Kistler, 1997). The human brain, unconsciously, learns to combine the results of these cues to associate each specific modification of the sound with a direction. To create synthetic binaural content, the effects of these cues are combined to create a filter representing each direction. As every individual has a different head and ears, these cues differ from one person to another. The filters, called Head Related Transfer Functions (HRTF), can be measured on a real or dummy head (Gardner & Martin, 1995), or extrapolated from measures of the head and ear (Algazi, Duda, & Satarzadeh, 2007; Zotkin, Hwang, Duraiswaini, & Davis, 2003). The filters are then applied to the sound to create a VAS with audio sources perceived as coming from various directions. This technology requires the sound to be played through headphones

The implications of such binaural listening, on the functioning of human sound localization, have been investigated since the late 1800's, and the first mention of binaural recording appears in the 1920's, also with the first dummy head recorder (named "Oscar"), and reproduction through headphones in the 1930's (Paul, 2009). This technology gained interest during the 1960's and 1970's with the development of dummy heads and the appearance of binaural microphones, opening the field to a new audience (musicians, sound engineers, etc.) (Usami & Kato, 1978).

Binaural technology has been used to produce and display music, although this technique is still marginal : the first to use it was Lou Reed on his album "Street Hassle" (Nusser, 1978), and then Pink Floyd (Loder, 1983), Pearl Jam (Pareles, 2000), and Radiohead ("Over your Shoulder - Binaural Listening", 2007)).

The perception of VAS and binaural technology is an active area of research, especially on the perception of the distance of the accuracy of sound localization (Begault, Wenzel, & Anderson, 2001; Blauert, 1997; Shinn-Cunningham, 2000). However affective studies, focusing on the overall impression, are more sparse. Studies suggested that 3D auditory spaces could improve the naturalness of perceived sound (Guastavino & Katz, 2004; Theile, 1991). Grani et al. (2014) showed that binaural rendering of audio is preferred over stereo in the case of a bimodal audio-visual context. For music, Fontana et al.'s (2007) results showed no particular preference for binaural mixes, compared with other mixing techniques.

The present study aimed to investigate issues of liking and appreciation of binaural musical content. This investigation was based on a comparison with a reference, as suggested by Nicol et al (2014). The reference was the same musical content, but created with stereo mixing methods, the most commonly-used method for recorded music. The hypothesis was that either 3D or stereo music will be preferred to the other. Both direction for preference could be inferred from the literature: stereo could be preferred due to a familiarity aspect (Fontana et al., 2007) ; preference could also be influenced by a "wow" effect, due to novelty of the content (Rumsey, Zielinski, Kassier &

Bech,2005). Both of these directions could also be influenced by the expertise of the listener (familiar or not with 3D audio, or musicians who would have a more critical approach to music listening).

In addition to comparing binaural and stereo music, the current study also compared the influence of using individualized HRTF and generic HRTF on the quality of experience for binaural music listening. Individualized HRTF is supposed to better suit each participant, and can be either measured, created from morphological data, or subjectively selected from a database (Xu, Li, and Salvendy, 2007). For this experiment, as measuring HRTF was not possible, HRTF was selected by the participants. Studies showed that individualized HRTF tends to be more accurate for the spatial localisation (Møller, Sørensen, Jensen, & Hammershøi, 1996), but not necessarily for the perception of the "naturalness" of the auditory content (Usher & Martens, 2007). Nonetheless, the preference for auditory content relies not only on the fidelity of the sound, but on many other factors (Begault, 2006). Also, as pointed out by Shinn-Cunningham (1998), the quality of experience and aesthetic aspects should prevail over the realism of the listening. The influence of individualized HRTF on affective aspects, such as preference, is still unknown. Determining if use of fitting HRTF enhances the preference for the content is one of the goals of the present research.

This study consisted of a comparison of three different mixing conditions for musical content : stereo, binaural music with generic HRTF, and binaural music with individualized HRTF. Comparison will be based on preference rating. The expertise of the listener, and its potential effect on the rating was also investigated.

METHOD

PARTICIPANTS

A total of 32 participants took part in the experiment (21 females), aged 19-34 (M=25.44). 25 were amateur musicians, 4 were professionals or semi-professionals, and 3 were not musicians. 14 had some experience in research in the field of music or audio, and 21 had some experience as a sound engineer (recording or mixing). Two of them reported a mild tinnitus. None of the participant was an expert in 3D audio, 16 were slightly familiar with the concept.

HRTF SELECTION

In order to obtain HRTF that fitted each participant, a listening task was implemented, based on the DOMISO (Determination method of OptimuM Impulse-response by Sound Orientation), developed by Iwaya (2006). This task is a tournament style approach to pick out of a database the more suitable set of HRTF for each participant.

A test stimulus was created, consisting of sequences of 13 bursts of white noise (300 milliseconds duration), separated by 100ms of silence. All the bursts were located on a 0° elevation plane and moved sequentially in +30° steps on azimuth, starting and finishing on the right side of the listener. This test stimulus was convolved with all the HRTF of the Listen database (Warusfel, 2003) to create 50 different test stimuli for comparison.

From those 50 stimuli, 32 were firstly randomly selected. Then, they were presented in pairs, where one was selected by the participant as the best to represent a circle around his or her head. Once a stimulus lost two duels, it was excluded from the set, and the duels continued until only one final winner remained. The winner, different for each participant, was used to create the individualized 3D condition.

The generic 3D version was created using the Kemar set of HRTF (Gardner & Martin, 1995).

STIMULI

Five songs were used for this study, all drawn from an open database (“The 'Mixing Secrets' Free Multitrack Download Library”, nd). In order to be mixed in 3 dimensions, the stimuli needed to be in a multitrack format, which allow a mixing phase and a placement of each instrument in the VAS. In addition, the songs were chosen to cover various genres, and variousalso because of their arousing potential. Excerpts of each song, ranging from 68s to 90s in length, were used as stimuli.

The stimuli were the following:

- -Jessica Childress: *“Slow Down”*, classified as Soul
- -Admiral Crumple: *“Keeps Flowing”*, classified as
- -Timboz: *“Pony”*, classified as Metal
- -Banned from The Zoo: *“Fish Tacos”*, classified as Pop-Folk
- -Leftover Salmon: *“River’s Risin”*, classified as Country

Each excerpt was then mixed in both stereo and 3D. Mixing was carried out using Protools 11.03 (Avid, nd) and VR Audio Work Station (3DSoundLabs, nd). The latter was also used to add a room effect to the music, similar for each version and coherent with the placement of the source, that followed Møller’s (1992) idea of a “binaural room simulator”

Once the mixing part was complete and the instruments distributed in the virtual space, they were convolved with HRTF (KEMAR and HRTF from the LISTEN database). For the stereo version, no HRTF convolution was performed. The left/right placement of the sources for the stereo version was consistent with the placement of instruments in the 3D versions.

PROCEDURE

Participants were asked to listen to five songs, each mixed in three different conditions: 1) stereo mix, 2) binaural mix created with generic (KEMAR) HRTF, 3) binaural mix created with individualized HRTF. For each song, participants rated their preference for each version. Stimuli were played through Audiotechnica ATH-M50x headphones (Audiotechnica, nd).

Ratings were obtained using a continuous, horizontally oriented, scale labelled with 5 descriptive, equidistant points on the scale (Dislike Extremely - Dislike Moderately - Neither like nor dislike - Like moderately - Like very much).

The rating interface was implemented using the Web Audio Evaluation Tools (Jillings et al.,2015), and based on a model developed by De Man & Reiss (2014). For each song, the three different conditions were presented at the same time (each condition assigned to a different cursor) and freedom was given to the participants to choose the listening order and to switch from one version to another whenever they wanted and for as long as they wanted before rating preference attribute. The transitions between versions were seamless and started at the same point where the previous excerpt was stopped. Once participants confirmed their ratings, they continued and repeated the same process for the next song. The different conditions (stereo, individualized 3D, and generic 3D) were randomly allocated to cursors. The order of the song presentation, and the initial position of each cursor on the scale were s also random. Prior to the actual task, a training phase helped the participants to get familiar with the process.

In addition to this task, participants answered to a questionnaire which included questions about their experience as musicians and, sound engineers, and in research in audio/music. From those data, participants were divided in two groups : 1) expert listener, who has their main activity related to music/audio; 2) naïve listener, including nonmusicians and amateur musicians. This expertise score was included as a between subject factor in the analysis. Out of the 32 participants, 15 were classified as naive listeners and 17 as experienced listeners

RESULTS

The differences in preference between conditions were calculated with a 3 × 5 × 2 mixed design ANOVA (5 songs, and 3 mixing conditions as within subject variables ; 2 levels of expertise as a between subject variable). The scale's range for preference rating was 0-100.

Mauchly's Test of Sphericity indicated that the assumption of sphericity had been violated, and therefore, a Huyn-Felt correction was used. There was a significant effect of the mixing condition on preference rating, $F(3.476, 107.752) = 16.303$, $p < .001$.

Pairwise comparisons between conditions, with Bonferroni adjustment for multiple comparisons, showed that individualized 3D versions (M=55.338) were significantly preferred (p=.041) over generic 3D (M=51.794), and also significantly preferred (p=001) over stereo versions (M=46.168). The difference of preference between generic 3D and stereo was also significant (p=.015).

The repeated measures ANOVA, with a Greenhouse-Geisser correction, also revealed a significant interaction between the Song and Mixing Condition factors, $F(4.815,149,265)=5.741$, $p<.001$.

Pairwise comparisons, with Bonferroni adjustments, showed different variations of preferences for the various Song*Conditions combinations. For the blues song, the individualized condition (M=75.138) was significantly (p=.026) preferred to the generic condition (M=67.150) and also significantly (p=.001) preferred to stereo (M=59.046). For the rap song, the individualized condition (M=52.761) was significantly (p=.019) preferred compared to the generic condition (M=45.755) and also significantly (p=.002) preferred compared to stereo (M=37.172). The generic condition was also significantly (p=.031) preferred than stereo. For the Folk-Rock song, the individualized condition (M=65.996) was significantly (p<.001) preferred than the generic condition(M=53.044) and significantly (p<.001) preferred than stereo(M=48.234). For both metal and country songs, the differences of preferences between conditions were nonsignificant. The results are summarised in Figure 1.

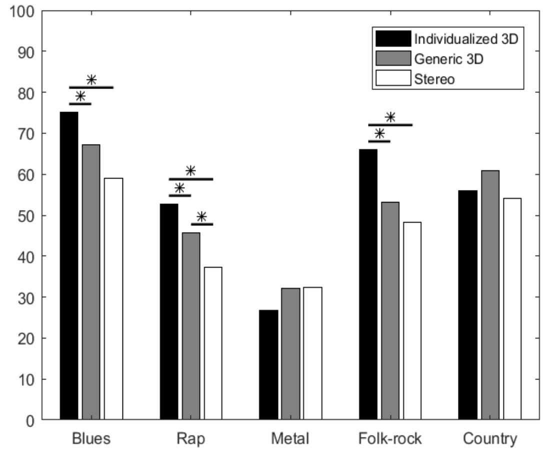


Figure 1. **Mean ratings of preference according to conditions and songs. * represents significant differences of ratings between versions, *p*<.05**

For the expertise factor, the ANOVA revealed a statistically significant effect of the expertise factor level on preference ratings: $F(1, 30) =5.651$, $p= .024$. However, there were no significant interactions between expertise and mixing conditions factors.

DISCUSSION AND CONCLUSION

This study was conducted to investigate an affective aspect of music listening with binaural technology, especially to examine listener's preference between stereo mixing and 3D mixing. The hypothesis was that either stereo or 3D versions would be preferred, but the direction of the preference was unclear. In the literature, both directions have been suggested: Fontana et al.(2007) suggested that stereo could be preferred over 3D music because listeners are more familiar with this type of mix; Rumsey, Zielinski, Kassier & Bech (2005) suggested that novelty associated with 3D music could produce a "wow" effect, which might influence preference.

The outcomes of this study show a significant preference for 3D music using binaural technology compared to stereo. This preference is higher when the 3D effect is created using individualized HRTF compared to generic HRTF. Those results suggest that the use of spatialization for recorded music content can create a more pleasant experience of music listening.

The interactions between song and mixing condition factors revealed that this effect was not significant for each stimulus individually. Although many factors can influence this effect, the mixing process also has to be kept in mind. As the mixing process of music involves subjective choices, those decisions could influence the music. For 3D music, as this technique is still marginal, there are no guidelines on how to mix and how to place instruments in a 3D VAS, even if some have reflected on this question (Barrett, 2002; Begault & Trejo, 2000). For the stimuli used in the present study, that could have played a role in the rating. Nonetheless, the stimuli used for this type of study – and the stimuli creation process is crucial – have to be considered and would need more reflection and research.

Expertise level of the participants was also studied. Fontana et al.(2007) suggested that experienced listeners, especially when familiar with 3D audio, would tend to be more critical on the listening task, and also to focus on the technical aspect rather than on the overall experience. The results of this study revealed a significant effect of the listeners' level of expertise on the preference, but only overall, and thus independently of the mixing conditions. No significant interactions were found between the expertise level and the mixing conditions factor. That suggests that expert and naïve listeners rated the three versions in a similar manner, even if the naïve tended to rate preference higher (regardless of the mixing condition).

Finally, it is important to point out that individualized 3D condition was preferred to generic 3D. This outcome supports the idea that personalized HRTF enhance the binaural effect and could help provide a more enjoyable listening experience.

The findings presented here contribute to the topic of quality of experience for binaural music listening. As a follow-up to this study, the work done here for binaural music could be extended in the future either to other technologies that could produce spatialized music, or to other binaural content (radio, audio-visual, interactive, etc.).

Acknowledgements. This study is a part of a master's thesis, done at the University of Jyväskylä

References

3D Sound Labs. (n.d.). VR Audio Work Station [computer software]. Retrieved from http://www.3dsoundlabs.com/category/developpeurs

Algazi, V. R., Duda, R. O., & Satarzadeh, P. (2007, May). Physical and filter pinna models based on anthropometry. In *Audio Engineering Society Convention 122*. Audio Engineering Society

Audio-Technica. (n.d.). Ath-m50x [apparatus]. Retrieved from http://www.audio-technica.com/cms/headphones/99aff89488ddd6b1/

Avid. (n.d.). Protools 11.03 [computer software]. Retrieved from http://www.avid.com/pro-tools

Barrett, N. (2002). Spatio-musical composition strategies. *Organised Sound, 7 (03),313–323*.

Begault, D. R., & Trejo, L. J. (2000). 3-D sound for virtual reality and multimedia.

Begault, D. R., Wenzel, E. M., & Anderson, M. R. (2001). Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, *49*(10), 904-916

Begault, D. R. (2006). Preference versus reference: Listeners as participants in sound reproduction. *Spatial audio and sensory evaluation technique*s, Guildford, UK.

Blauert, J. (1997). Spatial hearing: the psychophysics of human sound localization. MIT press.

Capra, A., Fontana, S., Adriaensen, F., Farina, A., & Grenier, Y. (2007). Listening tests of the localization performance of Stereodipole and Ambisonic systems. In *Audio Engineering Society Convention 123*. Audio Engineering Society.

De Man, B., & Reiss, J. D. (2014). Ape: Audio perceptual evaluation toolbox for matlab. *In Audio engineering society convention 136.*

Gardner, W., & Martin, K. (1995). HRTF measurements of a KEMAR. *The Journal of The Acoustical Society of America*, *97*(6), 3907-3908. doi: 10.1121/1.412407

Grani, F., Argelaguet, F., Gouranton, V., Badawi, M., Gaugne, R., Serafin, S., & Lecuyer, A. (2014). Design and evaluation of Binaural auditory rendering for CAVEs. In *Virtual Reality (VR), 2014 iEEE* (pp. 73-74).

Greenwood, J (2007) *Over your shoulder – Binaural Recording*. Retrieved from http://www.greenplastic.com/2007/05/07/over-your-shoulder-binaural-recording/

Guastavino, C., & Katz, B. (2004). Perceptual evaluation of multi-dimensional spatial audio reproduction. *The Journal of The Acoustical Society of America*, *116*(2), 1105-1115. doi: 10.1121/1.1763973

Jillings, N., Moffat, D., De Man, B., & Reiss, J. D. (2015, July). Web Audio Evaluation Tool: A browser-based listening test environment. *In 12th sound and music computing conference.*

Loder, K. (1983). The final cut. Rolling Stone. Retrieved from https://www.rollingstone.com/music/albumreviews/the-final-cut-19830414

Møller, H. (1992). Fundamentals of binaural technology. *Applied acoustics, 36 (3-4),171–218*

Møller, H., Sørensen, M. F., Jensen, C. B., & Hammershøi, D. (1996). Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society*, *44*(6), 451-469.

Nicol, R., Gros, L., Colomes, C., Noisternig, M., Warusfel, O., Bahu, H., ... & Simon, L. S. (2014). A roadmap for assessing the quality of experience of 3d audio binaural rendering. *10.14279/depositonce-4103*.

Nusser, D. (1978). Arista has 1st stereo/binaural disk. *Billboard, 18*.

Paul, S. (2009). Binaural Recording Technology: A Historical Review and Possible Future Developments. *Acta Acustica United with Acustica*, *95*(5), 767-788. doi: 10.3813/aaa.918208

Pareles, J. (2000). Binaural. *Rolling Stone*. Retrieved from https://www.rollingstone.com/music/albumreviews/binaural-20000608

Rumsey, F., Zielinski, S., Kassier, R., & Bech, S. (2005). Relationships between experienced listener ratings of multichannel audio quality and naïve listener preferences. *The Journal of The Acoustical Society of America*, *117*(6), 3832-3840. doi: 10.1121/1.1904305

Shinn-Cunningham, B. G. (1998). Applications of virtual auditory displays. *In Proceedings of the 20th international conference of the ieee engineering in biology and medicine society (Vol. 3, pp. 1105–1108).*

Shinn-Cunningham, B. G. (2000). Distance cues for virtual auditory space. In *Proceedings of the IEEE-PCM* (Vol. 2000, pp. 227-230).

The Mixing Secrets - Free Multitrack Download Library. (n.d.). Retrieved from http://www.cambridge-mt.com/ms-mtk.htm

Theile, G. (1991). On the naturalness of two-channel stereo sound. In *Audio Engineering Society Conference: 9th International Conference: Television Sound Today and Tomorrow*. Audio Engineering Society.

Usami, N., & Kato, T. (1978). *U.S. Patent No. 4,088,849*. Washington, DC: U.S. Patent and Trademark Office.

Usher, J., & Martens, W. L. (2007). Perceived naturalness of speech sounds presented using personalized versus non-personalized HRTFs. *Proceedings of the 13th International Conference on Auditory Display*, Montréal, Canada, June 26-29, 2007

Warusfel, O. (2003). *Listen hrtf database*. Retrieved from http://recherche.ircam.fr/equipes/salles/listen/

Wightman, F., & Kistler, D. (1997). Monaural sound localization revisited. *The Journal of The Acoustical Society of America*, *101*(2), 1050-1063. doi: 10.1121/1.418029

Xu, S., Li, Z., & Salvendy, G. (2007). Individualization of head-related transfer function for three-dimensional virtual auditory display: a review. In *International Conference on Virtual Reality* (pp. 397-407). Springer, Berlin, Heidelberg.

Zotkin, D. Y. N., Hwang, J., Duraiswaini, R., & Davis, L. S. (2003). HRTF personalization using anthropometric measurements. In

Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on. (pp. 157-160). Ieee.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Effect of Melody and Rhythm on the Perception of Nonadjacent Harmonic Relationships

Joanna Spyra[1] and Matthew Woolhouse[2]

[1] *Psychology, Neuroscience and Behaviour, McMaster University, Canada*
[2] *School of the Arts and Institute for Music and the Mind, McMaster University, Canada*

[1]spyraj@mcmaster.ca, [2]woolhouse@mcmaster.ca

Abstract

Temporally nonadjacent key relationships are ubiquitous in tonal-harmonic music although, the degree to which they are perceived beyond relatively short durations is uncertain. Previous research suggests that memory for a key remains active up to 20s after modulation. However, most previous work has been limited to the harmonic domain, typically employing homophonic textures or repeated arpeggios. Surface musical features—in addition to harmony—may be important in maintaining temporally nonadjacent key relationships. The current study aims to investigate this in a controlled manner by manipulating specific features of the musical surface and testing the effects of these manipulations on global perceptions. Two music-theoretically defined features were tested: melodic figurations (e.g. chordal skips and passing tones) and rhythmic figurations (e.g. anticipations and suspensions). Stimuli were segmented into three parts: X_1 (key establishing sequence), Y (second key), and X_2 (probe cadence in the key of X_1). We investigated the effects of the musical surface from three perspectives: (1) whether the presence of melodic and/or rhythmic figurations increased the global effect of X_1; (2) whether the type of harmonic progression in X_1 affected global retention; and (3) whether the probe cadence needed to be melodically and rhythmically similar to X_1 in order to maintain this global retention. The probe cadences in stimuli with rhythmic or melodic figurations, i.e. surface features, were rated significantly higher than unembellished control stimuli. Similarly, stimuli in which the probe cadence was similar to X_1 were rated higher than when they were different. No significant effect was found for type of harmonic progression. Our results are consistent with our hypothesis that surface musical features contribute to the establishment and maintenance of temporally nonadjacent key relationships within tonal harmonic music. However, 'deeper' structures, such as type of harmonic progression (for example, cycle of fifths), seem not to be crucial.

Introduction

The world is full of nonadjacencies and it is our ability to retain segments of information over time that is crucial to our perception of it. Few studies in the musical literature explore the processes that are involved in connecting nonadjacencies and it is unclear what our limits are in this domain, which factors effect perception, or even which cognitive processes are involved.

Cook (1987) was one of the first to explore musical nonadjacency by questioning our ability to perceive large-scale tonal structures—or the organization of music into larger structural patterns. His main finding was that participants were unable to perceive structures longer than a minute in length—a surprisingly short time frame considering the vast amount of organization found in large-scale musical compositions.

Woolhouse, Cross and Horton (2016) explored this finding using a concise 3-part stimulus paradigm. By keeping the first and last third of the stimulus in the same key, they were able to manipulate factors in the intervening section and test their influence on our ability to make nonadjacent connections. To this end, Woolhouse et al. (2016) tested the effect of time on nonadjacent key perception by transposing the intervening section into different keys and manipulating its length. They found that participants' abilities to perceive nonadjacent key relationships logarithmically decreased as an intervening key's length increased.

Farbood (2016) extended this research by testing the limits of perception in time. Using arpeggios rather than block chords, she found that participants could still successfully connect the nonadjacent key sections when the intervening key was up to 20 seconds in length.

Findings from Cook (1987), Woolhouse et al. (2016) and Farbood (2016) all provide evidence to suggest there are definite time limits to nonadjacent musical perception. This limit suggests key is represented in memory which decays within a short time frame of 10-20 seconds (Spyra, 2017). However, previous work on the subject used solely homophonic musical stimuli, which is rarely experienced in the real world. By manipulating the complexity of melody and rhythm, the current study aims to test a more ecologically valid set of stimuli with the intention of more closely resembling real music.

Hypothesis

The current study uses a goodness of completion probe stimulus paradigm. As such, we hypothesize that the presence of surface musical features such as melody or rhythm will result in higher goodness of completion ratings. These surface features create a more salient stimulus that is more representative of familiar musical experiences. As such, these stimuli may be more distinct in comparison to regular homophonic stimuli, thereby increasing attention or even memory for them.

We also manipulate the probe itself to further test the effect of these surface features. We hypothesize that the more similar nonadjacent sections are to each other, the easier it will be for participants to connect both sections and the higher sense of completion the probe will exude.

Methods

Participants

47 students from the undergraduate research pool at McMaster University (29 female, 18 male) participated in this experiment. 26 participants had 5 or more years musical training (Mean, 5 years; SD, 3.68), which was defined as those students with either formal lessons or self-taught musicians. Participants were rewarded for their participation with a course

credit under the McMaster Psychology, Neuroscience and Behaviour research participation system.

Apparatus

Stimuli were generated in MuseScore2. The timbre chosen for this experiment was a bassoon timbre as this provided a balance between fast attack rate and slower decay and gave a clear articulation to the musical lines. These were important qualities as certain stimuli included overlapping notes which were not perceptible in regular piano timbre. Stimulus presentation was controlled by a program created using the Python programming language with the Kivy package for graphic user interface. Participants listened to the stimuli through headphones and were able to adjust the volume freely. Responses were entered by adjusting a continuous 1-7 point slider presented by the Python program. A continuous measurement was chosen to ensure the participant interpreted the scale as interval and not ordinal.

Stimuli

Stimuli were made of three parts: (1) a nonadjacent section that was temporally nonadjacent to the probe sequence; (2) an intervening section that ended on a quarter note rest; and (3) a final probe cadence sequence. The nonadjacent section and probe sequence were always presented in the same musical key and could be written in one of any of the 12 possible major keys found in typical Western classical music. The intervening section was never presented in the same key as the nonadjacent and probe sequences but was instead transposed to any key that was 2, 4, or 6 semitones up or down from the nonadjacent key. All sequences were written using Western compositional voice-leading rules.

Table 1. Breakdown of all possible combinations of factors.

Sequence	Melodic Figuration	Rhythmic Activity	Cadence (same/diff)	Direction
C5	F1 (yes)	A1 (yes)	CS	2, 4, 6 -2, -4, -6
			CD	
		A0 (no)	CS	
			CD	
	F0 (no)	A1	CS	
			CD	
		A0	CS	
			CD	
Non-C5	F1	A1	CS	
			CD	
		A0	CS	
			CD	
	F0	A1	CS	
			CD	
		A0	CS	
			CD	

Table 1 provides a breakdown of five factors of interest: sequence, melodic figuration, rhythmic activity, cadence and direction. The first factor, sequence, represents the harmonic progression with which the stimuli were written. This factor had two levels: the nonadjacent and intervening sections could either be written using a cycle of fifths progression (C5: I-IV-vii°-iii-vi-ii-V-I) or a non-cycle of fifths progression (Non-C5: V-I-I-IV-ii-V6/4-V5/3-I). The nonadjacent and intervening sections were always presented in opposing progressions; if the nonadjacent section was a C5 progression, the intervening section would be Non-C5 and vice versa. The probe sequence was comprised of the last three chords of the nonadjacent section (C5: ii-V-I; Non-C5: V6/4-V5/3-I) and always matched the progression presented in that section. Previous research has suggested that a probe cadence composed of only two chords (e.g. V-I) does not provide enough context for participants to fully understand the key it was in (Spyra, 2017). For this reason, we decided to add a third chord to the beginning of the probe sequence, e.g. ii-V-I.

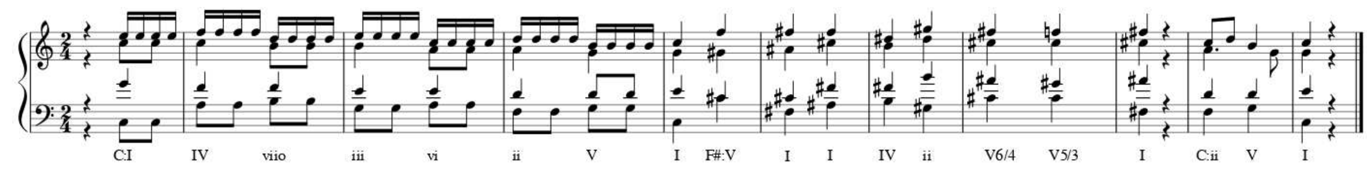
The second and third factors of interest were melodic figuration and rhythmic activity—whether the sequence was melodically and rhythmically complex or simple. These factors had two levels each: the nonadjacent sequence could either be melodically complex (F1; figuration) or simple (F0; no figuration); the nonadjacent sequence could either be rhythmically complex, in which more notes were added to the sequence (A1) or rhythmically simple (A0), which had the same number of notes as the regular sequence. For example, if a sequence was written in the most frugal way possible, it would be composed of only quarter-note chords and designated F0-A0—without figuration or rhythmic activity (Figure 1, first stimulus). If the sequence was melodically simple but rhythmically complex (F0-A1), the top (soprano) voice would be written in sixteenth notes but all of the notes would correspond to the original soprano note of the chord as it had no melodic figuration (Figure 1, second stimulus). However, if this sequence had both melodic figuration and rhythmic activity (F1-A1), these sixteenth notes would be arranged into a distinct melody (Figure 1, fourth stimulus). Lastly, the F1-A0 configuration adds suspensions and anticipations to provide melodic interest without adding any extra notes to the sequence (Figure 1: third stimulus). The intervening section was always written using straight quarter note chords (F0-A0).

Another factor of interest was the cadence. This refers to the probe-cadence sequence itself, which could either be the same (CS; cadence same) or different (CD) from the nonadjacent sequence in terms of figuration and rhythmical activity. If, for example, the nonadjacent sequence had both melodic figuration and rhythmic activity (F1-A1) then a cadence would either also have that factor combination (CS; F1-A1) or be arranged in the opposite fashion (CD; F0-A0).

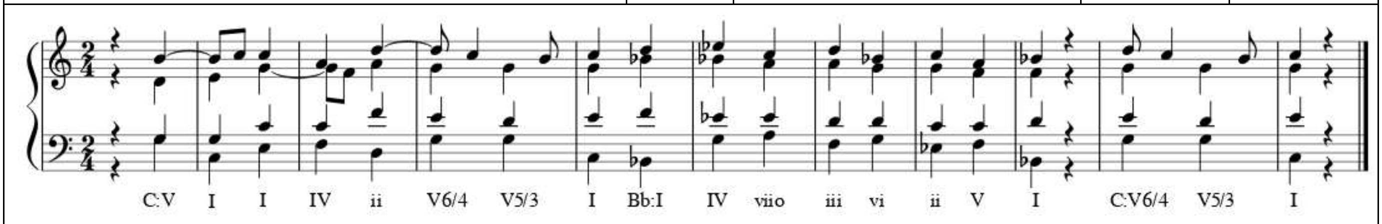
The last factor of interest in this experiment was the direction of modulation between the nonadjacent and intervening sections. As was previously discussed, the intervening section was never presented in the same key as the nonadjacent section, and was modulated up or down by 2, 4, or 6 semitones. This is designated by positive numbers for upwards modulation and negative numbers for downwards (see Table 1 and Figure 1).

Nonadjacent Section C	Intervening Section D		Probe C	
C5	2	Non-C5	C5	
F0-A0		F0-A0	CS	F0-A0

Nonadjacent Section C	Intervening Section F#		Probe C	
C5	6	Non-C5	C5	
F0-A1		F0-A0	CD	F1-A0



Nonadjacent Section C	Intervening Section Bb		Probe C	
Non-C5	-2	C5	Non-C5	
F1-A0		F0-A0	CS	F1-A0



Nonadjacent Section C	Intervening Section F#		Probe C	
Non-C5	-6	C5	Non-C5	
F1-A1		F0-A0	CD	F0-A0

Figure 1. Sample of stimuli showing four possible combinations of factors: sequence, melodic figuration, rhythmic activity and cadence (same or different). Stimuli above all begin in C major for clarity alone.

Procedure

Every participant gave informed consent and completed a questionnaire that gauged their prior musical experience. Questions included detailed inquiries as to their experience in formal musical lessons, as well as inquiries into informal musical experience (such as being self-taught on an instrument). These questions were carefully crafted to include musical expertise beyond formal classical training. For the purpose of this experiment, a 'musician' was defined as a participant with 5 or more years of musical training, regardless of whether it was formal or informal.

96 stimuli were presented in randomized order to every participant; one in every combination of factors discussed above (see Table 1). On each trial, participants listened to a musical passage followed by a probe sequence. Once the audio file had finished, participants were prompted to rate the goodness of completion of the probe sequence on a Likert-type scale from 1 (poor) to 7 (excellent). The entire experiment lasted for approximately 45 minutes.

Results

The raw data did not pass the Shapiro-Wilks test of normalcy, so raw scores were normalized for each participant. A factorial within-subjects ANOVA was then conducted to explore the effects of the five factors of interest described above while controlling for the nonadjacent key (any influence caused by stimuli beginning in different keys).

No significant difference was found between C5 and non-C5 sequences ($F_{1,44} = 3.66$, $p = 0.056$). The residual (difference in means) between them was only 0.08. There was, however, a significant effect of melodic figuration ($F_{1,44} = 6.08$, $p = 0.014$, residual = 0.15) and of rhythmic activity ($F_{1,44} = 27.395$, $p < 0.0005$, residual = 0.2). The main effect of cadence—whether it was the same as the nonadjacent section—was highly significant ($F_{1,44} = 88.949$, $p < 0.0005$, residual = 0.494). Lastly, the main effect of modulation direction was also highly significant ($F_{5,44} = 19.78$, $p < 0.0005$).

Discussion

Experimental results suggest that including musical surface features, such as melodic figuration and rhythmical activity, bolstered the retention in memory of the nonadjacent key, and thus enhanced the perception of global harmonic effects. Which is to say, that increasing the melodic or rhythmical complexity of a piece appears to create a stronger perceptual connection between nonadjacent sections.

Ratings for cadences that were melodically and rhythmically the same as the corresponding nonadjacent section were significantly higher than that for different cadences. What is interesting in this finding is the magnitude of the difference. Whereas the mean difference between ratings for melodic figuration or rhythmic activity did not surpass 0.2, the mean difference between same and difference cadences was more than twice that much (0.494). Despite both sections being presented in exactly the same key with exactly the same chords, changing melodic or rhythmic complexity had a staggering effect on participant ratings. This, perhaps most of all, supports the current finding that surface features exert a strong influence over the perception of nonadjacent harmonic relationships.

No significant difference was found between cycle-of-fifths (C5) and non-C5 sequences, suggesting that the type of harmonic sequence may not be as important as textural qualities in maintaining global relationships in composition.

Lastly, the direction of modulation was also significantly different. Though this significance is perhaps not surprising, the immerging patterns certainly bare remark. Figure 2 reveals an interesting pattern; modulating either 2 semitones up or down had similar mean ratings, as did 4 up and down, and 6 up and down. More specifically, mean ratings were similar for each distance pair and decreased as modulation distance increased. In other words, the direction itself was not the driving factor of the significant result, rather it was the distance of modulation. This symmetry is contrary to such works as Cuddy and Thompson (1992) which show a distinct asymmetry between the perception of modulations of the same distance but different directions on the cycle-of-fifths.

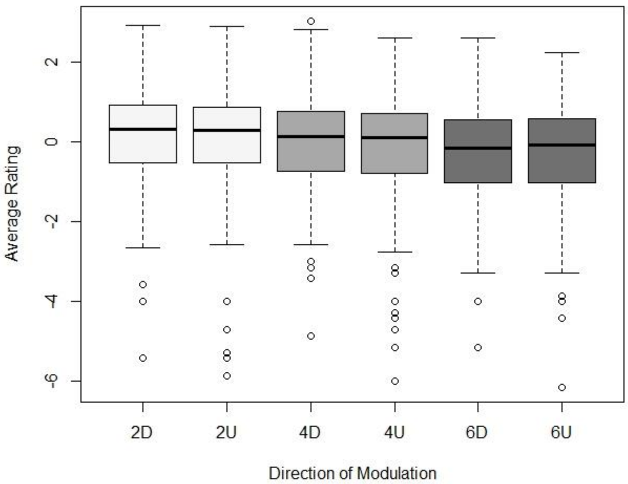


Figure 2. Mean ratings per distance and direction of intervening modulation

A limitation of this study was that 'time' was not included as a factor of interest. Participants found this experiment taxing at its current length and the addition of this factor would have extended the experiment beyond an optimal length of time. In a future study, it would be exciting to see whether the addition of melody and rhythm into the nonadjacent paradigm would extend the predicted times of previous research.

Conclusion

The current study provides further support for the findings of Woolhouse, et al. (2016) and Farbood (2016). But where they explored nonadjacent key relationships in musically frugal stimuli, the current study sought to extend this research towards the domain of surface features by employing stimuli that more closely resembled real-world music. To this end, melodic figuration and rhythmical activity was explored, and findings showed definite effects of both.

These results provide insight into the importance of surface features in music perception and understanding, bringing to question traditional musical practice which tends to overlook surface qualities. Traditional Western analysis, for example, brings the focus to chordal progressions while Schenkerian analysis further deepens musical examination to larger-scale (deeper) structures of the piece. Likewise, Western composition tends to group surface features as 'embellishments'; extra flourishes like passing or neighbouring tones are 'painted on' at the end. Though this way of thinking is certainly useful in understanding the larger-scale structure of a piece of music, surface features may be much more important for the maintenance of perceptual coherence than previously expected.

Acknowledgements. We would like to thank our students Tatiana Gayowsky, Vagisha Pruthi and Melissa Paul for their ideas, comments, and most of all, for their hard work. We would also like to thank Konrad Swierczek for his enthusiasm and invaluable input on this project.

References

Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception, 5*(2), 197-206.

Cuddy, L. L., & Thompson, W. F. (1992). Asymmetry of perceived key movement in chorale sequences: Converging evidence from a probe-tone analysis. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 971-684.

Farbood, M. (2016). Memory of a tonal center after modulation. *Music Perception, 34*(1), 71-93.

Spyra, J. (2017). The effects of time and events on the perception of nonadjacent key relationships. (Unpublished masters dissertation). McMaster University, Hamilton, Canada.

Woolhouse, M., Cross, I., & Horton, T. (2016). Perception of nonadjacent tonic-key relationships. *Psychology of Music, 44*(4), 802-815.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Designing Effective Auditory Interfaces: Exploring the Role of Amplitude Envelope

Sharmila Sreetharan,[1] Joseph Schlesinger,[2] and Michael Schutz[3,1]

[1] *Psychology, Neuroscience & Behaviour, McMaster University, Canada*
[2] *Vanderbilt University Medical Center, Vanderbilt University, USA*
[3] *School of the Arts, McMaster University, Canada*

[1]sreets@mcmaster.ca, [2]joseph.j.schlesinger@vanderbilt.edu, [3,1]schutz@mcmaster.ca

Abstract

A sound's shape over time or the amplitude envelope affects our cognitive and perceptual processing of sounds. Previous work has suggested that percussive tones, sounds with (i.e., naturally or exponentially decaying) amplitude envelopes, facilitate better encoding and retrieval of sounds than flat tones, sounds with (i.e., amplitude invariant) envelopes, during associative memory tasks. Flat tones are errantly used by most auditory interfaces, including the auditory alarms used by the global medical device International Electrotechnical Commission (IEC) 60601-1-8. The IEC alarms are criticized being difficult to be learnt and identified, easily confused, and are generally perceived as annoying. We manipulated the amplitude envelope of tones otherwise conforming to the IEC standard to explore its effect on learnability and annoyance. Forty participants were randomly assigned to learn alarm-referent associations composed entirely of either flat envelopes or percussive envelopes. Learning was measured by the number of training blocks required to learn the alarm-referent associations and the number of correctly identified alarms after a short break. Participants then evaluated alarm annoyance using a two-alternate forced task where trials varied by either alarm or envelope. Our results percussive tone sequences suggest reduce perceived annoyance without harming learning. These results highlight the importance of temporal considerations when creating auditory interfaces to reduce negative effects of prolonged exposure.

Introduction

Auditory Interfaces

Auditory interfaces rapidly convey information from machine to user using either speech or non-speech sounds and are particularly beneficial when visual interaction with devices is limited (Sodnik & Tomažič, 2015). From using Morse code during the 1800's to low battery levels and incoming email alerts on modern mobile devices, auditory interfaces have been employed for centuries using speech or non-speech sounds. Speech interfaces provide naturalness of communication and allow for multitasking during communication by requiring fewer attentional resources (Rosenfield, Olsen & Rudnicky, 2001). However, they are limited by language barriers, low robustness and accuracy in loud environments, a lack of privacy, and are slow forms of communication (Sodnik & Tomažič, 2015). Alternatively, non-speech interfaces use synthesized sounds conveying information using metaphors (i.e., auditory icons), abstract/symbolic sounds (i.e., earcons), or sped up spoken phrases no longer resembling speech (i.e., spearcons; Gaver, 1986). They can beneficially protect privacy, be universally understood, and rapidly conveyed (Gaver, 1986). However, unlike visual interfaces, it is difficult to create a set of concrete, intuitive metaphors (Sodnik & Tomažič, 2015).

Auditory Alarms

Non-speech interfaces are ubiquitous, but one prominent example is the use of auditory alarms in hospitals. The global medical device the IEC 60601-1-8 uses a standardized set of alarms to monitor patients and machine-related problems. These alarms are composed of three-note (i.e., medium priority) or five-note (i.e., high priority) melodic tone sequences with mnemonics (i.e., auditory icons) to aid in learning alarm-referent associations. Despite careful consideration by Block, Rouse, Hakala and Thompson (2000) on tone design specifications and mnemonics and Patterson (1990) on creating ergonomic auditory warnings in the creation of the alarm set, these alarms are riddled with problems. One study evaluating critical care nurses' abilities to learn these alarm-referent associations found only one nurse able to successfully identify 100% of associations after two training sessions, and that mnemonics did not improve identification association accuracy (Wee & Sanderson, 2008).

Overall, these alarms are frequently critiqued for being difficult to learn, easily confused, and poorly retained (Edworthy & Hellier, 2005; Sanderson, Wee & Lacherez, 2006). A large contributing factor in the poor discriminability and learnability of the IEC alarms is their lack of heterogeneity within the alarm set (Edworthy, Hellier, Titchener, Naweed & Roels, 2011). All tone sequences were synthesized within a narrow frequency range (262-523 Hz) in the key of C major, having similar contours, durations, rhythmic patterns, and starting on the note "C" (Edworthy et al., 2011). The large amount of similarity shared between the tone sequences is likely to contribute to alarm misidentification. Together, these identification errors contribute to the larger problem of alarm fatigue. Alarm fatigue results from sensory overload triggered by excessive alarms, lowering sensitivity and leading to missed alarms (Sendelbach & Funk, 2013). This negatively affects patient care due to the number of missed alarms and the noise pollution affecting recovery rates (Darbyshire & Young, 2013).

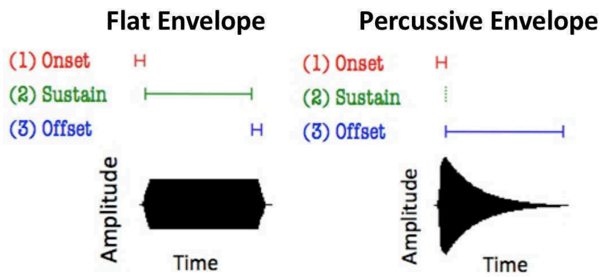


Figure 1. Sample wave forms of flat and percussive amplitude envelopes.

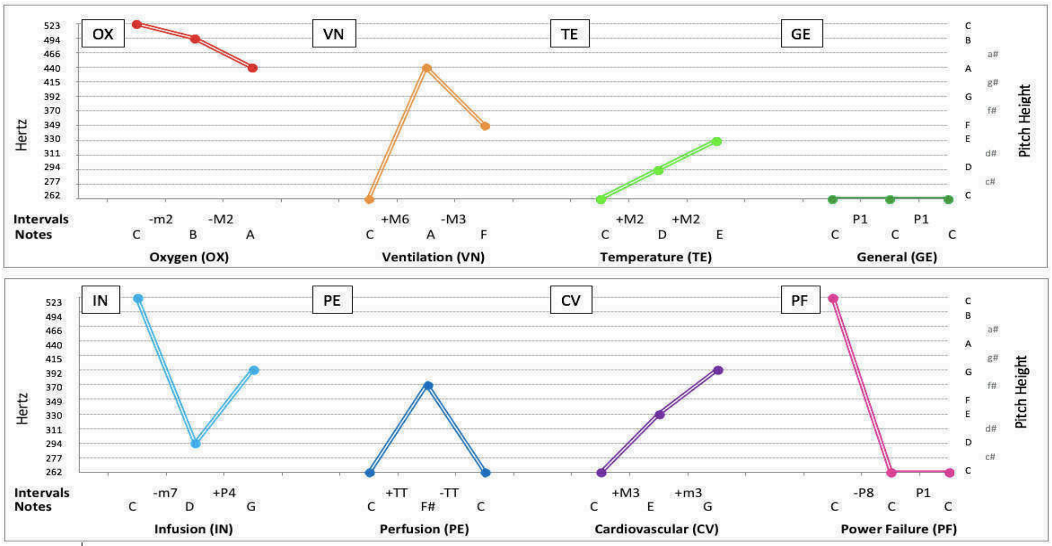


Figure 2. Contours of the medium priority IEC 60601-1-8 tone sequences used in the experiment. M = Major, m = Minor, P= Perfect, TT = Tritone, + = Ascending, - = Descending.

Amplitude Envelope

Amplitude envelope refers to a sound's temporal structure, specifically in reference to changes in a sound energy (i.e., intensity or loudness) over time. Flat envelopes are characterized by a rapid onset followed by a period of sustain, and a rapid offset, exhibiting minimal changes in a sound's energy (Figure 1A). Flat envelopes are characteristic of most artificially synthesized sounds, such as the emergency broadcast tone. Conversely, naturally produced sounds have percussive envelopes characterized by a rapid onset followed by an immediate, exponential decay (Figure 1B). Percussive envelopes are characteristic of everyday sounds including impact sounds, like the sound produced by two wine glasses clinking.

Amplitude envelope (hereafter referred to as "envelope") plays an essential role in giving details about impact events. For example, Lutfi (2001) demonstrated that the physical attributes of struck objects can be detected based on the varying acoustical features of percussive envelopes. Lutfi (2001) used a forced-choice task to determine specific acoustical features used in identification of a bar's hollowness (solid or hollow) and its material composition (iron, wood, or aluminium). Listener strategies that involved using envelope information aided in optimal performance levels (Lutfi, 2001). These findings, further extended by Giordano, Rocchesso, and McAdams (2010), suggest that envelope is a salient cue in determining the object hardness perception.

Envelope also plays an important role in audiovisual integration. Schutz and Lipscomb (2007) demonstrated in their "marimba illusion" that the length of a striking gesture can influence the perceived duration of percussive sounds produced by a marimba. This suggests visual information largely influences judgements of tone duration tasks, despite a popular notion that audition dominates such judgements (Fendrich & Corballis, 2001). A follow-up study by Schutz (2009) suggests this atypical pattern of integration stems from the amplitude envelope of percussive tones. When using flat tones, audition dominates tone duration judgements but when using percussive tones, visual dominates tone duration judgements. The over-reliance on flat envelopes in auditory perception research (Schutz & Vaisberg, 2014) seems to have led to conclusions from artificial tone beeps that fail to generalize to natural sounds.

Additionally, it is suggested that percussive sequences are rapidly processed compared to flat sequences during associative memory tasks (Schutz, Stefannuci, Baum & Roth, 2017). Schutz et al. (2017) examined the effect envelope on tone sequence-object associations and found that percussive tone sequence-object associations were learned more rapidly, were correctly recalled more frequently than flat tone-sequence-object associations. This suggests that amplitude envelope may play an important role in associative memory, especially during encoding and retrieval stages (Schutz et al., 2017).

Current Experiment

Like most non-speech interfaces, the auditory alarms used by the IEC 60601-1-8 exclusively uses tone beeps (i.e., tone sequences with flat envelopes). Our experiment aims to explore the potential benefits of amplitude envelope manipulation on the IEC alarm tone sequences while preserving the current alarm-referent associations. Specifically, we examined the effect of incorporating percussive envelopes into the IEC medium priority alarm tone sequences to assess the effect on learnability and annoyance.

Method

Participants

Forty undergraduate McMaster University students (35 female, 5 male) ranging in age from 18-23 years (M = 19.08, SD = 1.40) recruited from undergraduate psychology and linguistic research pools participated in this study and received 1 research course credit as compensation.

Stimuli and Apparatus

We synthesized eight tone sequences according to the frequency and amplitude envelope requirements in IEC 60601-1-8 standard for medium priority alarms using SuperCollider – a real time audio synthesis software (Figure 2). All tone sequences were within the range of 261.4 Hz (C3) - 523.4 Hz (C4). Unlike the IEC standards, here we used pure tones (sine waves) rather than more complex tones (an issue we will explore in future experiments). "Flat alarms" exhibited flat envelopes conforming to the temporal constraints outlined by the IEC standards (Figure 3A). "Percussive alarms" exhibited overlapping percussive envelopes (Figure 3B); since the envelopes overlapped for percussive alarms, the peaks amplitudes were set to be equivalent as this was sufficient to achieve an approximate equivalent perceived loudness for the two tone types. The total time sweep was 1040 ms for the flat alarms and 1180 ms for the percussive alarms. The last tone for the percussive alarms was approximately 1.6 times longer (after the onset) than the last tone for the flat alarms; this accounted for the shorter perceived duration of percussive envelopes and achieved an approximate equivalent perceived duration for the two tone types (Grassi & Darwin, 2006).

Tone sequences were appended together using Audacity – a free sound-editing software and stored on a MacBook Air laptop. Participants listened to these sounds through Sennheiser HDA-300 headphones. In addition, participants completed a consent form and a short demographic survey inquiring demographic information and musical background. Participants received cue cards with each of the eight alarm names and were permitted to re-arrange them to aid them in learning associations.

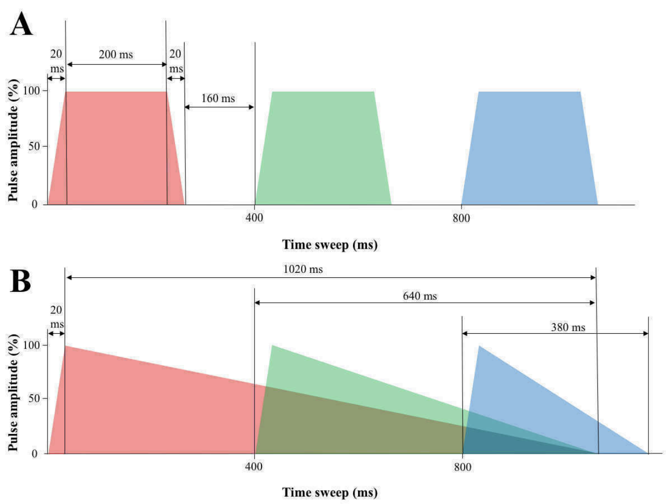


Figure 3. Temporal characteristics of flat (A) and percussive (B) tone sequences used in the experiment. For simplicity, exponential tone onsets and offsets are depicted to be linear.

Procedure

The experiment consisted of two separate assessments: the first examined learning of sequence-command associations and the second annoyance with the sequences themselves (Figure 4). We randomly assigned participants to either the "flat" or "percussive" conditions for the learning paradigm, dictating whether they heard "flat" or "percussive" alarm-referent pairings throughout the entirety of the learning paradigm. The learning paradigm was split into four stages: study phase, training phase, break, and evaluation phase.

Study phase. Participants heard the eight tone sequences-referent pairings twice. The alarms were presented in a random order to prevent order effects.

Training phase. During the training phase, participants heard an alarm and verbally indicated the referent. Regardless of their response, participants were reminded of the correct alarm-referent pairing. The end of the training phase occurred when participants correctly identified 7/8 alarms on two consecutive blocks (each block containing all alarms in the alarm set) to a maximum of ten blocks.

Break. After, participants had a five-minute break when they played a silent game of mini-putt, which was implemented to prevent mental rehearsal.

Evaluation phase. Participants completed a final identification task on one block of alarms during the evaluation phase in addition to reporting the confidence of their response on a 6 point Likert-type scale. Up to this point in the experiment, participants only heard sounds with one type of envelope (percussive/flat).

After the learning paradigm, participants completed an annoyance task when they identified alarm they perceived to be more annoying during a two-alternative forced choice task. The task contained 16 trials—with eight trials differing only in envelope type and eight trials differing only in alarm type. For example, trials comparing the percussive-general vs. flat-general alarms forced assessment based on envelope, whereas trials comparing the percussive-temperature vs. the percussive-cardiovascular alarms forced assessment based on melodic structure. All participants heard each alarm-envelope combination twice throughout the task and all trial types were mixed and randomized during the task. As the trials comparing alarm types served primarily as distractors to reduce task transparency, here we report only analysis of trials comparing envelopes.

Results

Learning Paradigm

A between-subjects t-test conducted on the number of blocks to reach criterion based on condition during the training phase revealed a non-significant effect of condition, $t(38)=0.52$, $p>.05$. A between-subjects t-test conducted on the number of correctly identified alarms based on condition during the evaluation phase revealed a non-significant effect of condition $t(38)=1.94$, $p>.05$.

Annoyance Task

Overall. A one-sample chi-square goodness-of-fit test revealed flat alarms are more annoying than percussive alarms, $X^2(1, N=320) = 99.01, p<.05; w=.56$ (Figure 5A).

Percussive alarm training. A one-sample chi-square goodness-of-fit test revealed participants training on percussive alarms found flat alarms considerably more annoying, $X^2(1, N=160) = 115.60, p<.05; w=.85$ (Figure 5B). Furthermore, a series of one-sample chi-square goodness-of-fit tests with Holm-Bonferroni corrections indicate that participants found flat versions of each alarm type (e.g., general alarm, power failure alarm) more annoying than their respective percussive counterparts (all p's<.05).

Flat alarm training. A one-sample chi-square goodness-of-fit test revealed participants training on flat alarms also found flat alarms considerably more annoying $X^2(1, N=160) = 11.03, p<.05; w=.26$ (Figure 5C). A series of one-sample chi-square goodness-of-fit tests were conducted assessing whether the flat or percussive versions of each alarm type were equally perceived to be annoying for those trained on flat alarms. All tests were non-significant, suggesting that annoyance ratings for flat and percussive versions of all alarm types are not statistically different (all p's>.05).

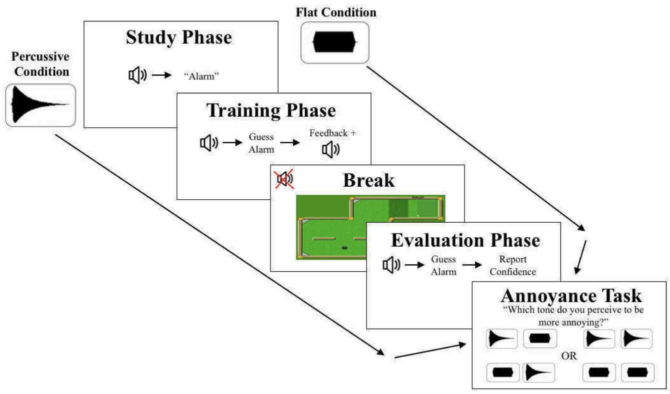


Figure 4. Overview of experimental paradigm. Condition assignment was reflective of envelope type of the sounds heard during the learning paradigm (i.e., study phase, training phase, break, evaluation phase). The annoyance task, common to both experimental conditions, was comprised of sounds with both percussive/flat envelopes.

Discussion

Learning paradigm

Performance during the learning paradigm suggests that our percussive tone sequences do not affect training required to learn alarm-referent associations nor their recall rate after a short break. We failed to replicate the results by Schutz et al. (2017) suggesting percussive tone sequences aided learning and recall of tone sequence-object associations. A distinguishing feature between our experiments is that Schutz et al. (2017) measured associative memory using concrete objects, whereas the alarm-referent associations we used were comprised of abstract commands/states. According to the dual-coding theory during encoding and retrieval, concrete words use both verbal and multi-modal sensory systems whereas abstract words only use the verbal system (Paivio, 1991). Due to the abstractness of alarms, it is possible that a greater training period would be required for any observable differences between the flat and percussive alarms.

Annoyance task

Annoyance task performance suggests that flat alarms are more annoying than percussive—regardless of sounds heard during training. However, those trained on percussive alarms found flat alarms particularly annoying. Breaking down responses by eight alarms types, all of the flat versions were chosen as more annoying by 80%-100% of participants. Consequently, the flat version of the eight alarms appears significantly more annoying than the percussive—despite each participant providing a single rating per alarm type.

Those trained on flat alarms still found flat more annoying, although to a lesser degree than when analyzing ratings of those trained on percussive alarms. When broken down by the eight alarm types, we found no statistically significant differences based on envelope for any individual alarm. Nonetheless, flat versions were chosen to be more annoying in by 50%-70% for individual alarm sequences. Together these findings suggest percussive alarms are generally less annoying than flat, although the degree of difference varies based on exposure and training.

Novelty effect

We believe the effect of training on annoyance ratings to reflect novelty's effect on annoyance ratings. When trained on flat alarms and exposed to the percussive alarms, participants are newly exposed to overlapping tone sequences, characteristic of our set of percussive tone sequences. It is possible that this overlapping tones may accentuate consonance and dissonance Western harmonic structure of the tone sequences. According to Western musical classifications of consonance and dissonance, simple frequency ratios (e.g., unison, major third, and perfect fifth intervals) are considered consonant whereas complex frequency ratios (e.g., minor second, tritone, minor seventh intervals) are considered dissonant (Schellenberg & Trehub, 1994). We noticed a slight trend towards percussive versions of alarms with dissonant intervals being rated as more annoying than consonant intervals. When individual notes in the tone sequence are separate, it is likely the interval relationships go unnoticed, especially for those lacking musical training. Since we did not control for musical training, it is possible that varying amounts of Western musical training may have contributed to the observed pattern of results.

Implications

Our envelope manipulation had no effect on learnability; however, it did affect short-term annoyance ratings of the alarm tone sequences. Together these results suggest there is great potential benefit to exploring more varied envelope shapes when designing auditory alarms in medical devices. Although learnability is a concern when introducing medical

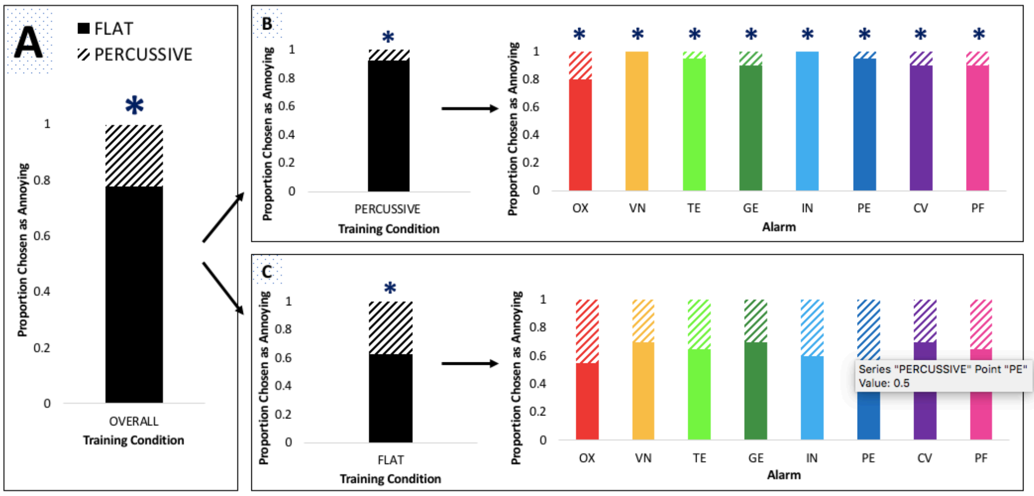


Figure 5. Annoyance task results for all participants (A), broken down by those trained on percussive alarms (B) or flat alarms (C) during the learning paradigm. *$p < .05$

professionals to current alarms, alarm annoyance is a long-term persisting problem affecting both clinicians across a career and patients over the duration of their hospital stay.

Auditory alarms. An observational study found that on average there about two critical alarms missed per hospital, per day in the United States (Donchin et al., 2003). Extrapolating this number across every hospital in entire country over the span of one year roughly equates to about 4.2 million alarms missed annually in the US. Edworthy (1998) suggests as alarms become increasingly irritating, there is a higher likelihood of the device being shut down or severe volume reduction. Although there are many contributing factors to missed alarms, it is likely that alarm fatigue is one (Cvach, 2012). Currently, there are implementations in practise to reduce alarm fatigue, yet there are few strategies in place to reduce alarms missed critical alarms (Rayo & Moffatt-Bruce, 2015). Our envelope manipulation offers one solution to reduce annoyance in hopes of reducing the number of missed critical alarms.

Auditory Interfaces. Although the advancement of technology has led to widespread use of higher quality speakers and tone generators, the sounds used in medical devices routinely fail to take full advantage of these possibilities. Presently, many auditory interfaces employ “tone beeps” that become incessant over time. The most basic example to illustrate this is a cashier who receives auditory feedback of a “beep” each time an item passes over the scanner. These sounds create a cacophony of beeps between cashiers become both annoying and confusing (Edworthy, 1998). Amplitude envelope manipulation can create more aesthetically pleasing sounds with greater heterogeneity. One study evaluating effect on envelope on product value found that participants were willing to pay more for cell phones with percussive ring tones than flat ring tones (Schutz & Stefanucci, under review). Despite the practical applications, those designing the alarms are focused on the nuances of the machine itself and the importance of the design of the sounds is lost. Our results emphasize the importance of temporal considerations of sounds in the creation of effective auditory interfaces.

Future Directions

Our current experimental design did not take into account the effect of variable alarm exposure on annoyance due to the variable amount of blocks required to learn the alarm-referent associations. It is likely that exposure is positively correlated with annoyance; future experiments will control for amount of exposure prior to assessments of annoyance. Furthermore, we did not assess annoyance prior to the beginning of the experiment. It is possible that annoyance ratings may have been influenced by pre-existing biases. It is unclear whether individuals with high exposure to sounds with flat envelopes present reduced or greater sensitivity to such sounds affecting their annoyance ratings. We aim to account for this by assessing annoyance prior to the beginning of the experiment to understand how envelope manipulation may influence any pre-existing biases. In consideration of ecological validity, sounds with acoustically richer structure should be implemented to combat masking from environmental noise (Patterson, 1990). Our stimuli consisted of pure tones to determine how effective amplitude envelope manipulation is at the most basic level. In the future we plan on using complex tones with greater number of harmonics in subsequent experiments.

To the best of our knowledge, these results represent a novel approach to improving alarm design with great potential to assist the larger problem of auditory alarm annoyance in medical settings. As such, we invite feedback and comments from the research presented on this extended abstract.

Acknowledgements. None of this work have been possible without Carmen Dang, Cameron Anderson, Rebecca Benjamin, and Neerjah Skantharajah for their assistance with data collection in addition to Aimee Battcock and David Prete for their ongoing support. We would also like to acknowledge Drs. Pratik Pandharipande and Matthew Weinger from the

Vanderbilt University Medical Center Department of Anesthesiology. This wok was supported from the grants to Michael Schutz from the Natural Sciences and Engineering Research Council of Canada and the Canadian Foundation for Innovation [grant number CFI-LOF 30101].

References

Block, F.E., Rouse, J.D., Hakala, M., & Thompson, C.L. (2000). A proposed new set of alarm sounds which satisfy standards and rationale to encode source information. *Journal of Clinical Monitoring and Computing, 16,* 541-546.

Cvach, M. (2012). Monitor alarm fatigue: An integrative review. *Biomedical Instrumentation & Technology, 46(4),* 268-277.

Darbyshire, J.L., & Young, J.D. (2013). An investigation of sound levels on intensive care units with reference to the WHO guidelines. *Critical Care, 17(5),* R187.

Donchin, Y., Gopher, D., Olin, M., Badihi, Y., Biesky, M., Sprung, C. L., … Cotev, S. (2003). A look into the nature and causes of human errors in the intensive care unit. *Quality and Safety in Health Care, 12(2),* 143–148.

Edworthy, J. (1998). Does sound help us to work better with machines? a commentary on Rauterberg's paper "About the importance of auditory alarms during the operation of a plant simulator." *Interacting with Computers, 10(4),* 401–409. https://doi.org/10.1016/S0953-5438(98)00036-8

Edworthy, J., & Hellier, E. (2005). Fewer but better auditory alarms will improve patient safety. *Quality and Safety in Health Care, 14(3),* 212–5. https://doi.org/10.1136/qshc.2004.01305

Edworthy, J., Hellier, E., Titchener, K., Naweed, A., & Roels, R. (2011). Heterogeneity in auditory alarm sets makes them easier to learn. *International Journal of Industrial Ergonomics, 41(2),* 136–146. https://doi.org/10.1016/j.ergon.2010.12.004

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63(4),* 719–725.

Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human–Computer Interaction, 2(2),* 167–177.

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance, 36(2),* 462–476.

Grassi, M., & Darwin, C. J. (2006). The subjective duration of ramped and damped sounds. *Perception & Psychophysics, 68(6),* 1382–1392.

Lutfi, R. A. (2001). Auditory detection of hollowness. *The Journal of the Acoustical Society of America, 110(2),* 1010–1019.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45(3),* 255-287.

Patterson, R. (1990). Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society of London, 327(1241),* 485–492.

Rayo, M. F., & Moffatt-Bruce, S. D. (2015). Alarm system management: evidence-based guidance encouraging direct measurement of informativeness to improve alarm response. *BMJ Quality & Safety, 24(4*), 282–286. https://doi.org/10.1136/bmjqs-2014-003373

Rosenfield, R., Olsen, D., & Rudnicky, A. (2001). Universal speech interfaces. *Interactions, 8(6),* 34-44.

Sanderson, P., Wee, A. N., & Lacherez, P. (2006). Learnability and discriminability of melodic medical equipment alarms. *Anaesthesia, 61(2),* 142–147.

Schellenberg, E.G., & Trehub, S.E. (1994). Frequency ratios and the perception of tone patterns. *Psychonomic Bulletin & Review, 1(2),* 191-201.

Schutz, M. (2009). *Crossmodal integration: The search for unity.* University of Virginia.

Schutz, M., & Lipscomb, S. D. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36(6),* 888–897. https://doi.org/10.1068/p5635

Schutz, M., & Vaisberg, J.M. (2014). Surveying the temporal structure of sounds used in music perception. *Music Perception, 31(3),* 288-296.

Schutz, M., Stefanucci, J., Baum, S., & Roth, A. (2017). Name that percussive tune: Associative memory and amplitude envelope. *Quarterly Journal of Experimental Psychology, 70(7),* 1323-1343.

Schutz, M., & Stefanucci, J.K. (under review). Hearing value: Exploring the effects of amplitude envelope on consumer preference. *Ergonomics in Design: The Quarterly of Human Factors Applications.*

Sendelbach, S., & Funk, M. (2013). Alarm fatigue: A patient safety concern. *AACN Advanced Critical Care, 24,* 378-386.

Sodnik, J., & Tomažič, S. (2015). *Auditory Interfaces. In Spatial Auditory Human-Computer Interfaces* (pp. 33–44). Cham: Springer International Publishing.

Wee, A. N., & Sanderson, P. (2008). Are melodic medical equipment alarms easily learned? *Anesthesia & Analgesia, 106(2),* 501–508. https://doi.org/10.1213/01.ane.0000286148.58823.6c

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Empirical Findings on the Influence of Artist Image in Music Evaluation

Alexander Stamatiadis

Department of Music, University of Sheffield, England

alex.stamatiades@gmail.com

Abstract

In musicological theory of artist image it is maintained that a coherent artist image, along a display of authenticity and authorship, will aid to a positive evaluation of an artist's music by listeners (Mäkelä, 2004; Ahonen, 2008). This idea has been developed through case study analyses of the images of popular artists, and the reactions of their fans on those as found on music magazines and internet forums. The present work sought to test the artist image theory empirically, namely whether the presentation of a fine-tuned image, authenticity and authorship can influence the evaluation of music positively. A derivative aim was to test if the music listeners are actually sensitive to clues related to the projection of the image parameters of authenticity and authorship. The study used a survey-based experiment that presented 8 different fictional artist images (specifically: artist name, picture, short biography) and 8 corresponding music clips. After the presentation of an image, the survey asked the responders to answer questions related to authenticity and authorship clues presented in the image, and to rate the artist's music clip. The music clips were studio recordings that were edited in order to sound like live recordings acquired through a handheld device, in order to control for the sound quality variable which is confound in the context of comparing professional artists (image and sound handled by experts) with non-professional artists (image and sound potentially handled by individual or non-experts). The results indicated that explicit mention of the "success" of an artist does not influence the evaluations of the music. On the other hand, the results seemed to confirm what has been suggested by the musicological research about the positive effect of presented authenticity and authorship.

Introduction

In popular music, the idea of artist image plays an important role in multiple ways and on different aspects of the industry. Significantly, an image can act as an element of unification and classification for a body of work that an artist has produced (Goodwin, 1992: 103). At the same time, a distinctive image can make the artist differentiate from other artists in the crowded industry of popular music. Another advantage of the maintenance of a public image is its "trademark" quality, that helps the association of the performed songs with the artist (Brackett, 2000:14-15; Toynbee, 2000:61), potentially making the song more attractive to the artist's existing fan base. Therefore, the creation and the support of a coherent and unique author image is a desirable function for any popular music artist, and it is commonly handled by a team of expert marketing personnel and image making professionals, operating in order to enhance an artist's vision of their own image (or create it, if the artist is not interested in the matter) (Negus, 1992: 64-65).

An artist image acts as a way to rise the public awareness and extend the reach of the artist to a constantly bigger audience. As such, it is in favour of the artist and his company to create an image that follows the tradition of the images already established in the specific musical genre in which the artist operates. For example, pop artist images generally focus on looks, fashion and commercial achievements, embracing the superficiality that is connected to commercialism (Butler, 2003). On the other hand, the electronic music oriented, less-known artists, generally maintain an "anti-man" and robotic style that is based on the hiding of the real identities of the artists (Reynolds, 1998; McLeod, 2003). They achieve this through wearing masks and manipulating their photo shootings in order to cover their faces, and by preserving an identical appearance between band members in attempt to minimise each member's individual characteristics. This direction extends to the album covers and general use of graphics, where they prefer to show pictures that do not contain the band members or humans in general (Tandt, 2004).

Despite the differences in image formation that are found between musical genres, we can identify some specific properties of artist images which are thought to be linked with specific positive traits that the artists seem to possess. By writing positive, I refer to traits that make music listeners to value the corresponding artists high in the hierarchy of popular music creators. These properties can be categorised in two themes, the one being *authorship*, and the other *authenticity*.

Authorship is a notion related to the creation of one's own music, and the control over the technical aspects of the production process of the music (Ahonen, 2008). If their work is followed by positive evaluation by the audience and media, artists with authorship over their music are highly regarded and thought of as carrying their own style and personality over their work, through following their personal artistic vision. This justifies authorship as positive trait.

The other theme of property categorisation is the one of authenticity. While this term is traditionally linked to the faithful reproduction of a pre-existing musical material (Young, 2001), in popular music it considers the public image of the artist (Dyer, 1986). Even if authenticity can be seen as a historical and social construction that may undergo many changes through time, it is a notion of focus for artists, the media and the audiences (Moore, 2002) and through its consideration as a property of respected artists, it can also act as a tool for the evaluation of music (Mäkelä, 2004). In her analysis of the public image of the artist Jewel, Laura Ahonen (2008) performs a meta-analysis of the views of music critics and Jewel's audience, in order to understand how Jewel's radical change in music style and general image (from a folk-rock storyteller living in a van in Antarctica to a dance-pop star) influenced their perception of her authenticity. In the conclusion of the corresponding chapter, she considers the criticism of Jewel's authenticity to focus on the opposition of notions relating to what the artist would do if she operated from a position of creating music from her personal aesthetic preferences and life experience, contrasting with the position of publishing music that is meant to be pleasing to a large, commercial music oriented fan base. Specifically, Ahonen views the criticism as a comparison between the creation processes of her early folk-rock, storytelling oriented albums, and the later dance-pop, technologically complicated and densely arranged music. She mentions the criticism that was raised as a result of Jewel's use of songwriters, beat makers and other collaborators, her "slick" pop sound, and in general, her commercially oriented work process and musical direction. Conclusively, we could describe authenticity at minimum as a combination of originality and of creative output without focus on commercial viability.

It must be noted that authorship and authenticity are not strictly defined, and their nature can change through different contexts and time periods. Also, the very image of an artist is dependent not only on the artist's own presentation of their image, or the presentation by the responsible marketing team, but it is a combination of this presentation with the one of the media (through e.g. album reviews and interviews) and the reception of these two sources by the music listeners. It can be said that the image of an artist is different for every music fan who receives and combines all the different sources of information about the artist, in order to create their own personal view. Nevertheless, as defined above, authorship and authenticity are thought to positively influence the image of an artist, which possibly further influences the reception of the artist's music. In the present study I examine this hypothesis; that the desired factors in an artist image influence the perception of the songs of the artist by listeners. Specifically, I examine the link between the evaluation of songs by listeners and artist image properties. Thus, the examination must precisely consider the songwriting, and not further properties of the published music, such as production quality and marketing efforts.

Method

The experiment of this essay was based on an online questionnaire. What it tried to find is what happens to the rating of a song by a music listener when the author of the song is thought to possess certain artistic virtues in varying levels. In order to achieve this, it presented to the subjects 8 different artists. Specifically, there was presented one picture of the artist, followed by a short biography. After these, it was also presented a song clip of 30 seconds duration. The subjects were asked to rate the artists on the basis of how much they possess certain artistic properties (this information is to be derived from the short biography), and to also rate how much they liked the accompanying song clip. There were used 5 questionnaires for randomisation of the presentation order.

Scale Questions

As discussed in Introduction, there exist specific properties that are thought to positively influence an artist's image appreciation. They were divided into two themes, namely authorship and authenticity. The differentiation of the two themes comes from the fact that authorship requires a control over the processes of the production, in order for the artist to bring their unique artistic vision to existence. On the other hand, authenticity is a more abstractly defined notion that can mean different things. However, it revolves around the idea of originality and artistic "purity" of the output. The study addressed how the subjects feel regarding the acquisition of these properties by the presented artists by asking the subjects to rank the artists and their song using relevant scales. After exposing them to the artist's picture (photo shooting or graphic design) and an accompanying short biography, they were asked to what extent do they agree with some following statements, "having read the artist's biography".

The first one was about authorship and it read:
"[Artist Name] is handling most technical aspects of his artistic output." and the scale ranged from 0 ("Not at All") to 10 ("Very Much"). In accordance to what was discussed in Introduction, this scale gathered data relating to a characteristic exclusively corresponding to authors, in order to find how much did the artist fulfill the criterion to be considered as such.

The next statement considered the theme of authenticity:
"[Artist Name] is an authentic artist." and it used the same scale as the first question. Because authenticity is an ambiguous notion, the statement was presented in a general way in order to let every subject interpret it according to their own view. This did not harm the experiment, because the analysis wanted to find how the perception of an artist as authentic influences the perception of their music, and not to force an agreement between subjects about what constitutes authenticity. Finally, the order of the statements was made this way in order to present the hardest to answer question, the one about authenticity, last.

After the assessment of the artist's image there was presented a song written by the artist. The corresponding questionnaire title was "One of [his/its] songs (randomly chosen clip)", in order to make the song seem like a random example of the artist's output, and not an exemplary work. This could potentially help the subjects to not feel bad in case they wanted to give a low rating, which could alter the responder's honesty (Stathakopoulos, 2007). After listening to the 30-seconds song clip, the responders answered the question "How much did you like the song in the clip?" by rating it in a scale which is identical to the ones of the artist image evaluation statements. The question was expressed this way in order to emphasize that what is of importance is the song itself and not the sound quality of the clip.

Artist Images

The artist images used in the questionnaire were designed with the purpose of assessing the different image parameters that are thought to influence the perceived status of an artist as author and authentic. At the same time, I utilised the name and the picture of the artists in order to aid in their classification either as professional and successful artists, or hobbyist and less successful. The creation of the artist images consisted of two different parts which included all the information that the responders did get about the artists.

The first part was about the visual presentation of the artists, and specifically about a picture that shows the artists, or a graphic design about them, that intended to present them in a positive way and to bring their individual style and talents into focus. The division of the 8 artists into 4 professional or good and 4 hobbyist or bad ones, meant that their pictures should differ accordingly. The good artists had pictures that were made by professional photographers and graphic designers, whereas the bad artists had ones that were made by non-professionals, and the photo editing or the graphic design was supposedly done by the artists themselves. In order to achieve these effects, for the 4 good artists I chose professionally situated and captured photographs of real artists or models. For the bad artists I chose pictures for which the photographer did not have an artistic intention, or alternatively I chose some graphics, and then I edited both kinds of pictures in order to make them look like they were edited by someone with little graphic design knowledge, who tried to make the pictures more artistic and interesting. Other than the artist pictures, the second part of the image creation had to do with their biographies.

For the biographies, a common theme was that of artist's origin and current place of residence, or place of higher education. In order to divert the responders' attention from the fact that I was examining how the images of the presented artists influence their ratings about the songs, I presented the study and wrote the texts of the questionnaire as if the study was about how living in a country of foreign culture influences the music creators in their artistic output. In this context, in the biographies I was mentioning that the artist is either Armenian or partly Armenian, and the place in which they now live or studied at some point. After this part, I wrote material that corresponded to the characteristics that influence the perception of the position of an artist in the hierarchy of popular music quality (authorship, authenticity and success, or lack thereof). In accordance to this purpose, in the biographies of the artists that had the X characteristic I wrote elements that made these specific characteristics prevail (e.g. record sales, production autonomy), and in the ones of the artists that did not have the X characteristic I wrote elements that had the opposite effect in their status as possessors of X, or (in 1 case) I entirely omitted discussion about the specific characteristic.

Some parts of the biographies were real, some were based on the real biographies but were altered to fit the experiment's objectives, and some were fictional. But, either in the case of the pictures or of the biographical elements, whether the information was real or not did not matter for the purpose of the experiment, as long as it did not make the artists' existence to be not believable by the responders. In the pilot phase, there was no feedback about the nature of any artists' existence looking as fake. When asked, one subject said that she thought all artists were real. Lastly, the names of the artists of the "good" category were either names of real successful artists, or fictional with the purpose of looking as such. The names of the other category were more difficult to pronounce, or were reminiscent of names with characteristics of the ones that frequently belong to less successful artists (i.e. not exciting to imagination).

As noted, the artist images presented in the questionnaires were 8, with 4 of them seeming professional or good, and the other half seeming bad or hobbyist. Independent of that, in the images there was given information about the possession of specific characteristics by the artists, which corresponded to the strengthening or weakening of their status as authors and authentic creators. For authorship, the information given concerned the artist's handling of the technological aspects of the artistic output, such as production, arrangement, mixing and mastering of the music tracks. For authenticity the information was about how original did the output seem to be, which derived from creation of the music by the artists instead of other songwriters, the want to create material from their own perspective instead of following the mainstream trends or artists that they admire, and the absence of focus on the potential for chart success of their music in general. To recap, Table 1 shows all the different combination of image properties and which artist corresponded to which combination (artists separated by comma).

Table 1. Distribution of presented characteristics of the 8 artists.

Positive Presentation	Negative Presentation
Successfulness	
Daron Malakian, Viza, Buckethead, Harry Styles	Chnages, The Hard, Caveboy, Narek Hagopian
Authorship	
Daron Malakian, The Hard, Viza, Buckethead, Narek Hagopian	Chnages, Harry Styles, Viza, Caveboy
Authenticity	
Daron Malakian, Viza, The Hard, Caveboy	Chnages, Harry Styles, Buckethead, Narek Hagopian

Sound Clips

The assessment of the songs should have a clear focus on the songs themselves, as in songwriting, as opposed to good sound quality, use of high-end studio equipment etc. These variables can be obstacles when one wishes to assess the preference differences of subjects between the songs of hobbyists and professional artists, because they can act as confounding variables that distort the perception of the subjects on the quality of the songwriting itself. Professional artists generally have a bigger budget at their disposal, which can allow them to work with more skilled music producers and engineers, record with more expensive equipment and achieve a more impressive sound than the one of do-it-yourself, low budget musicians. The production and sound quality of the music tracks is an aspect that can influence the subjects in their appreciation of songs. Therefore, I eliminated this aspect with the introduction of an editing process which made the songs sound like they were recorded at a live concert setting through a handheld recorder (e.g. a mobile phone of a member of the crowd). To achieve this I emulated the chain which the sound goes through during such a recording process. In the pilot phase the responders thought that the clips contained recordings made through a handheld device and that they were not studio recordings. Therefore, this process could potentially be of interest to someone who would like to eliminate the high-end production elements of songs and make them sound like live recordings captured by a handheld device.

Firstly, we have to consider what happens to the sound when a handheld device records it at a concert: The sounds of the instruments come out of the PA speakers of the venue at a high volume (also producing a strong reverberation). The small microphone of the device captures this output, along the crowd noises. The bass is too strong for the microphone, making a booming sound and overriding the other elements in degrees influenced by the amount of presence of the bass. When this occurs, the rest of the sounds become ducked (i.e. they lose volume when the bass rises in volume). Also because of the high volume, the recording can occasionally get louder than the digital 0 dB and the sound gets distorted. The combination of these factors create the sound we are used to hear when listening to such recordings.

The emulation of this chain was made in the computer using the following steps:

1. On the studio version of the song, I used the effect of an Impulse Response (IR) that emulates the PA speaker output at a club that hosts live concerts. I used an IR captured in the club "Besame Mucho" in Zurich, Switzerland which is a square shaped club with a capacity of approximately 250 people. The use of the same IR in all song clips ensured that they will all get a similar sound, no matter the prior sound design and quality.

2. I blended in the sound of crowd noises recorded in a club with high quality sound.

3. I inserted a compressor effect on the output of sound sources 1. and 2. to emulate the compression occurring in the phone speaker. Since almost all studio-produced music recordings are at a level of about -5 dB to -1 dB, keeping the threshold of the compressor at about -25 dB and ratio at 1.4:1 was a safe setting that would not need to change from song to song, forcing all clips to have not only a similar sound but similar sound dynamics as well.

4. Lastly, a subtle distortion effect on the final signal was applied, in order to emulate the distortion occurring in these recordings due to the high volume of the sound inserted in the microphone.

This is how the song clips were created, in order to overcome the confounding aspects of sound quality and equipment differences between professional artists and hobbyist music makers. In fact, the songs selected were all professionally produced. But in case the results of the study would imply that a "successful" artist image did not seem to play a role in the appreciation of the songs by the responders, our focus should shift to the fact that other than the image and status, the music production and general sound quality is also a factor of difference between successful and less successful artists. Therefore, the appreciation of more successful artists with the same songwriting skills (or songwriters) as less successful ones could be considered at least partially as a product of the sound quality superiority, and not of the assumed influence of the "successful status" of the artist.

The length of the clips was ~30 seconds, in order to prevent fatigue that could influence the attention and the responses of the subjects. The songs used were either relatively unknown songs of (commercially) successful artists ("Scars on Broadway – Hungry Ghost", "Viza – Sans Red", "Buckethead – Moss", "The Script – Walk Away"), relatively successful songs of less successful artists ("Immaculate Machine – So Cynical", "Expert Medicine – Perfect Maniac") or songs released close to the time of writing by artists reviewed by media as good but are not well known ("Cyanna Mercury – Nothing We Can Do", "The Orbellion – Imposter"). With this selection procedure I tried to keep the songs at a comparable level in terms of songwriting quality, although each subject can have a quite difference appreciation of the same song (Christenson & Peterson, 1988; Rentfrow & Gosling, 2003, 2006).

Sample

The questionnaire was sent to University of Sheffield and University of Piraeus students of randomised years and subjects of studies, and of random ethnicities and genders (N=32).

Results and Discussion

The matters that were analysed in the study can be divided in three distinct questions:

- How does the possession of author and / or authentic status by an artist influences the ratings of their songs by listeners?
- How does the possession of a professional vs. unprofessional presented image by an artist influences the ratings of their songs by listeners?
- To what extent can subjects evaluate the possession of such statuses when given an artist's short biography?

For the first question I analysed N=256 individual answers for the 8 artists. In order to examine this link I wanted to test if there existed a correlation between these factors and the song ratings. The ratings in all scales did not seem to follow a normal distribution, with sig. = ,000 for all variables and on both Lilliefors-corrected Kolmogorov-Smirnov (Lilliefors, 1967; Conover, 1999) and Shapiro-Wilk (Shapiro & Wilk, 1965) tests. Because of this and even though the sample was relatively large, instead of using the Pearson correlation test which is not robust (Wilcox, 2011) to the normality violation (Hufthammer, 2014), I chose the Kendall's Tau-b (Kendall, 1938) and Spearman's Rho (Spearman, 1904) tests.

It was found that Authorship and Authenticity did correlate significantly and positively with Song Ratings: Kendall's Tau-b resulted in 0,124 and 0,205 correlation coefficients respectively, while the values of Spearman's Rho were 0,160 and 0,262. While the coefficients did not indicate a very strong correlation, they did indicate a 0,01 level highly significant, weak to moderate positive association. This result can be interpreted as the effect of influence of Authorship and Authenticity Ratings on Song Ratings. Even though correlation is a measure of association and not of causal relationships (Mukaka, 2012), the design of the experiment allows us to further interpret this association as an indicator of statistically significant effects, because of the order of the questions in the questionnaire. Authorship and Authenticity always preceded Song Ratings, so the last could never influence the answers on the two image factors.

For the second and third questions I followed a two-step procedure. At first I assessed the Means and Variances for all variables (Song Ratings, Authorship Ratings and Authenticity Ratings) to decide how to proceed with the analysis. The results indicated a mixed outcome in the various ratings between the positive and negative artist images. As such, the analysis would not benefit from a pairwise comparison of the results of each artist. In order to answer the questions #2 and #3 I proceeded with the creation of 6 new variables which consisted of the aggregate Ratings of each subject, for the artist attributes that were presented in a positive vs. negative way.

Specifically, I created the variables:

Successful Image Songs = Daron Malakian Song + Viza Song + Buckethead Song + Harry Styles Song

Unsuccessful Image Songs = The Hard Song + Chnages Song + Caveboy Song + Narek Hagopian Song

Positive Presentation Authorship = Daron Malakian Authorship + The Hard Authorship + Buckethead Authorship + Narek Hagopian Authorship

Negative Presentation Authorship = Viza Authorship + Chnages Authorship + Caveboy Authorship + Harry Styles Authorship

Positive Presentation Authenticity = Daron Malakian Authenticity + Viza Authenticity + The Hard Authenticity + Caveboy Authenticity

Negative Presentation Authenticity = Chnages Authenticity + Buckethead Authenticity + Narek Hagopian Authenticity + Harry Styles Authenticity

With this procedure there was no loss of data even though I did not use the original variables, other than the distribution of the Ratings inside each subject's questionnaire, which was not useful data for the aims of this study.

Because the sample was not particularly big and the data was not normally distributed, I chose the non-parametric Wilcoxon Signed Rank (Wilcoxon, 1945) test over the Related Samples T-test, in order to see if the distributions of Authorship, Authenticity and Song Ratings differ significantly under positive vs. negative presented artist images. The results are presented in Table 2.

Table 2. Differences in the distribution of aggregate Ratings.

Wilcoxon Signed Ranks Test			
	Succ. Image Songs – Unsucc. Image Songs	**Pos. Pres. Authorship – Neg. Pres. Authorship**	**Pos. Pres. Authenticity – Neg. Pres. Authenticity**
Z	-1,767[b]	-4,795[c]	-3,842[c]
Asymp. Sig. (2-tailed)	,077	,000	,000
b. Based on positive ranks. c. Base on negative ranks.			

As shown, the song ratings were not affected significantly for ,05 sig. by whether the artists had a “professional” or “good” vs. a “hobbyist” or “bad” image. Also, Authorship and Authenticity ratings did reflect an understanding of these qualities by the subjects who read the biographies, at a ,01 sig. level. (Note: The Z score is based on negative ranks for c, which means that the artists who were presented as having higher Authorship and Authenticity received higher ratings than their counterparts.)

Conclusively, the results seem to replicate what has been suggested by the literature about the positive effect of such presented image properties. On the other hand, the presentation of the artists as successful or not so successful did not seem to impact the judgments of the songs. This could seem to contrast a common conception that presenting achievements and maintaining a “polished” image can benefit the judgments of an artist's output. If such an effect exists, this study suggests that it is to be found in a confounding variable and not at the presented image per se. For example, artists who have achievements and a professionally maintained image also have financial access to better music equipment and award-winning music producers, who can help the artists achieve a sound of a significantly higher standard than of the artists with fewer available resources (this is the reason why I emulated live recordings instead of presenting album versions). In both cases, it seems that the aesthetic judgments intend to follow a kind of artistry evaluation over factors of influence that are unrelated to the procedure of creation of the music (e.g. sales and industry awards). This is emphasised by the fact that when artists were presented as authentic or authors, they had a better rating. This result can suggest that subjects choose how they are influenced by opinions: That they are not affected by prizes, achievements and “polished” images, but they are influenced by the artist's (perceived) skills and honesty.

Acknowledgements. I would like to thank Dr Renee Timmers for her remarks.

References

Ahonen, L. (2008). *Constructing Authorship in Popular Music*. Saarbrücken, Germany: VDM Verlag Dr. Mueller e.K.

Brackett, D. (2000). *Interpreting Popular Music: With a new preface by the author*. Oakland, CA: University of California Press.

Butler, M. (2003). *Taking it seriously: intertextuality and authenticity in two covers by the Pet Shop Boys. Popular Music, 22(1)*, 1-19.

Christenson, P. G., & Peterson, J. B. (1988). Genre and gender in the structure of music preferences. *Communication Research, 15(3)*, 282-301.

Conover, W. J. (1999). *Practical nonparametric statistics*. New York, NY: John Wiley & Sons, Inc.

Dyer, R. (1986). *Stars*. London, UK: BFI Publishing.

Goodwin, A. (1992). *Dancing in the distraction factory: Music television and popular culture*. Minneapolis, MN: University of Minnesota Press.

Hufthammer, K. O. (2014) *How robust is Pearson's correlation coefficient to violations of normality?* Retrieved from https://stats.stackexchange.com/questions/38273/how-robust-is-pearsons-correlation-coefficient-to-violations-of-normality

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30(1/2)*, 81–93.

Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association, 62(318)*, 399-402.

McLeod, K. (2003). Space oddities: aliens, futurism and meaning in popular music. *Popular Music, 22(3)*, 337-355.

Moore, A. (2002). Authenticity as authentication. *Popular music, 21(2)*, 209-223.

Mukaka, M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal: The Journal of Medical Association of Malawi, 24(3)*, 69–71.

Mäkelä, J. (2004*). John Lennon Imagined: Cultural History of a Rock Star*. Pieterlen, Switzerland: Peter Lang Inc., International Academic Publishers.

Negus, K. (1992). *Producing Pop: Culture and Conflict in the Popular Music Industry*. London, UK: Edward Arnold Publishers Ltd.

North, A. C., & Hargreaves, D. J. (1995). Subjective complexity, familiarity, and liking for popular music. *Psychomusicology: A Journal of Research in Music Cognition, 14(1-2)*, 77.

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: the structure and personality correlates of music preferences. Journal of personality and social psychology, 84(6), 1236.

Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: The role of music preferences in interpersonal perception. *Psychological science, 17(3)*, 236-242.

Reynolds, S. (1998). *Generation ecstasy: Into the world of techno and rave culture*. London, UK: Psychology Press Limited.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52(3-4)*, 591-611.

Spearman, C. (1904). The proof and measurement of association between two things. *The American journal of psychology, 15(1)*, 72–101.

Stathakopoulos, V. (2005). *Μέθοδοι Έρευνας Αγοράς [Market Research Methods]*. Piraeus, Greece: Stamoulis Press.

Tandt, C. D. (2004). From craft to corporate interfacing: rock musicianship in the age of music television and computer-programmed music. *Popular Music and Society, 27(2)*, 139-160.

Toynbee, J. (2000). *Making Popular Music. Musician, Creativity and Institutions*. London, UK: Edward Arnold Publishers Ltd.

Wilcox, R. (2011). *Introduction to robust estimation and hypothesis testing*. Waltham, MA: Academic Press.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics bulletin, 1(6)*, 80-83.

Young, J. O. (2001). Authenticity in Performance. In B. Gaut & D. Lopes (Eds.), *The Routledge Companion to Aesthetics* (pp. 383-394). Abingdon, UK: Routledge.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Effects of Cultural Background and Musical Preference on Affective Social Entrainment with Music

Jan Stupacher[1,2] and Guilherme Wood[1,3]

[1] *Institute of Psychology, University of Graz, Austria*
[2] *Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Denmark*
[3] *BioTechMed Graz, Austria*

stupacher@clin.au.dk

Abstract

When people come together to interact with music, they might not only synchronize their movements (*temporal* social entrainment), but also harmonize their emotional states (*affective* social entrainment). Here, we investigate how culturally typical and atypical music influences affective social entrainment. Sixty-one participants with a Western cultural background watched videos of two walking stick-figures. One figure represented the participants themselves and the other figure represented an unknown person. Two variables were manipulated: Musical pattern (Western vs. Indian, three pieces of instrumental music each) and synchrony (both figures in phase with the music vs. self-figure in phase and other-figure out of phase). Participants rated how connected they felt with the other person and how much they enjoyed the different types of music. Connectedness with the other person was higher with Western compared to Indian musical patterns. Additionally, a difference of connectedness with the other person between participants with higher vs. lower enjoyment of Indian musical patterns was found when the two stick-figures were walking in synchrony with Indian music, but no significant difference was found when the other-figure was walking out of phase. Similar effects were found for videos with Western musical patterns. In general, music with Western patterns was enjoyed more than music with Indian patterns. The findings suggest that high familiarity with and enjoyment of the music one interacts with lead to a higher predictability of movements and a stronger presence of social norms that would be fulfilled when moving in interpersonal synchrony, or violated when another person moves asynchronously. Additionally, the higher enjoyment of Western musical patterns might have led to a more positive mood, which, in turn, might have positively influenced social connectedness.

Introduction

When moving together in time with music, individuals not only synchronize their motor outputs, but may also syntonize their emotional experiences and judgements, such as trust and cohesion (Demos, Chaffin, Begosh, Daniels, & Marsh, 2012; Freeman, 2000; Koelsch, 2010). One reason for this unifying effect of music might be that movement synchronization as well as the act of listening to music appear to involve endorphins and the endogenous opioid system in general (Tarr, Launay, & Dunbar, 2014, 2016). The neuropeptide oxytocin seems to be especially important for the mediation of emotional musical experiences, interpersonal bonding, and entrainment (Chanda & Levitin, 2013; Gebauer et al., 2016). Another reason for the prosocial and emotional effect of music might be that it has played a major role in human evolution (Cross, 2001; Cross & Morley, 2008; Huron, 2001; Loersch & Arbuckle, 2013) and that the basics of social entrainment are learned early in life (Phillips-Silver & Keller, 2012). Moving together in synchrony promotes interpersonal affiliation and cooperation (Hove & Risen, 2009; Wiltermuth & Heath, 2009). Two recent studies have shown that these prosocial effects are strengthened when moving together with music (Stupacher, Maes, Witte, & Wood, 2017; Stupacher, Wood, & Witte, 2017).

Lucas, Clayton, and Leante (2011) showed that in an Afro-Brazilian Congado performance (a ritual with different types of musical ensembles), groups of the same community are more likely to entrain than groups of different communities. Additionally, a recent study by Vuoskoski, Clarke, and DeNora (2017) showed that for individuals with high trait empathy, listening to music from a specific culture increased the preference for facial pictures of individuals from that culture. These two studies provide some evidence that cultural background is an important factor in interpersonal interactions with music. However, detailed information about how the socio-cultural background of an individual affects social entrainment to music is still needed.

Here, we used a recently introduced experimental paradigm (Stupacher, Maes, et al., 2017) to investigate whether the socio-cultural background of an individual affects his/her social evaluations in interpersonal entrainment with culturally typical and atypical musical patterns.

Method

Participants

We report data of 61 participants (42 female, 19 male, mean age 22.0 years, $SD = 3.4$ years) with Western cultural backgrounds. Based on the Goldsmiths Musical Sophistication Index (Müllensiefen, Gingras, Stewart, & Musil, 2013), musical training of the participants varied between 7 and 43 (1^{st} and 93^{rd} percentile, respectively; $M = 21.97$, 36^{th} percentile, $SD = 9.64$).

Procedure

Based on the experimental design of Stupacher, Maes, et al. (2017), participants watched videos of two walking stick-figures with a duration of 20 seconds. One figure represented the participants themselves and the other figure represented an unknown person (Figure 1). Two variables were manipulated:

1) Synchrony: Both figures in phase with the music (*sync*, Figure 1A) vs. self-figure in phase and other-figure out of phase (*async*, Figure 1B). Each stride consisted of 21 frames. In the *async* videos, the steps of the other-figure were delayed by eight frames.

2) Musical pattern: The musical accompaniment of the videos was typical for either Western or Indian music styles.

For each musical pattern, three instrumental pieces with clear beats were selected. The tempo of all instrumental pieces was aligned to the stride length of 636 ms / 94.3 beats per minute (either originally or slightly adjusted). A metronome accompaniment served as a control condition.

Ratings

Participants rated the connectedness with the other-figure on an adapted Inclusion of Other in the Self Scale (IOS; Aron, Aron, & Smollan, 1992; Figure 1C) with values ranging from 0 to 100. At the end, participants rated how much they enjoyed each piece of music, also on a scale from 0 to 100.

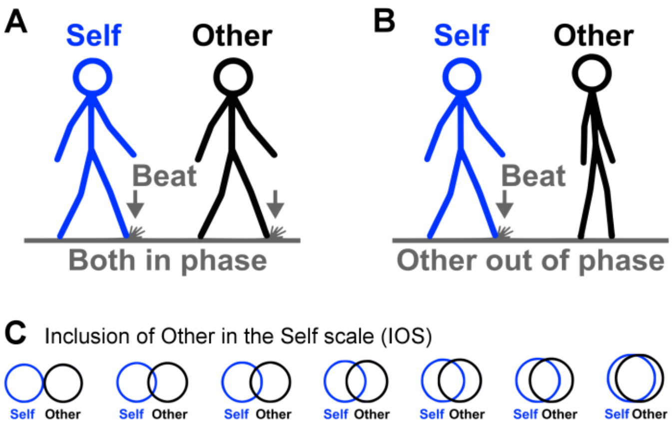


Figure 1. A) Both figures walking in phase. A dust cloud highlights the time when the foot hits the ground. B) The self-figure is walking in phase, while the other-figure is walking out of phase. C) Adapted Inclusion of Other in the Self Scale (IOS, Aron et al., 1992).

Results

A 2x2 ANOVA on IOS ratings revealed main effects of synchrony ($F(1,60) = 124.27$, $p < .001$; *sync*: $M = 60.35$, $SD = 26.03$; *async*: $M = 33.93$, $SD = 16.45$) and musical pattern ($F(1,60) = 24.28$, $p < .001$; *Indian*: $M = 44.22$, $SD = 19.97$; *Western*: $M = 50.06$, $SD = 20.53$), and no significant interaction.

To investigate the effect of musical preference, we divided the participant sample into two groups based on the median of enjoyment ratings of Western (Figure 2A) and Indian music (Figure 2B), respectively.

A 2x2 ANOVA on IOS ratings of videos with Western musical patterns with the within-subjects factor synchrony and the between-subjects factor music preference (higher vs. lower enjoyment of Western music) revealed a main effect of synchrony ($F(1,59) = 123.11$, $p < .001$; *sync*: $M = 63.30$, $SD = 26.67$; *async*: $M = 36.82$, $SD = 17.85$), a significant between-subjects effect ($F(1,59) = 5.50$, $p = .022$), and a significant interaction ($F(1,59) = 5.30$, $p = .025$; Figure 3A). Following up this interaction, two t-tests comparing participants with higher vs. lower enjoyment of Western music revealed a difference of IOS ratings when the two stick-figures were walking in synchrony ($t(59) = 2.68$, $p = .010$), but no significant difference when the other-figure was walking out of phase ($t(59) = 1.41$, $p = .164$).

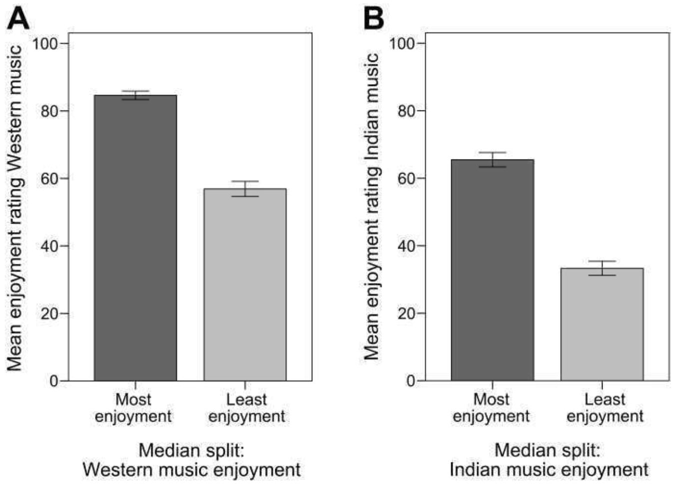


Figure 2. Depiction of the music enjoyment ratings after median splits based on A) enjoyment ratings of Western music, and B) enjoyment ratings of Indian music. In general, music with Western patterns (A) was enjoyed more than music with Indian patterns (B; $t(60) = 6.69$, $p < .001$).

Similar effects were found for videos with Indian musical patterns. A 2x2 ANOVA on IOS ratings of videos with Indian music with the within-subjects factor synchrony and the between-subjects factor music preference (higher vs. lower enjoyment of Indian music) revealed a main effect of synchrony ($F(1,59) = 124.92$, $p < .001$; *sync*: $M = 57.40$, $SD = 26.57$; *async*: $M = 31.04$, $SD = 16.52$), a significant between-subjects effect ($F(1,59) = 6.00$, $p = .017$), and a significant interaction ($F(1,59) = 4.27$, $p = .043$; Figure 3B). Following up this interaction, two t-tests comparing participants with higher vs. lower enjoyment of Indian music revealed a difference of IOS ratings when the two stick-figures were walking in synchrony ($t(59) = 2.61$, $p = .012$), but no significant difference when the other-figure was walking out of phase ($t(59) = 1.71$, $p = .093$).

Discussion

Effects of Synchrony

The main effects of synchrony underscore previously found prosocial effects of interpersonal movement synchronization (e.g., Hove & Risen, 2009; Wiltermuth & Heath, 2009). The findings also validate the experimental paradigm introduced by Stupacher, Maes, et al. (2017), by showing that active movement is not necessary for evaluating social components of interpersonal synchronization.

Effects of Musical Patterns and Musical Preference

By comparing interpersonal synchronization with music and a metronome, Stupacher, Maes et al. (2017) concluded that "especially with music, unsynchronized movements of another person contradict one's own affective state and violate social expectations." Here, we expand this suggestion by showing that the connectedness to another person is higher when moving to culturally typical music, as compared to less typical music. Additionally, we found that the enjoyment of the music one interacts with plays an important role in social evaluations. Findings suggest that individuals who enjoy the accompanying music more show larger differences in their connectedness to another person depending on whether this person moves synchronously or asynchronously (Figure 3).

Referring to Stupacher, Maes et al. (2017), we provide two potentially interdependent explanations for these findings: One based on social expectations related to the familiarity with the stimuli and the other related to music-induced mood.

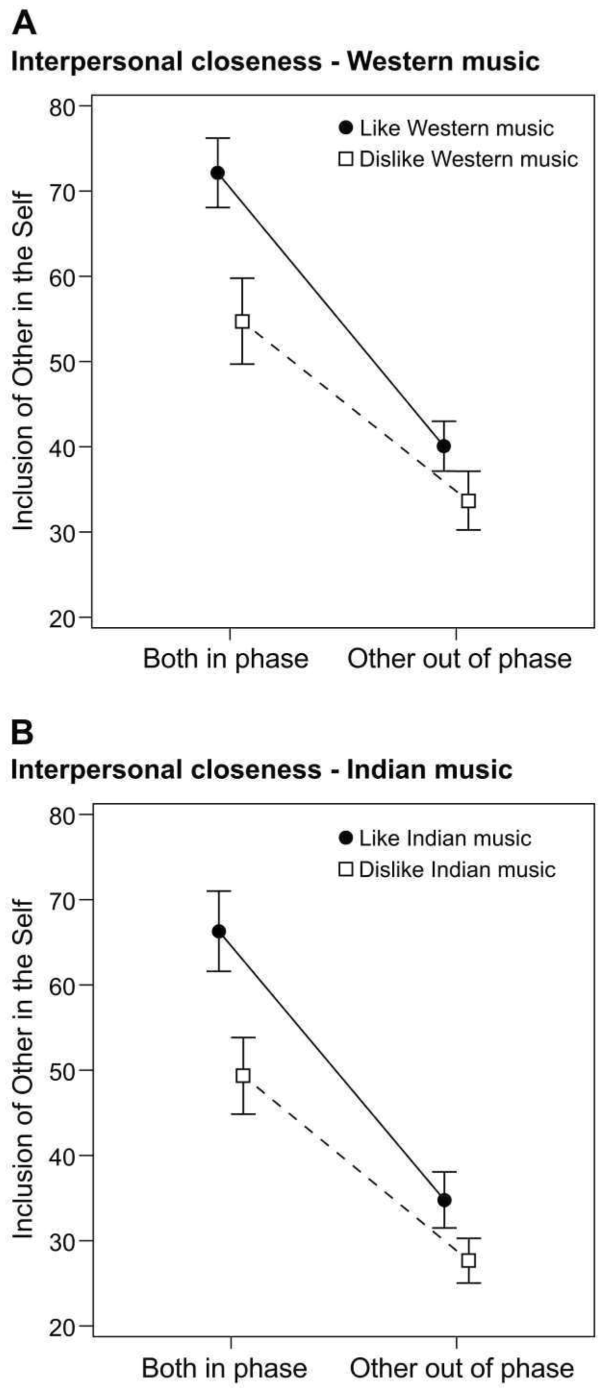


Figure 3. Means (+/- 1 *SE*) of participants' ratings of interpersonal closeness between the self- and other-figure. A) Ratings of videos accompanied by Western musical patterns for participants who liked vs. disliked Western musical patterns (based on a median split). B) Ratings of videos accompanied by Indian musical patterns for participants who liked vs. disliked Indian musical patterns (based on a median split).

Individuals from the same community are more likely to entrain (Lucas et al., 2011) and listening to music from a specific culture can promote affiliation with individuals from that culture (Vuoskoski et al., 2017). Therefore, our findings suggest that higher familiarity with musical patterns in interpersonal synchronization might have resulted in higher ratings of interpersonal closeness when moving to culturally typical compared to untypical music. High familiarity with and enjoyment of the music one interacts with might lead to a higher predictability of movements and a stronger presence of social norms that would be fulfilled when moving in interpersonal synchrony, or violated when the other person moves out of phase, as suggested by the interactions between synchrony and music preference depicted in Figure 3.

When moving together in time with music, individuals not only synchronize their movements, but also their moods and emotions (Demos et al., 2012; Freeman, 2000; Koelsch, 2010). Listening to "uplifting" compared to "annoying" music during physical exercise increases the likelihood of prosocial behavior (North, Tarrant, & Hargreaves, 2004). Here, participants rated Western musical patterns as more enjoyable than Indian musical patterns. This might have led to a more positive mood when listening to Western musical patterns, which, in turn, might have positively influenced the ratings of social connectedness.

Limitations and Future Directions

The current experimental design does not enable a clear separation of the effects of cultural familiarity with music and enjoyment of music. Our findings suggest that the effects are interdependent. However, future research on social entrainment should try to disentangle the influences of (cultural) familiarity and enjoyment of music in more detail.

The highly controlled nature of task and stimuli in the current experiment simplifies social entrainment with music and might reduce the external validity of the results. Thus, our findings should be validated in more ecologically valid settings, such as real-time face-to-face interactions with music related to different cultures or subcultures.

References

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*, 596–612.

Chanda, M. L., & Levitin, D. J. (2013). The neurochemistry of music. *Trends in Cognitive Sciences*, *17*, 179–191.

Cross, I. (2001). Music, cognition, culture, and evolution. *Annals of the New York Academy of Sciences*, *930*, 28–42.

Cross, I., & Morley, I. (2008). The evolution of music: Theories, definitions and the nature of the evidence. In S. Malloch & C. Trevarthen (Eds.), *Communicative Musicality* (pp. 61–82). Oxford, UK: Oxford University Press.

Demos, A. P., Chaffin, R., Begosh, K. T., Daniels, J. R., & Marsh, K. L. (2012). Rocking to the beat: Effects of music and partner's movements on spontaneous interpersonal coordination. *Journal of Experimental Psychology. General*, *141*, 49–53.

Freeman, W. J. (2000). A neurobiological role of music in social bonding. In N. Wallin, B. Merkur, & S. Brown (Eds.), *The Origins of Music* (pp. 411–424). Cambridge, MA: MIT Press.

Gebauer, L., Witek, M. A. G., Hansen, N. C., Thomas, J., Konvalinka, I., & Vuust, P. (2016). Oxytocin improves synchronisation in leader-follower interaction. *Scientific Reports*, *6*, 38416.

Hove, M. J., & Risen, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, *27*, 949–960.

Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Science*, *930*, 43–61.

Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences*, *14*, 131–137.

Loersch, C., & Arbuckle, N. L. (2013). Unraveling the mystery of music: Music as an evolved group process. *Journal of Personality and Social Psychology*, *105*, 777–798.

Lucas, G., Clayton, M., & Leante, L. (2011). Inter-group entrainment in Afro-Brazilian Congado ritual. *Empirical Musicology Review*, *6*, 75–102.

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. J. (2013). *Goldsmiths Musical Sophistication Index (Gold-MSI) v1.0: Technical report and documentation. Revision 0.3*. London: Goldsmiths, University of London.

North, A. C., Tarrant, M., & Hargreaves, D. J. (2004). The effects of music on helping behavior: A field study. *Environment and Behavior*, *36*, 266–275.

Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience*, *6*, 26.

Stupacher, J., Maes, P.-J., Witte, M., & Wood, G. (2017). Music strengthens prosocial effects of interpersonal synchronization – If you move in time with the beat. *Journal of Experimental Social Psychology*, *72*, 39–44.

Stupacher, J., Wood, G., & Witte, M. (2017). Synchrony and sympathy: Social entrainment with music compared to a metronome. *Psychomusicology: Music, Mind, and Brain*, *27*, 158–166.

Tarr, B., Launay, J., & Dunbar, R. I. M. (2014). Music and social bonding: "Self-other" merging and neurohormonal mechanisms. *Frontiers in Psychology*, *5*, 1096.

Tarr, B., Launay, J., & Dunbar, R. I. M. (2016). Silent disco: Dancing in synchrony leads to elevated pain thresholds and social closeness. *Evolution and Human Behavior*, *37*, 343–349.

Vuoskoski, J. K., Clarke, E. F., & DeNora, T. (2017). Music listening evokes implicit affiliation. *Psychology of Music*, *45*, 584 –599.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, *20*, 1–5.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Functional Nature of Theoretically Decorative Common-Tone Chords

Konrad Swierczek,[1] Joanna Spyra,[2] and Matthew Woolhouse[1]

[1] *School of the Arts, McMaster University, Canada*
[2] *Psychology, Neuroscience, and Behaviour, McMaster University, Canada*

[1]swierckj@mcmaster.ca, [2]spyraj@mcmaster.ca, [3]woolhouse@mcmaster.ca

Abstract

A goal of music theory is to classify patterns in order to better understand the structure of music. While these principles quite often coincide with perception, as is arguably the case with mixture and applied dominants, complex cases require more nuanced explanations. The common-tone diminished seventh (CTº7) is such an example, being typically interpreted as non-functional decorative harmony, despite its use within otherwise well-formed musical phrases. More generally, the ambiguous nature of diminished seventh chords presents an obstacle for theoretical classification. Through the principal of "tritone substitution" of pitches belonging to the major subdominant harmony, we identify the CTº7 as a predominant chord, relative to the key region of the common tone. The same principal, when applied the minor subdominant, yields a related French sixth chord. In the same way that the subdominant can approach the dominant or return to the tonic, this theory relates CTº7 to its dominant enharmonic equivalents. In order to test this theoretical notion, this study relates the CTº7 to its parent subdominant harmony through a psychological experiment, testing the validity of the theory against perception. The three unique possible (enharmonically equivalent) diminished seventh chords and their French sixth counterparts were related to simple diatonic tonic-subdominant and tonic-dominant structures using a two-alternative forced choice paradigm. Participants selected one of two diminished seventh or French sixth chord progressions, which they judged to be most similar to the presented stimulus. Data analysis indicated that participants found CTº7 chords more similar to their subdominant theoretical parent harmonies than other diminished sevenths, but did not find any French sixth more similar to a given chord. The results indicate that tritone substitution with respect to CTº7 and subdominant chords may indeed be perceivable , and a powerful tool for relating chromatic regions which otherwise might seem tonally non-functional. .

Background

The study and pedagogy of western tonal harmony was for many years concerned mainly with the classification and labeling of sonorities and their resolutions as to create a framework for identification, and a vocabulary for the student and theorist alike. Relatively arbitrary names were assigned to chromatic harmonies, and both regular and irregular behaviors were documented based on the masterworks of the 18th and 19th century composers. However, in the 20th century,theorists such as Arnold Schoenberg became more concerned with questions of hierarchy, attraction, and derivation (e.g. monotonality; Schoenberg, 1969). , Despite this, many contemporary harmony texts still use traditional explanations of complex chromatic sonorities, which arguably provide an unsatisfactory framework. The advent of the field of music cognition has provided an opportunity to rely not on relatively arbitrary labels and classifications, but no the observation of how high level structural features of music are perceived. The issue of non-dominant diminished sevenths chords (also known as the common tone diminished seventh, CTº7, or the raised supertonic/submediant in older literature (see Aldwell & Schachter, 2011; Piston, 1987) is an outmoded theory of particular interest. Usage of the CTº7 has changed greatly, as composers found new ways to integrate it as a decoration, prolongation, and preparation for modulation. In its earliest form, it evolved contrapuntally as a collection of 3 neighbour notes: scale degrees 3-#2-3, 5-#4-5, and the diatonic 5-6-5. With the addition of a common tone (the root of the succeeding chord), a diminished seventh chord is formed (Fig 1). The common tone is traditionally notated as the seventh of the chord, thus making the CTº7 an apparent seventh chord, not a true seventh chord (discrepancies in notation will be addressed below). The CTº7 is rarely used to elaborate a minor triad in the literature of the common practice era due to one chromatic leading tone (scale degree #2) becoming a common tone with b3. Beginning in the Classical period, the likes of Mozart exploited the CTº7 as a brief decoration of the tonic, or any major chord, often shorter than the harmonic rhythm of the rest of the phrase (Fig 1). Composers of the Romantic period such as Schubert, Tchaikovsky, and Chopin, began to use the CTº7 as a largely independent harmony. For example, in Tchaikovsky's Symphony No. 5 (Fig 1), CTº7 prepares the dominant, then elaborates the dominant, finally resolving to the tonic. It was during the 19th century that composers began to exploit the intense attraction of the chromatic leading tones for less conventional purposes such as modulation. For instance, the leading-tone seventh chord of any key can be re-interpreted as CTº7 to the key area of the leading tone, supertonic, subdominant, or lowered submediant. Songwriters in the 20th century often used CTº7 in a chain of predominant chords. For example, in "Witchcraft" by Cy Coleman, CTº7 is approached by chord I, but resolves to chord ii (Fig 1). It is important to note here that CTº7 is often spelled in jazz lead sheets as a diminished seventh built on the scale degree b3. In jazz improvisation, this practice is so common that it has become a common substitution or reharmonization technique (Felts, 2002). This broad development of the CTº7 is the the rationale for a new functional explanation to describe chords from this family: while in its earliest form it may have been a simple elaboration of a triad, modern usage in the literature is much more independent and flexible. It is therefore hypothesized that non-dominant diminished seventh chords do indeed have function; specifically, predominant function. The CTº7 can be viewed as a substitution of the subdominant, preparing the tonic, dominant, cadential 6/4, or preparing other predominant harmonies.

Figure 1. Examples of conventional usage of common-tone diminished sevenths.

The process of tritone substitution is often used in jazz, where any dominant seventh chord can be substituted for another dominant seventh whose root is a tritone higher (Felts, 2002). The dominant seventh is unique due to the presence of one tritone: by performing tritone substitution on all its pitches, the resultant chord will have two common tones with its parent chord. Similarly, a diminished seventh with all its pitches put through tritone substitution will be identical to its parent chord. Tritone substituted dominant seventh chords, or augmented sixth chords as they are referred to in the classical literature, rely on the tritone relationship from their parent, common tones, and semitone voice leading relationships in the circle of fifths to create a coherent sense of functional harmony. The combination of these elements enables a chord with a great deal of covert dissonance to operate effectively in a tonal context (Komar, 1992). Other chords without a tritone, such a major and minor triad or sevenths, will invert symmetrically, but produce no common tones with the parent chord. While these chords might theoretically have a structural relationship with their parent chords, they are less commonly used than their dominant seventh counterpoints. In classical music theory, a common chromatic alteration is the German sixth, which is enharmonically equivalent to a dominant seventh. These chords can be interpreted as a tritone substitution of the applied harmony V7/V (similarly, the Italian sixth is an incomplete version of the same chord). However, tritone substitution need not be limited to an entire chord. By taking two pitches of a chord (specifically the third and fifth), performing tritone substitution while keeping the original notes in the chord, we can derive diminished seventh chords from a minor triad and French sixth chords from a major triad (Fig 2). Consider the French sixth of C minor resolving to the dominant. Using the parent triad F minor, the subdominant, we perform tritone substitution on the third and fifth of the chord, include the original pitches, and omit the root. The resultant chord is the French sixth of the key. The major dominant of the key put through the same process yields the leading tone seventh chord. Finally, the same tritone substitution can be performed on the major subdominant to create the common tone diminished seventh of the tonic. The study conducted here uses a psychological discrimination task to acquire evidence based on perception to support the theory of expanded tritone substitution. Music is filled with apparent patterns, but many of them are irrelevant if they are not represented in cognition. The tools of music cognition can help enable music theory to better describe phenomena such as tritone substitution.`

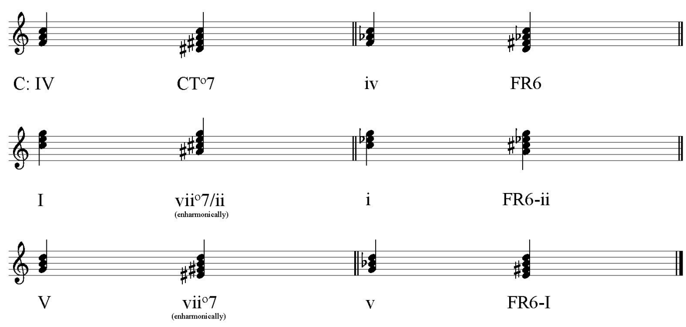


Figure 2. Parent harmonies and their respective tritone substitutions using the procedure described above. The first system shows the subdominant harmony of C major and minor and their corresponding diminished seventh and French sixth. The second system shows the same chords for the tonic of C major and minor, and the third system shows the dominant of C major and minor.

Methods

480 trials were produced in MuseScore 2, using a grand piano timbre. The trials consisted of all possible combinations of the three acoustically unique diminished seventh chords and their related French sixth chords. In addition, all of these combinations were coupled with both a tonic-dominant and tonic-subdominant focus stimulus. This design controls for confounds of psychoacoustic similarity, as the participants are not only presented with an opportunity to match the tritone substituted chord with its parent, but also to reject its similarity to a different chord. Each trial consisted of three chords: in both the focuses and the standards, the first and last chord were tonic triads. In focus stimuli, the middle chord was either a subdominant or dominant triad related to the tonic. In standard stimuli, the middle chord was a tritone substituted diminished seventh or French sixth chord. Each unique trial was presented twice, with standards presented in reverse to control for the effect of order. In addition, each trial was randomly cycled through four key regions related by minor thirds, controlling for the effect of acoustic familiarity. All progressions were written in closed position, following the

conventions of common practice voice-leading (Aldwell & Schachter, 2011). 29 undergraduate students (22 female) were recruited through the McMaster University Department of Psychology Research Participation System. 11 participants reported having knowledge of music theory, 22 reported having at least 1-5 years of musical training, and only two participants reported having perfect pitch. All participants reported listening to music for at least 5 hours a week, with genres varying between participants. Participants first provided informed consent and completed a questionnaire on general demographics and musical experience. Following this, 120 stimuli were presented binaurally at 70BPM through Sennheiser HD 280 headphones. A two-alternative forced-choice paradigm similar to an ABX design was used (Hautus & Meng, 2002). Rather than discriminating the focus stimulus to one of two standard stimuli, a focus is presented followed by two standards (Fig 3). Participants were asked to select the standard which sounds most similar to the focus. Stimulus presentation and response recording was controlled by a Python based program on a Windows computer. The entire experiment lasted approximately 30 minutes.

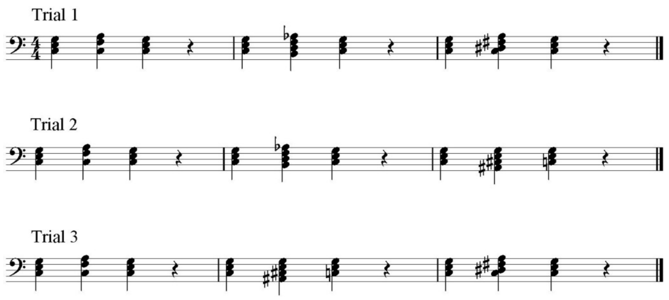


Figure 3. Stimuli examples. Each system represents one trial. The first measure of each system represents the focus stimulus. The following two measures of each system represent standard stimuli: in this case, every possible combination of diminished seventh chord.

Results

Actual responses for each trial were grouped together to create a proportion of responses for each chord and its variables (See Appendix 1 for actual values). For the purposes of data analysis, all trials with a subdominant focus were separated from all trials with a dominant focus as the hypothesis would expect opposite results. Two within-subject factorial ANOVAs (one for dominant focus, one for subdominant focus) were performed to determine the effect of parent harmony (tonic, subdominant, dominant) and chord quality (diminished seventh or French sixth). Four post-hoc Tukey's honest significant difference analyses were then performed to compare individual differences between parent harmonies for a given chord quality. Analysis of the dominant focus condition revealed a highly significant main effect of both parent harmony ($F_{2,29}$ = 41.675, p = 0.0003) and chord quality ($F_{1,29}$ = 93.765, p < 0.0005), as well as a significant interaction between the factors ($F_{2,29}$ = 18.103, p = 0.0029). Comparison of means between parent harmonies revealed the dominant diminished seventh was deemed significantly more similar to the dominant focus than the subdominant diminished seventh and tonic diminished seventh (p = 0.046, p = 0.00095). The subdominant diminished seventh was deemed significantly more similar to the dominant focus than the tonic diminished seventh (p = 0.0022). No significant difference was found in judgement of the dominant French sixth and either the subdominant or tonic French sixth (p = 0.1963, p = 0.6435). However, the subdominant French sixth was deemed significantly more similar to the dominant focus than the tonic French sixth (p = 0.0914). Similarly, analysis of the subdominant focus condition revealed a highly significant effect of both parent harmony ($F_{2,29}$ = 26.460, p = 0.0011) and chord quality ($F_{1,29}$ = 17.081, p = 0.0061), as well as a significant interaction between the factors ($F_{2,29}$ = 30.792, p = 0.0007). No significant difference was found between subdominant diminished sevenths and dominant diminished sevenths in the subdominant focus condition (p = 0.2013). However, both the dominant and subdominant diminished sevenths were deemed significantly more similar to the subdominant focus than the tonic. (p = 0.0146, p = 0.0062). No significant difference was found between any of the French sixth chords in the subdominant focus condition (p = 0.2682, p = 0.9945, p = 0.2935) (fig.4).

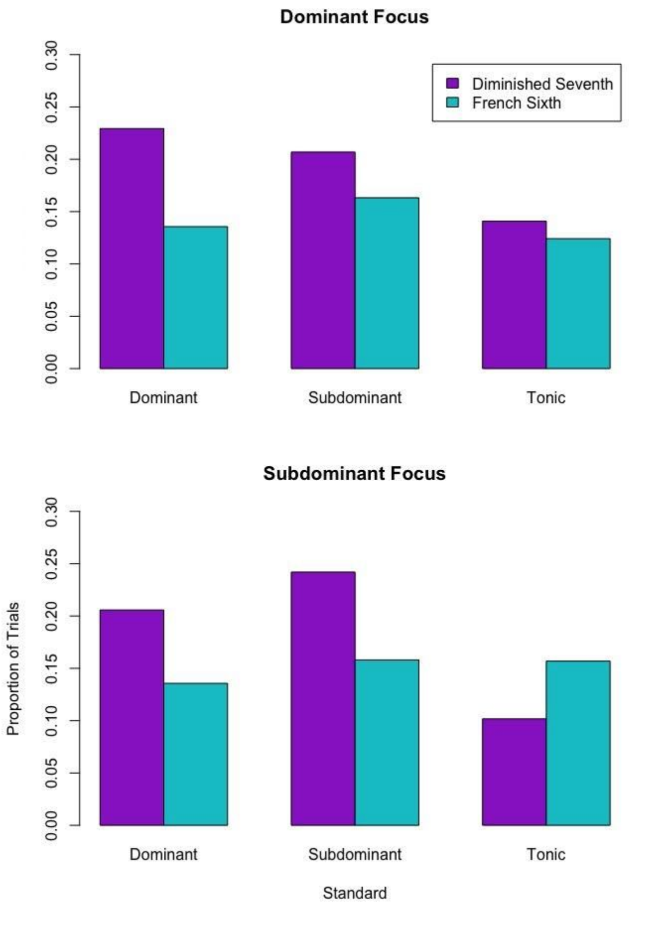


Figure 4. Visual representation of the proportion of trials each standard stimulus was deemed more similar to dominant or subdominant focus conditions. Total amount of trials analyzed was 3,480.

Discussion

The results presented here are a promising indicator that these expanded forms of tritone substitution are perceived structurally and preferred over similar chords without this relationship. Tukey's HSD analysis revealed that all but one diminished seventh chord was considering significantly more similar to its parent chord than any other diminished seventh. This is compelling evidence that tritone substitution indeed has a perceptual basis. The lack of a significant similarity between dominant and subdominant diminished sevenths to the subdominant focus indicates a design limitation; the instructions given to the participants are the root of this problem. While harmonic progression and structure is known to be an important feature of music, it can be confusing to ask people without a knowledge of harmony to discriminate musical grammar. As such, the lack of a significant judgement of similarity of subdominant diminished sevenths to a subdominant focus indicates that participants may have responded based on psychoacoustic similarity. Simply put, the dominant diminished seventh has more pitches in common with the subdominant than the subdominant diminished seventh. However, the inverse relationship showed that participants chose the dominant diminished seventh more than the subdominant for a dominant focus. This further reinforces that these substitutions depend on the tritone relationship, voice leading, and psychoacoustic similarity equally to maximize congruence.

Only one French sixth chord produced a significant judgement of similarity. The common tone French sixth, as used here, is an unusual harmony as it features a large degree of sensory dissonance, and a lesser amount of chromatic leading tones in comparison to its common tone diminished seventh and common tone German sixth counterparts. Additionally, it is seldom used in popular music, and handled with care in classical and jazz music. In jazz, the French sixth is most often used as a dominant substitution by composers such as Antonio Carlos Jobim, where it is referred to as a "dominant seventh flat five". In the classical repertoire, it is most often used by composers of the late 19th century such as Richard Wagner, Richard Strauss, and Arnold Schoenberg for its peculiar sound and powerful dissonance. Here, we found that participants only found the subdominant French sixth more similar to the dominant focus than the tonic French sixth. This is likely due to psychoacoustic similarity to the key region. The chromatic leading tone scale degree #1 in the tonic French sixth is more distantly related to the tonic than the chromatic leading tone #4 in the subdominant French sixth, which can tonicize the dominant region: the similarity to the applied harmony V7/V brings to mind more common progression than the tonic French sixths similarity to V7/ii. Furthermore, tonic tritone substitutions surrounding tonic harmonies may cause perceptual confusion. While the theorized tritone substitution may be recognized by the brain, the chord still represents a departure from the tonic. As such, it may be that the cognitive process is conflicted by the presence of a tonic derived, dissonant harmony alongside a regular tonic harmony.

This early exploration of expanded tritone substitution does not fully test the implied function of these chords. In order to effectively imply function to these chords, they must be tested in the larger context of a phrase unit. For instance, the cadential six-four is a harmony with apparent tonic function due to its component elements, but is in practice an elaboration of the dominant. Similarly, common tone harmonies always precede the chord with which they have common tones with: they are "prelongations" of the chord. Assigning predominant function to the common tone diminished seventh is not unreasonable. Within the framework of tritone substitution, I-CT°7-I resembles the plagal progression I-IV-I. It is also enharmonically equivalent to vii°7/V. A similar experiment involving the rating of progressions on likert-type scales would help elucidate possible functions, while also providing more information as to the goodness of fit of these chords in larger contexts. The original experimental design included a factor of context mode (whether the context mode of the focus and tonic harmonies are major or minor). However, incorporation of this factor would have required a much longer experiment for each subject. Future work might analyze this factor as a between-subject variable to avoid time constraints. A next step might also include other tritone substitutions. The chords considered here followed a specific procedure of tritone substitution, but many additional options also exist. Three additional unique French sixths exist, each with a multitude of derivational contexts through tritone substitution. Harmonies with additional extensions are also plausible candidates for experimentation. These less commonly used (or previously unknown) chords might provide composers and improvisers with a new vocabulary.

The existence of a functional relationship from the tritone may traditionally be considered unusual: the tritone is often associated with dissonance and harshness (at one point in history even being associated with the devil). However, the tritone is a critical feature in diatonic pitch perception. It is the only interval with an octave cycle of two, and its presence defines a key region due to its uniqueness within a scale (Woolhouse, 2010). A final consideration of the chords discussed here is spelling. Conventionally, common tone diminished seventh chords have been notated with leading tones to chord tones of the resolution harmony. Augmented sixth chords are notated with accidentals appropriate to their resolution direction. However, in jazz, a tritone substituted dominant seventh is typically notated as a diminished fifth above the parent root. The same method applied to a French sixth yields a dominant seventh with a flattened fifth. The identity of a diminished seventh chord put through this procedure remains unchanged, but adopts a different root. The desired spelling of these chords is ultimately not an issue of perception, but must be considered if this theory is to be applied pedagogically. Indeed, outside of the generation of new harmonic possibilities, the next objective of this investigation is to provide a theoretical framework for all tritone substituted harmonies, presently referred to as "other" chromatic chords or otherwise by confusing historical labels. Regional labels can be particularly confusing for the student, as they have no connection to the concept of a chord being innately "Neopolitan" or "German".

An expanded theory of tritone substitution informed by cognition and perception provides a theoretical framework for students, a generative vocabulary for composers and improvisers, and insight into the organization of tonal structure in the brain. This investigation is a preliminary effort into building an encompassing theory which would account for the generative possibilities of tritone substitution in western tonal harmonic music.

References

Aldwell, E., & Schachter, C. (2011). *Harmony & Voice Leading*. Schirmer, Cengage Learning.

Felts, R. (2002). Reharmonization Techniques. Berklee Press.

Hautus, M.J., & Meng, X. (2002). Decision strategies in the ABX (matcing-to-sample) psychophysical task. Perception & Psychophysics, 64(1).

Huron, D. (2001). Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principals. Music Perception, 19(1).

Komar, A. (1992). *Linear Derived Harmony*. Ovenbird Press.

Piston, W. (1978). *Harmony*. W.W. Norton & Company.

Schoenberg, A. (1969). *Structural Functions of Harmony*. W.W. Norton & Company.

Woolhouse, M., & Cross, I. (2010). Using interval cycles to model Krumhansl's tonal hierarchies. Music Theory Spectrum, 32(1).

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Positioning Learning Support System for Violin Beginners using Multimodal Information

Yoshinari Takegawa,[1] Rui Saito,[2] and Keiji Hirata [3]

[1,2,3] *Future University Hakodate, Japan*

[1]yoshi@fun.ac.jp, [2]g2116020@fun.ac.jp, [3]keiji@fun.ac.jp

Abstract

When playing the violin, it is difficult to master accurate positioning, since any slight deviation of string-pressing position causes the pitch to change. In order to solve this problem, in the conventional learning method, understanding of correct pressing position is supported by visual auxiliary information, such as stickers stuck on the fingerboard to mark frets. Alternatively, a tuner may be used. On the other hand, to perform accurate positioning it is necessary not only to memorize correct pressing position, but also to have an excellent ear for music, or 'sense of sound'. However, conventional learning methods and previous research do not support development of sense of sound. In our research, we aim to solve this problem by designing and implementing a positioning learning support system that presents multimodal information (visual and auditory auxiliary information) and has a function to promote understanding of correct pitch position and a function to train learners' sense of sound. This system aims to make learning more efficient than conventional methods. Evaluation experiments were conducted on 14 subjects who were not experienced at the violin. There were three sets of three practice procedures: using the system's learning support to practice a musical scale, playing an assigned song with no learning assistance, and reviewing practice. The learning assistance used by the control group was visual aid only, and that used by the experiment group was multimodal information. When calculating the average score of each group based on the pitch difference between the sounds produced by the subjects and the correct sounds of the assigned piece of music, the average score for the song in the first set was lower in the experiment group than in the control group, while in the third set the average score was higher in the experimental group than in the control group. Also, there was a significant difference between groups regarding the increase in the average score from the first performance of the assigned piece to the third performance.

Introduction

Correct positioning is essential in violin playing. Positioning means pressing strings onto the fingerboard to produce the pitch indicated by the notation on the score at the current playing position, as shown in Figure 1. As the violin fingerboard does not have frets, and mistaking pressing position by even a few millimeters causes a large deviation in pitch, acquiring correct positioning is extremely difficult. Conventional positioning learning methods include indicating pressing position on the fingerboard with stickers (hereafter: fingerboard stickers) and visualizing the pitch difference between played sound (the sound produced when the player plucks the string) and correct sound (the sound corresponding to the notation on the score at the current playing position). Fingerboard stickers enable visual comprehension of the position to be pressed, but learners do not become able to notice slight deviation of pitch.

Figure 1 Positioning

Learners can find out positioning while looking at the tuner indicators which are a real-time visual representation of pitch difference, but this operation is complicated.

As mentioned previously, because positioning requires a high level of accuracy, it is necessary not simply to memorize correct pressing positions, but to have the ability to notice and correct positioning errors. To develop such an ability, we focused on sense of sound. In our research, sense of sound is defined as the ability to determine how played sound compares to correct sound (too high, correct, too low) by listening alone. When sense of sound improves, not only can the player instantaneously notice and correct positioning errors, he/she also becomes able to correct positioning dynamically when the violin is out of tune.

Fingerboard stickers and tuners are forms of visual assistance that support memorization of pressing position, yet they do not support development of sense of sound. Therefore, the target of our research is the construction of a positioning learning support system with multimodal information presentation, focused on memorization of pressing position through visual assistance, and development of sense of sound through auditory assistance.

Related Research

Research relating to musical instrument learning support

One work of research on learning support for violin positioning is a system (Jian-Heng et al., 2012) which visualizes and feeds back the pitch difference between played sound and correct sound in real-time, based on sound data

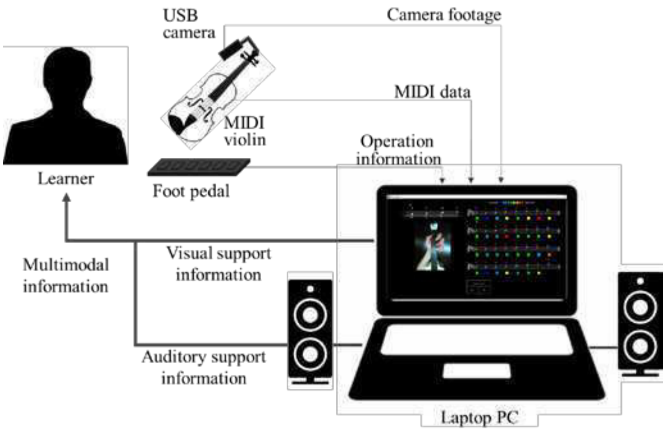


Figure 2 System Structure

obtained while playing. This system includes functions to record playing errors and support bowing. A system (Kuamki et al., 2017) that displays pressing position on the fingerboard utilizes the psychology that 'when lied to, a person ceases to trust the liar', and presents not only correct pressing position but also vague or false position information, to encourage learners not to rely on the system and thus promote their graduation from system support. Furthermore, this system has a function to feedback and record pitch difference between played sound and correct sound, based on MIDI data obtained while playing. In addition, there are systems that present key(s) to be pressed, fingering and example videos etc. (Takegawa et al., 2011).

Research utilizing multimodal information

Multimodal refers to the combining of several types of modality information when transmitting certain information. Modality means a cognitive method or style used to convey information (Nagaya et al., 1996). As modality information that can be used multimodally it is common to use information that can be obtained through the five senses, e.g. sight or sound. In recent years, systems constructed utilizing multimodal information have been developed in various fields. For example, a multimodal interaction system with an emphasis on dialogue, or a multimodal interface focusing on human and computer interfaces. These systems use various forms of modality information, such as graphics which are the basis of direct operation, and audio or text that presents words, as both simultaneous and mutual supplementation. In this way, users can interact by exercising the modality of words and actions (Oviatt et al., 2000).

In this way, research comprising multimodal technology is gradually increasing, however, organizations related to musical instrument learning support are still aiming to increase the efficiency of learning by providing visual assistance for cases where it is difficult for learners to judge by ear. In contrast, there are few cases of supporting learning through provision of multimodal support information, as in our research.

Design and Implementation

The positioning learning support system proposed in this research is targeted at violin beginners. The proposed system

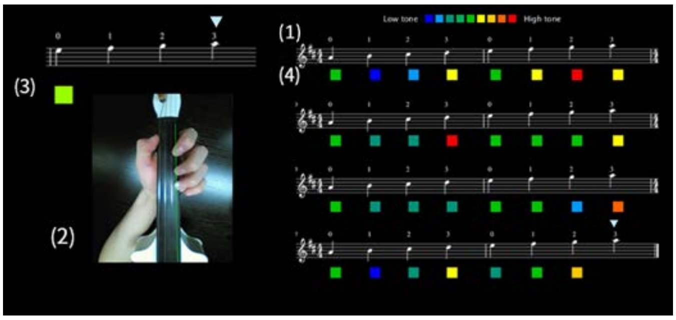


Figure 3 Screen snapshot of visual support mode

makes learners memorize appropriate pressing position and provides diverse learning support functions to cultivate sense of sound. Our aim is for learners, using the learning support provided by the system, to practice a piece of music that they cannot play at all, improve as quickly as possible, and ultimately be able to play without making any positioning errors, even without using system learning support.

System Structure

The system structure of the proposed positioning learning support system is shown in Figure 2. A USB camera is attached to the scroll of a MIDI violin. The fingerboard footage recorded by the USB camera, foot pedal operation information, and MIDI data taken during performance (note number, string number, velocity, pitch bend) are sent to a PC. Based on this information, the PC creates and presents learning support information.

Support Information

This subsection explains the support information provided by the system. The proposed system manages two different types of modal (sight and sound) support information, as modes. Learners use the two modes selectively to learn positioning. Figures 3 and Figure 4 are screenshots of each mode. We will now explain the functions of each mode. The numbers in the figures correspond to the numbers in the itemization given below.

Visual Support Mode.

(1) Musical Score Presentation Function for Scale Practice

This is a function that presents a musical score. The light blue triangle shown above the notes (in Figure 3, above the right-most notation on the bottom-most staff) is an indicator of current playing position and will move to the next note when the designated switch is pressed on the foot controller. Additionally, finger numbers are written above all the notes. Finger number refers to the number representing the finger that should be used when playing a certain note (1: Index finger, 2: middle finger, 3: fourth finger 4: little finger). An enlarged view of playing position is displayed in the upper left of the screen.

(2) Pressing Position Presentation Function

The pressing position corresponding to the notation under the light blue triangle explained in (1) is indicated by a point of light. In this way, learners can intuitively understand the position they should press.

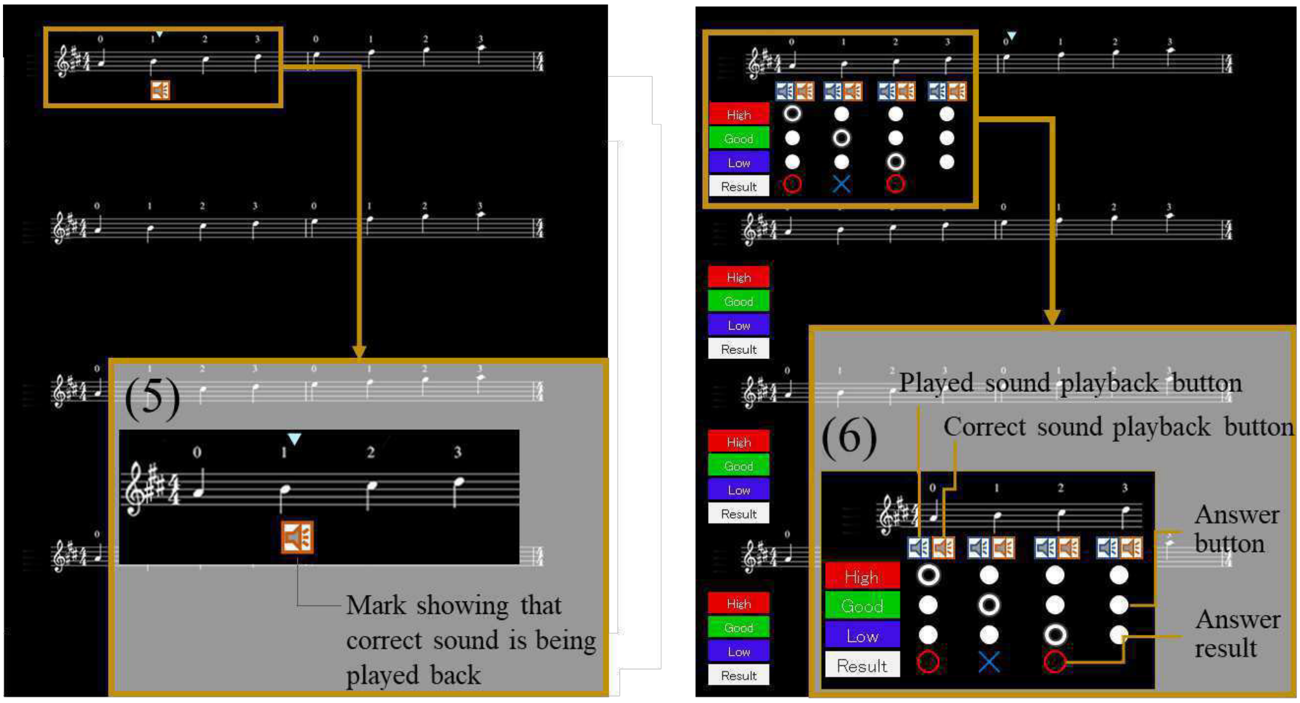


Figure 4 Screen snapshot of auditory support mode

Pitch difference (cent)	- -100	-99 - -75	-74 - -50	-49 - -25	-24 - +24	+25 - +49	+50 - +74	+75 - +99	+100 -
Rectangle color									

Figure 5 Pitch difference feedback

(3) Pitch Difference Presentation Function

The pitch difference between played sound and correct sound is presented in real-time using nine colors of rectangle. Figure 5 shows the criteria that decide the color of the rectangle. When correct sound and played sound are mostly consistent, a yellow-green rectangle is displayed. That is, if the pitch difference between played sound and correct sound is within ±25 cents, the rectangle will be yellow-green. When one pitch in a unison (two sounds of the same pitch) is gradually shifted, the sound becomes jarring. The pitch difference at which the jarring can be clearly perceived is approximately 25 cents, thus we decided on ±25 cents as the standard pitch difference for displaying a yellow-green rectangle. The lower played sound is compared to correct sound, the further the color of the rectangle shifts from yellow-green towards blue, becoming blue when pitch difference exceeds a semitone (100 cents). In contrast, the higher played sound is compared to correct sound, the further the color shifts towards red, in the same way becoming red when pitch difference exceeds a semitone. In this way, learners can comprehend pitch difference visually.

(4) Pitch Difference Log Presentation Function

This function presents a log of the pitch difference explained in (3). The pitch difference recorded in the log is the pitch difference from the time of first plucking the note corresponding to the light blue triangle mentioned in (1). Learners correct their positioning while using the functions detailed in (1)-(3). Accordingly, by leaving a log at the time when a note is first played, learners can check the places where they are prone to make mistakes, as well as the degree of error.

Audio Support Mode.

(5) Correct Sound Presentation Function

As shown in Figure 2, when the foot controller switch corresponding to correct sound presentation function is selected, correct sound is played. Pressing the switch a second time stops sound playback. During correct sound playback only, a speaker icon is displayed directly beneath the note. This function enables learners to adjust pressing position while listening to and comparing correct sound and played sound, or else listen to correct sound before playing then confirm pitch difference using sense of hearing when correct sound is muted, thus memorizing correct sound by ear. In this way correct sound presentation function supports cultivation of sense of sound.

(6) Correct Sound Presentation Function

As shown in the right-hand side of Figure 4, a sound of the same pitch as played sound is played, the learner is made to judge whether the sound is the same as correct sound, or

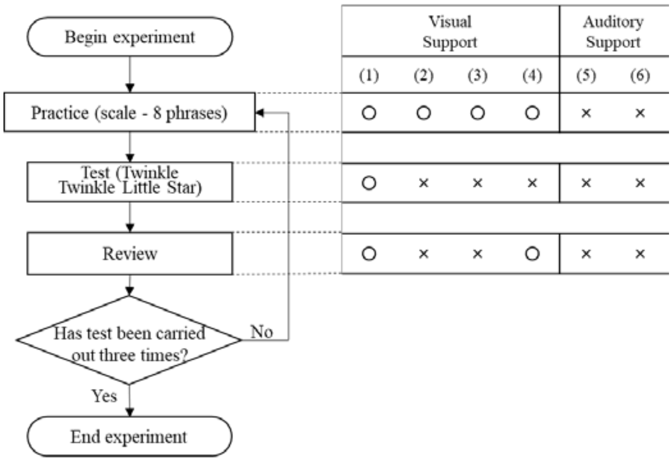


Figure 6 Flow of experiment and available functions for control group

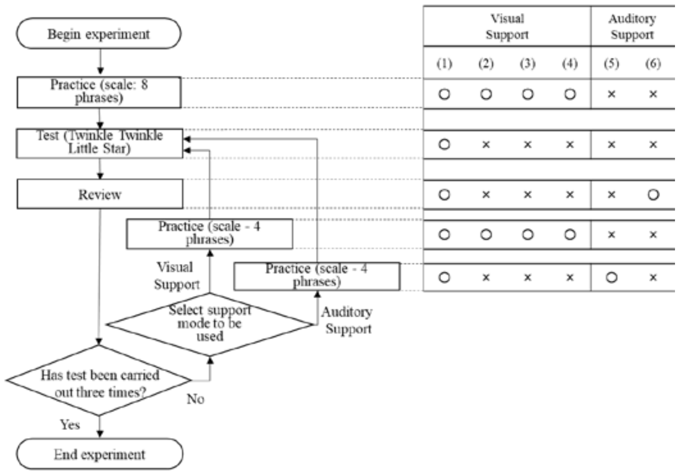


Figure 7 Flow of experiment and available functions for experiment group

higher or lower, then if the learner's answer is correct a red circle is displayed and if it is incorrect a blue x is displayed. In the case of function (5) we suppose that learners use the function while playing the instrument. In the case of function (6) we suppose that it will be used when not playing. Furthermore, (5) can only play the correct sound for the note at current playing position, whereas (6) enables the listening to and comparing of the played sounds and correct sounds of all previously played notes. Accordingly, we suppose that this function will be used to review performance. Due to having to concentrate on positioning, fingering and plucking while playing, it is difficult for learners to concentrate on played sound itself. By comparing with correct sound in an environment that allows them to concentrate on listening to played sound, learners can measure their own sense of sound skills from a different viewpoint. For example, when a learner uses playback, when not playing, to judge whether pitch is correct, and, in the case of pitch being incorrect, misjudges whether it is too high or too low, that learner will have even more difficulty judging while playing. Therefore, it is necessary to listen to and compare correct sound and played sound to hone sense of sound. In contrast, even if a positioning error causes played sound and correct sound to differ, if the player can judge the fact that something is wrong by his/her ears alone, it is possible to correct the positioning error. By using function (6), learners can correctly judge whether it is sense of sound or memorization of positioning that they need to focus on, and progress learning efficiently. It should be noted that the answer button can change up to 9 levels, but the default is 3.

Evaluative Experiment

In the evaluative experiment, taking violin beginners as our target users, we assessed the effectiveness of the proposed system by measuring accuracy of positioning, which uses pitch difference between played sound and correct sound as an evaluation index.

Experiment Procedure

Subjects. There were 14 subjects, all of whom had no experience of the violin. How to use the system and how to play the violin were explained to the subjects in advance. Specifically, the subjects were instructed to play 'pizzicato': using the index finger of the right hand to pluck the strings. Usually, the violin would be played by bowing: using the bow, held in the right hand, to produce a sound. However, even how to hold and position the bow and basic bowing present a high difficulty level for beginners and this would possibly prevent them from being able to concentrate on positioning. For this reason, we used pizzicato.

Assigned Piece. As an assigned piece of music for practice, we used a scale of one octave of A major. There were two types of assigned practice piece; a piece involving playing the scale of one octave of A major eight times (hereafter, scale - 8 phrases), and a piece involving playing the scale four times (hereafter, scale - 4 phrases). As the assigned test piece, Twinkle Twinkle Little Star (A major) was chosen. The reason for this is that it is suited to beginners, as it uses all the notes from the practice piece and has a simple rhythm and fingering. We confirmed that all subjects had heard the test piece but had never played it before.

Control Group and Experiment Group. The 14 subjects were divided into two groups of 7: the control group and the experiment group. The control group used only visual support, while the experiment group used both visual and auditory support.

Flow of the Experiment. The experiment consisted of practice, test and review. Figure 6 presents the flow of the experiment for the control group, as well as the functions they were permitted to use at each stage. The numbers 1-6 in the figure correspond to the numbers of the functions detailed in the former section. Numbers with a circle are functions permitted for use, while numbers with an X are forbidden. The subjects in the control group used all the functions of the visual support mode introduced in the former section. For practice they played scale: 8 phrases. The subjects were allowed to practice each note as many times as were necessary for them to be satisfied, but once they proceeded to a new note they could not practice the preceding notes again. The maximum time allowed for practice was 30 minutes. During the test, the score of the assigned test piece was displayed using the score

Pitch difference (cent)	- -100	-99 - -75	-74 - -50	-49 - -25	-24 - +24	+25 - +49	+50 - +74	+75 - +99	+100 -
Point	1	2	3	4	5	4	3	2	1

Figure 8 Mapping of pitch difference and score

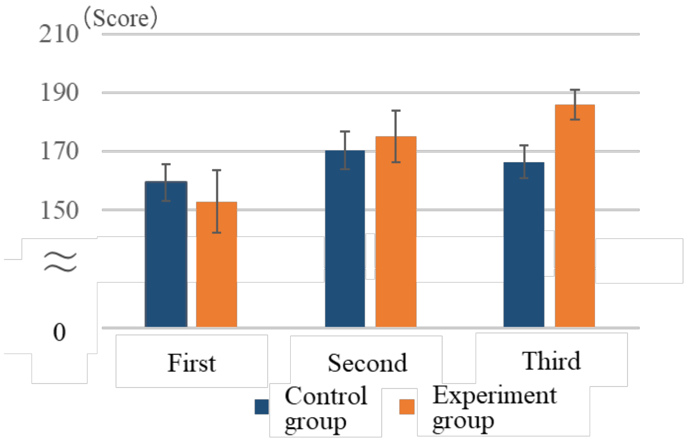


Figure 9 Average score for each test

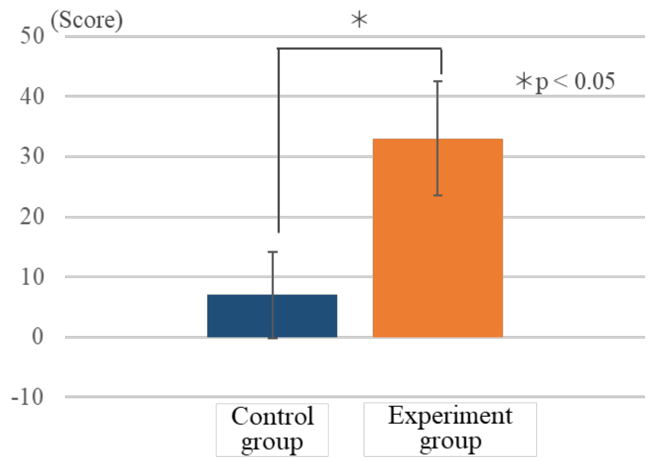


Figure 10 Increase in the average score between the first test and the third test

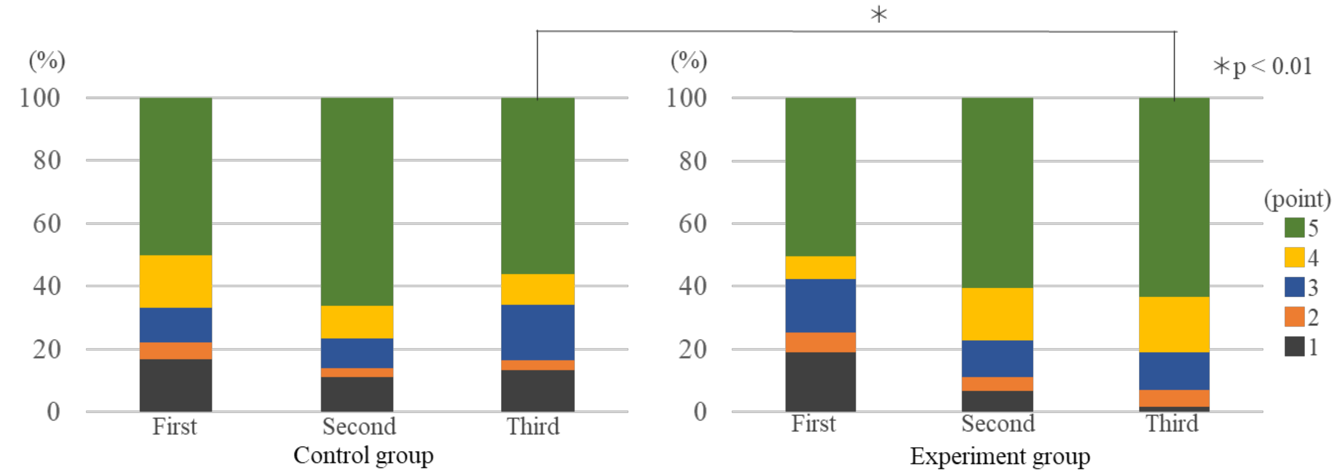


Figure 11 Ratio of score for each test in each group

presentation function (1). The subjects played the test piece only once, from beginning to end, with a time limit of 5 minutes. Finally, in the review stage, the pitch difference of correct sound and played sound for each note of the test piece was presented using the pitch difference log function (4). The subjects had a time limit of 5 minutes to carry out review operations. The whole process from practice to review constituted one set, and a total of 3 sets were carried out.

Figure 7 presents the flow of the experiment for the experiment group, as well as the functions they were permitted to use at each stage. The numbers 1-6 correspond to the numbers of the functions detailed in the former section. Numbers with a circle are functions permitted for use, while numbers with an X are forbidden. The subjects in the experiment group selectively used all the functions of visual support mode and auditory support mode. The subjects selected whether to use visual support mode or auditory support mode before beginning practice.

Subjects were free to choose which mode to use in each set but were not permitted to change midway through practice. Using their chosen modes, the subjects played the practice piece from beginning to end only once. They were allowed to practice each note multiple times until satisfied, but once they proceeded to a new note they could not practice the preceding notes again. During the test, as with the control group, the score of the assigned test piece was displayed using the score presentation function (1). The subjects played the test piece only once, from beginning to end, with a time limit of 5 minutes. Finally, in the review stage, the subjects used the sense of sound check function (6) to assess the test results of Twinkle Twinkle Little Star. The subjects had a time limit of 10 minutes to carry out review operations. The assigned practice piece differed between the first set and the second set onwards. In detail, in set 1 subjects practiced 8 phrases of the scale as the assigned practice piece, whereas from set 2 onwards they only practiced 4 phrases. Accordingly, the maximum time allowed for practice was 30 minutes for set 1, and 20 minutes for set 2 onwards. The reason for altering the time and amount of practice was that the maximum time allowed for reviewing was increased, because subjects used the sense of sound check function to review. Once the experiment was finished, the subjects were asked to complete a questionnaire about their impressions and the effects of each function, as shown in tables 6.1 and 6.2. The subjects were told to comment freely on the reasons for their answers to the questionnaire. The questionnaire used the 5 level Likert scale method (1: completely disagree - 5: completely agree). In cases where additional investigation was necessary, we conducted a semi-structured interview.

Evaluation Index. Positioning accuracy level is evaluated based on the pitch difference (played sound-correct sound) for each note of the assigned test piece. As shown in Figure 8, a pitch difference within 25 cents between correct sound and played sound earns the maximum score (5 points). As the difference increases, the amount of points that can be obtained decreases (the minimum score is 1 point). The test piece contains a total of 42 notes, meaning that the maximum score is 210 points (5 points x 42 notes). Based on this scoring standard, the average score for each group and each test was calculated. It should be noted that, because this research focuses on positioning learning, velocity and accuracy of rhythm during performance are not taken into account.

Results

Average Test Scores

The average test scores of each group for sets 1 to 3 are presented in Figure 9, and the increase in the average score between first test and third test is presented in Figure 10. Comparing the transition of the average test score from set 1 to set 3, it can be seen that the average score in the control group increased from set 1 to set 2 but decreased in set 3. In contrast, the average score in the experiment group increased from set 1 to set 3. Conducting a two-sided t-test at a significance level of 5% on the score of each test revealed that in set 1 and set 2 a significant difference was not observed, whereas in set 3 a significant difference was observed. Furthermore, when conducting a two-sided t-test at a significance level of 5% on the increase value of the average score of each group from test set 1 to set 3, a significant difference was observed. These results demonstrate that the experiment group was able to learn positioning more efficiently than the control group.

Score Ratio per Note

Figure 11 expresses the ratio of the points obtained by each group for all 42 notes in the assigned test piece. There was no notable trend in the increase and decrease of the rate of points obtained in the control group, thus slow progress can be observed. In the experiment group, the rate of obtaining 1 point decreased greatly with each test, while the rate of obtaining 4 points and 5 points increased. Conducting a Chi-squared test on the ratio of obtained points for each group in test set 3 produced the result $X^2(4) = 38.47$, $p < 0.01$, showing that the difference was significant.

Conclusion

In this research, we constructed a positioning learning support system with multimodal information presentation, for violin beginners. Comparison of the average test scores obtained in the evaluation experiment showed that the score of the experiment group was lower than that of the control group in set 1, whereas the score of the experiment group was higher in set 3. The difference between the value of the increase in the average score of each group from set 1 to set 3 was significant. Additionally, analysis of the points obtained for each note in the test shows that there was no clear trend in the increase and decrease of points scored by the control group, while in the experiment group, instances of obtaining 1 point greatly decreased and in turn instances of obtaining high scores of 4 or 5 points increased. The difference in the ratios of points obtained for each note in test set 3 was also significant. In accordance, it became evident that, rather than using visual support alone, using multimodal support produced better learning effects.

As future work, we propose evaluating the system in the case of not using the sense of sound check function, as well as additional experiments to inspect the specific characteristics of each function. Also, as the evaluation experiment in this research was short-term, it is necessary to verify the usability of the system in the case of long-term use. Furthermore, we will examine the time required to learn positioning for pieces of music of varying difficulty.

Acknowledgements. This work was supported by JSPS KAKENHI Grant Numbers JP16K12560, JP15K00279.

References

Jian-Heng, W., Siang-An, W., Wen-Chieh, C., Ken-Ning, C., & Herng-Yow, C. (2012). Real-Time Pitch Training System for Violin Learners, *Proceedings of Multimedia and Expo Workshops* (pp. 163-168).

Kumaki, M., Takegawa, Y., & Hirata, K. (2017). Evaluation of Positioning Learning Support System using True Information and False Information and Vague Information for Violin Beginner, *Proceedings of International Computer Music Conference* (pp. 150-155).

Nagaya, S., Seki, S., & Oka, R. (1996). A Theoretical Tonsideration of Pattern Space Trajectory for Gesture Spotting Recognition; Automatic Face and Gesture Recognition, *Proceedings of International Conference on Automatic Face and Gesture Recognition* (pp. 72-77).

Oviatt, S., & Cohen, P. (2000). Perceptual User Interfaces: Multimodal Interfaces that Process What Comes Naturally, *Communications of the ACM*, 43, 45-53.

Takegawa, Y., Terada, T., & Tsukamoto, M. (2011). Design and Implementation of a Piano Practice Support System using a Real-Time Fingering Recognition Technique, *Proceeding of International Computer Music Conference* (pp. 1-8).

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Design and Implementation of a Support Tool for Piano Teachers to Identify Bad Fingering Habits

Yoshinari Takegawa,[1,*] Asami Hasegawa,[2,*] and Keiji Hirata[3]

[1,2,3] *Graduate School of Systems Information Science, Future University Hakodate, Japan*

[1]yoshi@fun.ac.jp, [2]g2117038@fun.ac.jp, [3]hirata@fun.ac.jp

Abstract

The aim of this research is the proposal of a piano playing fingering analysis tool for piano teachers. Fingering involves subtle alterations of the strength and speed of keying and is vital as the basis of high level performance technique. However, in regard to flexibility of joints, strength of the connection of tendons and etc., the structure of the hand varies from person to person. In certain performers bad fingering habits can be seen, such as straining when pressing keys or distancing the fingers from the keyboard. In the majority of face-to-face lessons, a piano teacher can only see the pupil's hands and fingers from one direction and cannot thoroughly observe fingering. Furthermore, even if pupils have bad habits, there are cases in which it sounds as though they are playing well, meaning that the presence or absence of bad habits cannot be judged at one hearing. For this reason, it is difficult for piano teachers to point out all the bad fingering habits pupils exhibit while playing. We conducted a fingering observation experiment on subjects who had won prizes in piano competitions and identified the functions necessary for analysis of fingering. The proposed tool has functions that enable users to view video from multiple viewpoints simultaneously and to adjust playback position by clicking on the notations on the musical score.

Introduction

Piano playing demands high level performance techniques such as subtle alterations of the strength and speed of keying. Fingering is a basic skill essential to high-level performance. In this research, we define fingering as three general terms: movement of the fingers, independence of the fingers and shape of the hands.

- Movement of the fingers: the correspondence of finger numbers to musical notations (Fig. 1). Good fingering is that in which the distance travelled by the fingers is short and the proceeding notes can be played smoothly. When playing the notes in Fig. 1 using descending finger numbers, keying will be performed in the order of little finger – thumb. In this order, because the distanced covered by the fingers is long, keying is delayed and the player cannot perform with correct rhythm. Accordingly, keying can be performed more smoothly using ascending finger numbers.
- Independence of fingers: moving each finger individually (Fig. 2). To reduce muscle fatigue, a state in which the player moves each finger without exertion is desirable. In the example shown in Fig. 2, in which the little finger is not independent, when the ring finger presses a key the little finger not being used for keying is raised, unlike in the case of the little finger being independent. The operation of raising the finger uses strength. By reducing the fatigue incurred by a single keystroke, it becomes possible to perform a greater number of keystrokes.
- Shape of the hands: the angle of each finger in relation to the keyboard, at the time of keying. A state in which the fingers are vertical to the keyboard is good. In the case of 'frog-like' fingers (Fig. 3) or slanting fingers (Fig. 4), the power to press keys becomes weaker. Therefore, when playing a loud sound, it is better to have the fingers vertical to the keyboard as though holding an egg, as shown in Fig. 3.

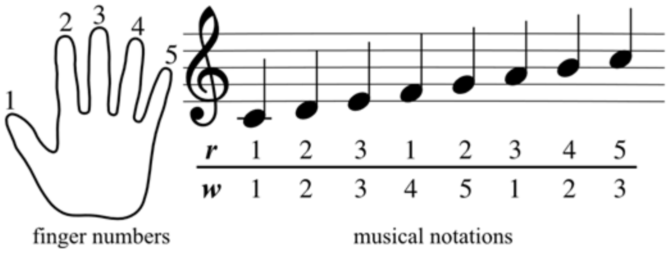


Figure 1. Finger numbers and example of musical notations. A row of 'r' means right movement of the fingers, a row of 'w' means wrong one.

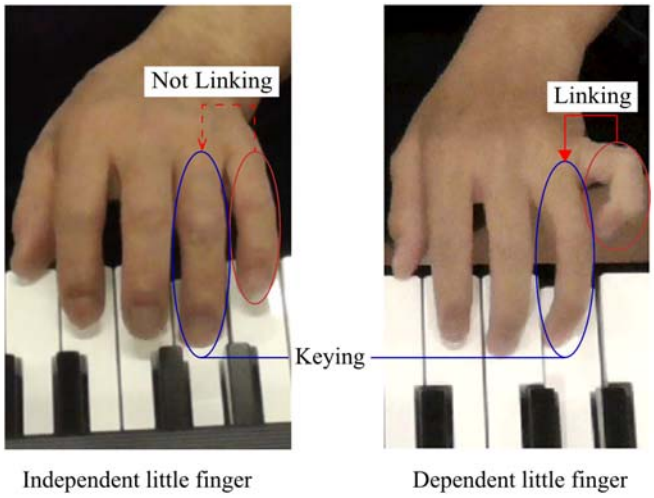


Figure 2. Examples of independence and dependence

If fingering is undeveloped, this appears as unnecessary movement in fingering while playing. For example, as in the case where the little finger is not independent (Fig. 2), the finger not used for keying is left raised or, as shown in Fig. 4, the fingers become slanted when keying. This kind of unnecessary movement during playing is not coincidental, but rather is established as a bad habit. In particular, children, because their bodies are undeveloped and their joints are flexible, are liable to develop bad habits if they play only loud and fast pieces. Establishing bad habits results in the problem that pupils cannot move their fingers smoothly and thus

* Joint first author

$ Naofumi Kaneshige (Faculty of Education, Mie University, Japan) is the fourth author.

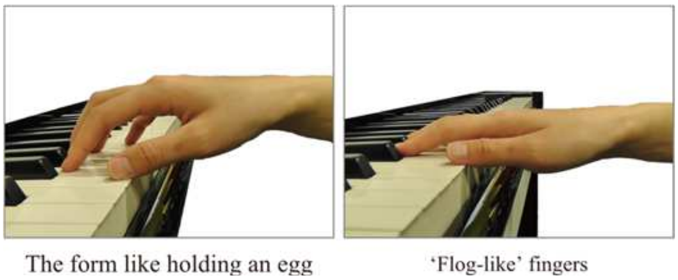


Figure 3. Example of shape of the hands

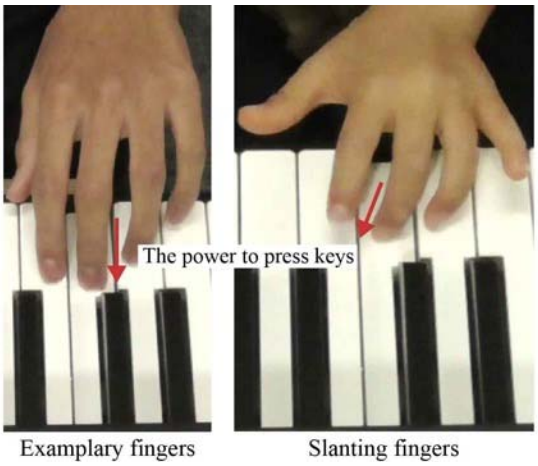


Figure 4. Example of keystrokes

cannot play as they wish. Therefore, it is vital to improve bad habits according to each pupil's body.

However, for piano teachers to discover all their pupils' bad playing habits is difficult, due to the problems of visibility and method of evaluating pupils' performance. The following are two problems with visibility. In an ordinary lesson, the teacher can only see the pupil's hands from one direction at a time. Also, depending on the piece of music, the movement of each finger may be fast, making it impossible to watch them carefully. In regard to the problem of evaluation method, teachers have a tendency to evaluate pupils' playing based on sound. Accordingly, there are cases in which, even if pupils have bad habits, at first hearing it sounds as though they are playing well, meaning that the presence or absence of bad habits cannot be judged at only one hearing.

To solve the aforementioned problems, this research aims to design and implementation of a support tool for piano teachers to find bad fingering habits for piano teachers. Through the proposed tool, we expect to support piano teachers' analyses of pupils' fingering and to simplify and reduce the time taken to discover bad habits. Firstly, we conducted a fingering observation experiment using experienced pianists, in order to determine the proposed functions of the fingering analysis tool, from the analysis method used by the 4th author, a piano teacher. After that, 4th -author analysed the experiment data and observed various bad habits. Additionally, we obtained verbal statements about the analysis method from 4th author. Based on these statements we designed a fingering analysis tool.

Related Research

There are many studies of systems to learn piano performance, but most of them are for beginners (Dannernbarg et al, 1990; Raymaekers, Vermeulen, Luyten, & Coninx, 2014; Rogers et al, 2014; Takegawa, Terada, & Tsukamoto, 2012; Takegawa, Tsukamoto, & Terada, 2011). Beginners can improve their performance by using these systems instead of piano teachers. Although, advanced pianists typically take lessons from their teachers and improve their performance. These systems can not support teachers who have intermediate and advanced students since they target beginners.

On the other hand, there are automatic exercise generation systems (Mukai, Emura, Miura, & Yanagida, 2007; Kitamura & Miura, 2006) to overcome weaknesses of pianists. As in these studies, the way to infer weaknesses of pianists from MIDI is useful for this research.

Research has been made to visualize performance information (volume, tempo, movement of fingers, etc.) for analysis (Hiraga, Mizaki, & Fujishiro, 2002). They designed and implemented a system that three-dimensionally expresses performances in 3D space. The method of expressing performance information using graphics that teachers do not usually see takes a learning cost for the piano teacher to understand the figure. In this research, taking into consideration the learning cost which becomes a burden of the piano teacher, keep it to a simple display such as a graph or a heat map.

Measurement and visualization of movement of fingers during playing have been studied by various methods. For instance, methods using gloves with sensors (Furuya, Flanders, & Soechting, 2011; Rahman, Hossain, Rana, & Mitobe, 2013), a method of installing high speed cameras and capacitive sensors on a piano (MacRitchie & McPherson, 2015) a motion capture with a marker (Goebl & Palmer, 2008). Either method can measure movement of fingers. In this research, we think that teaching methods of each teacher and finger recognition method suitable for the lesson environment are different, we will not limit the method or propose a new proposal.

There is pianoFORTE to support piano lessons (Smoliar, Waterworth, & Kellock, 1995). The system can record the student's performance as a MIDI file and display strength, tempo, articulation and key timing on the score. With this system, the teacher can see the performance of the students as objective data. However, the system is unsuitable for finding bad habits of fingering since it does not record finger movements.

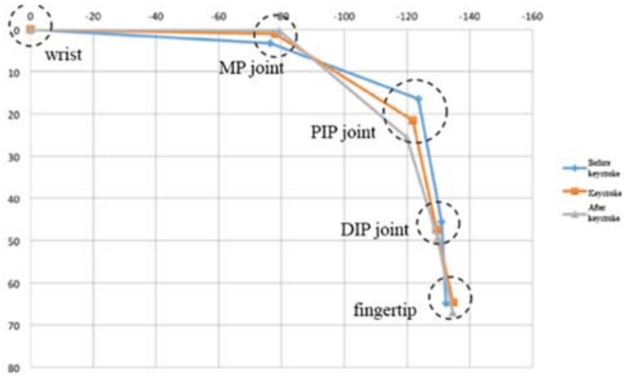


Figure 5. Example of bone stick

Figure 6. Assigned piece

Figure 7. Places of markers

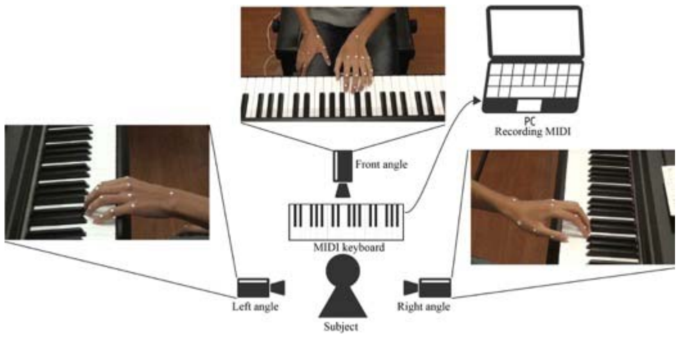


Figure 8. Experiment environment

Advanced Pianists' Fingering Observation Experiment

To find out what functions are necessary for the proposed tool, we conducted an observation experiment on the fingering of advanced pianists. The subjects, to whom markers were attached to enable movement following of their fingers during performance, were filmed from three viewpoints playing the assigned piece of music. After the experiment, fingering was analyzed based on IOI(Inter-Onset Interval) and velocity, finger motion data and performance video obtained during the experiment. IOI and velocity were visualized by the visualization tool, details of which are given hereafter. From performance video, the markers attached to the subjects were analyzed by the movement analysis software MOVIAS Neo (nac Image Teachnology Inc.) and bone sticks (Fig. 5) were created, at the states of before and after keying.

Experiment Environment

To enable movement following of the fingers during performance, white stickers were stuck on the subjects' hands to acts as markers. A white marker was stuck on every joint of the right and left hands (Fig. 7). As shown in Fig. 8, the arrangement during the experiment was 3 cameras, placed to the right and left and in front of the subjects, and a MIDI keyboard connected to a computer.

Introductions to Subjects

The subjects were given the following instructions before beginning the experiment:

- Play with a tempo of between 60 and 72 beats per minute
- Play with consistent keying strength
- When playing with only one hand, place the other hand on your knee
- Try to play looking at the score

Experiment Flow

The flow of the experiment was as follows:

(1) Subjects played the assigned piece with the right hand 3 times, starting over from the beginning if a keying mistake was made.

(2) Then, in the same way as in (1), subjects played with the left hand.

Analysis Tools

To analyze experiment data, we used the 3 tools described below.

Visualization Tool. This tool records MIDI input from the MIDI keyboard, and, as shown in Fig. 9, has a function to display IOI and velocity on the musical score. It is possible to visualize data as either a graph or a heatmap. A heatmap enables simultaneous visualization of IOI and velocity, whereas a graph can only visualize one aspect at a time. A graph can be displayed a plurality of IOI or velocity by putting on the other.

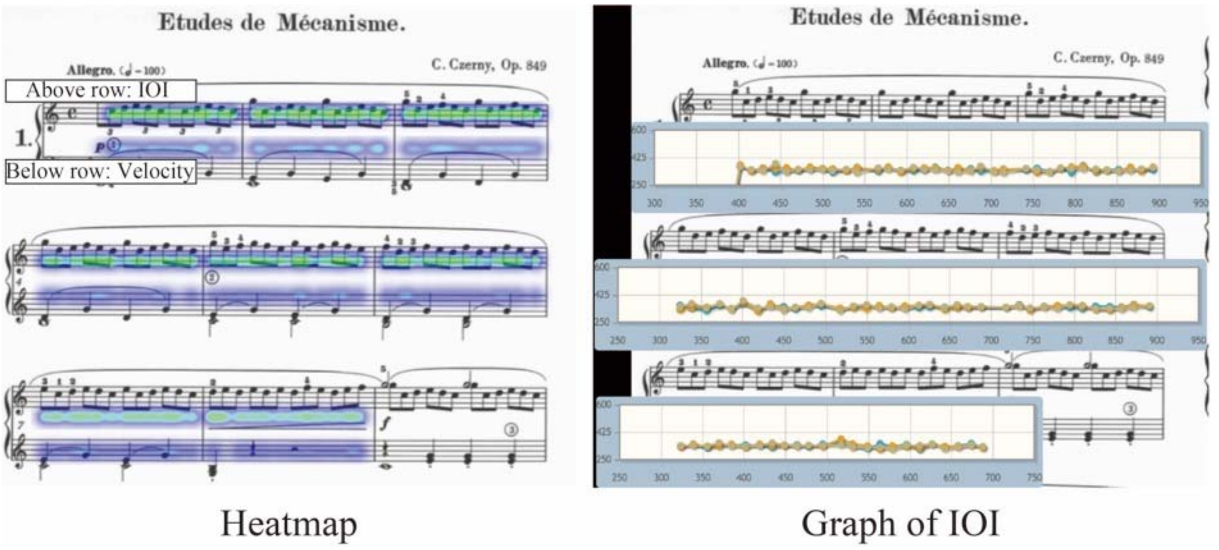


Figure 9. Visualization methods

Video Player. We used the video editing software TMPGEne Video MasterringWorks 6 (Pegasys Inc.) as a video player to view performance footage. This software enables not only ordinary playback, but also seamless enlargement and reduction of video. Video can be enlarged or reduced during playback and we used these functions for analysis.

Movement Analysis Tool. We used MOVIAS Neo (nac Image Technology Inc.) in the movement analysis of advanced pianists fingering. When markers are attached to the target for movement analysis, this software can automatically follow movement from video images. In our research, we analyze in three dimensions and convert the movement of each joint to a numerical value.

Results and Considerations

As a result of the experiment, we observed multiple bad fingering habits. For example, the little finger was extending when keying with the middle finger (Fig. 4), and keying with the fingers curled (Fig. 2).

The Piano teacher (4^{th} author) observed fingering by sight, from the performance video, focusing on projecting points on the graphs for IOI and velocity. He primarily played performance video of the viewpoint filmed from directly in front of subjects. When observing, the teacher played video of the same part of performance over and over again. Additionally, regarding fingers that not constant IOI or velocity, after confirming this on the performance video, the teacher created bone sticks from motion data and analyzed whether bad habits could be observed objectively.

Fingers with bad habits could be determined from IOI and velocity. Bad habits could be confirmed from video filmed from directly in front, but we considered that more accurate judgement could be made by also viewing video filmed from the right and left. As the teacher were playing back the same part of performance over and over, it is evident that, for analysis by sight, frame by frame playback would be effective. Furthermore, the time required for analysis could be reduced by making it possible to carry out playback of the same performance position with a single operation.

Proposed Tool

Proposed Functions

Based on the results of the experiment, we propose the following 5 functions. Fig. 10 gives an overview of the proposed tool when being used by a user.

Multiple Viewpoint Performance Video Playback Function. As shown in Fig 10(a), this is a function that synchronizes and plays back performance video filmed from 3 viewpoints. When performance video is viewed using the proposed tool, it is played back by this function. The function is operated by clicking on notations on the score or clicking the play button, as presented in Fig 10(b). Rather than watching only video filmed from the front, also watching video filmed from the right and left enabled more accurate judgement of certain bad habits, demonstrating that simultaneous viewing of multiple viewpoints is effective. It can be expected that the problem of visibility, mentioned earlier, can be solved by this function.

Note-by-note Playback Function. As shown in Fig. 10(b), this is a function that, when each note is clicked on, can play back video from the key approach to key departure of the finger playing the relevant note. In this function, key approach refers to the action of the finger approaching the key and key release refers to the action of the finger releasing the key. Bad habits appear not only at the moment of keying, but also during key approach and key release, thus we design the function to allow observation of the movement from key approach to key release. As, in fingering analysis, it is necessary to view the same part over and over, we aim to make operation more efficient by removing the need to adjust a seek bar on the video player.

Playback Speed Adjustment Function. This function that can make performance video playback speed slower or faster. Moving the knob of the slider shown in Fig 10(c) to the left, as viewed by the user, makes playback speed slower, while moving the knob to the right makes it faster. Making playback speed slower allows the user to observe carefully.

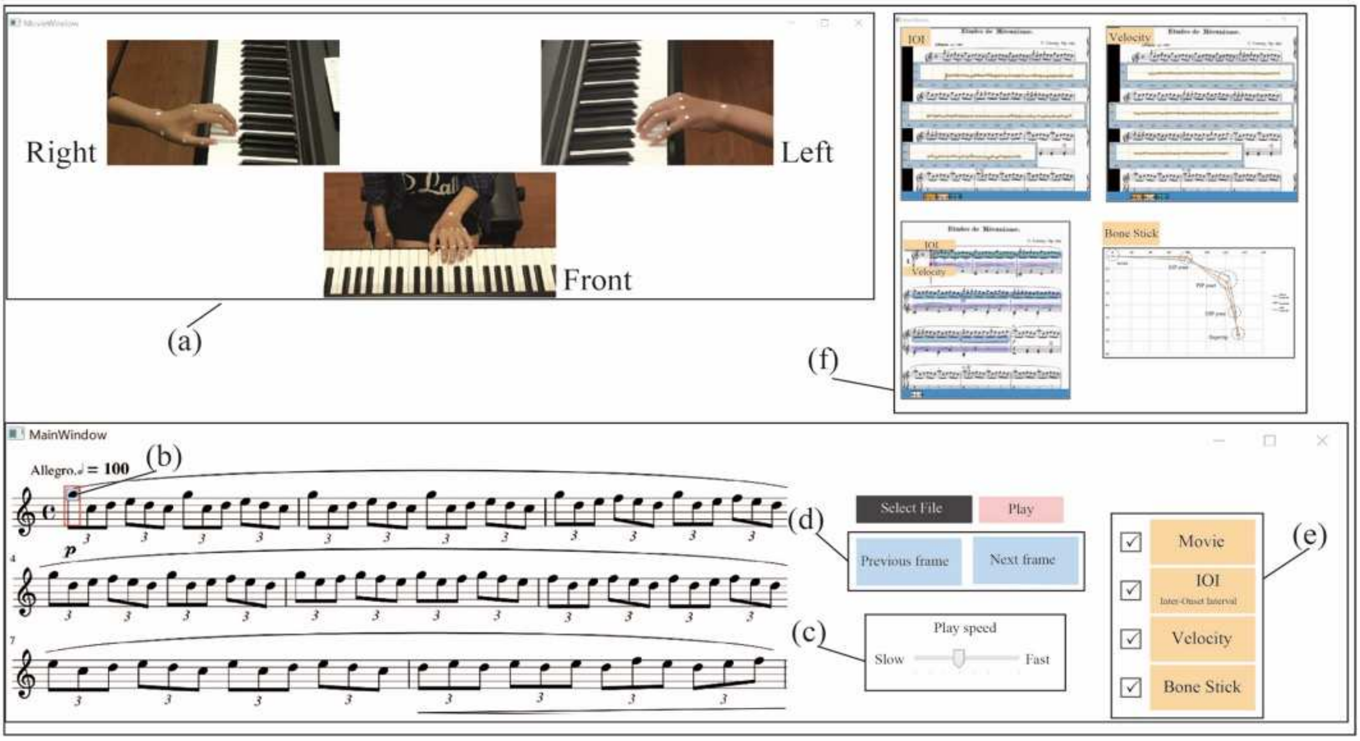


Figure 10. Proposed tool's overall view

Making playback speed faster allows the user to save time when he or she wants to view the entire performance several times. This solves the problem of visibility and results in a reduction of bad habit discovery time.

Frame-by-frame Playback Function. This function returns or advances stills from the performance video 1 frames at a time. Looking at Fig 10(d), clicking the button on the right, as viewed by the user, returns by 25 milliseconds while clicking the button on the left advances by 25 milliseconds. The fingering corresponding to a certain playback time can be confirmed two-fold by combining bones with viewing of still performance video. It can be expected that the problems of visibility and reconfirmation will be solved by this function.

Data Switching Function. This is a function that switching show or hide data. Data mean 5 items: performance movie, graph of IOI, graph of velocity, heatmap, bone stick. When a box in Fig 10(e) is checked, showed in Fig 10(f). When a box in Fig 10(e) is unchecked, hide in Fig10(f). Only for heatmap, it is shown when IOI or velocity is checked. Graphs and heatmap that showing IOI and velocity are created using visualization tools. In chapter 2 related research we mentioned that recognition method of fingers movement is not limited in this research. Fig. 10 (f) The figure of the bone stick is an example using the same method as the observation experiment of fingering.

Usage

Table 1 shows an example of how to use proposed tool. Next sentences are supplementary explanation for Table 1. In Step 2, a user think piano performance is sound flow, so play from the beginning to end. And the user narrows down points expressing bad fingering. Next Step 3, the user checks the effect on data. In Step 4, the user visually checks again. In Step 5, the user checks another place to judge finger's movement whether is always happened movement or rarely happened movement. We expect a usage like this example, using piano teacher's experience and objective data to find bad fingerings.

Conclusion

In this research, we proposed a support tool for piano teachers to find bad fingering habits. By conducting an observation experiment on the fingering of advanced pianists and getting verbal statements about analysis methods from the piano teacher (4th author), we proposed 5 functions to support the analysis of fingering. Hereafter we plan to evaluate and improve these functions.

Acknowledgements. In the progression of this research, I was greatly aided by the advice of associate professor Asuka Terai. Kentaro Ueda, graduates of Hirata/Takegawa laboratory, also contributed observation experiment and development of the visualization tool. This work was supported by JSPS KAKENHI Grant Numbers JP16K12560, JP15K13226, JP15K00279.

Table 1. Example of usage

Step	Action	Obtained data
1	Click Select File button in Fig10 Select performance video	-
2	Click Play button in Fig10 Watch and listen to whole video	Visual data Auditory data Intuition
3	See graphs of IOI and velocity that are place the user thought problem at step 2	MIDI data
4	Slow play speed Play place same step 3 using Note-by-note playback function	Visual data
5	Play another place same movement of fingers with step 4 using Note-by-note playback function	Visual data
6	Judge fingering whether bad fingering or not from step2 to step5	-

References

Dannenberg, R. B., Sanchez, M., Joseph, A., Capell, P., Joseph, R., & Saul, R. (1990). A computer-based multi-media tutor for beginning piano students. *Interface*, 19(2-3), 155–173.

Furuya, S., Flanders, M., & Soechting, J. F. (2011). Hand kinematics of piano playing. *Journal of Neurophysiology*, 106(6), 2849–2864.

Goebl, W., & Palmer, C. (2008). Tactile feedback and timing accuracy in piano performance. *Experimental Brain Research*, 186(3), 471–479.

Hiraga, R., Mizaki, R., & Fujishiro, I. (2002). Performance visualization: a new challenge to music through visualization. In Proceedings of the tenth ACM international conference on Multimedia (pp. 239–242).

Kitamura, T., & Miura, M. (2006). Constructing a support system for self-learning playing the piano at the beginning stage (pp. 258–262). Presented at the 9th International Conference on Music Perception and Cognition.

MacRitchie, J., & McPherson, A. P. (2015). Integrating optical finger motion tracking with surface touch events. *Frontiers* in Psychology, 6, 702.

Mukai, M., Emura, N., Miura, M., & Yanagida, M. (2007). Generation of suitable phrases for basic training to overcome. Presented at the 19th, International Congress on Acoustics, Madrid.

Mostafizur, R., & Aowlad, H., & Masud, R., & Kazutaka, M. (2013). Hand motion capture system in piano playing. Paper presented at International Conference on Informatics, Electronics Vision (ICIEV), Dhaka.

nac Image Technology Inc. MOVIAS Neo. Retrieved from https://www.nacinc.jp/analysis/software/movias-neo/

Pegasys Inc. TMPGEnc Video Mastering Works 6. Retrieved from http://tmpgenc.pegasys-inc.com/en/product/tvmw6.html

Rahman, M. M., Hossain, A. B. M. A., Rana, M. M., & Mitobe, K. (2013). Hand motion capture system in piano playing. In 2013 International Conference on Informatics, Electronics Vision (ICIEV) (pp. 1–5). Dhaka: IEEE.

Raymaekers, L., Vermeulen, J., Luyten, K., & Coninx, K. (2014). Game of tones: learning to play songs on a piano using projected instructions and games. Poster presented at CHI '14 Extended Abstracts on Human Factors in Computing Systems (pp. 411–414).

Rogers, K., Röhlig, A., Weing, M., Gugenheimer, J., Könings, B., Klepsch, M., … Weber, M. (2014). P.I.A.N.O.: Faster Piano Learning with Interactive Projection. In Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (pp. 149–158). New York, NY, USA: ACM.

Smoliar, S. W., Waterworth, J. A., & Kellock, P. R. (1995). pianoFORTE: A System for Piano Education Beyond Notation Literacy. In Proceedings of the Third ACM International Conference on Multimedia (pp. 457–465). New York, NY, USA: ACM.

Takegawa, Y., Terada, T., & Tsukamoto, M. (2012). A Piano Learning Support System considering rhythm. Presented at the International Computer Music Conference.

Takegawa, Y., Tsukamoto, M., & Terada, T. (2011). Design and Implementation of a Piano Practice Support System using a Real-Time Fingering Recognition Technique (pp. 1–8). Presented at the International Computer Music Conference.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Something to tell: musical work interpretations of the performer as narrator

Matías Tanco[1]

[1] *Laboratory for the Study of Musical Experience (LEEM-FBA-UNLP)*

[1]matiastanco@fba.unlp.edu.ar

Abstract

The relationship between music and narrative has been extensively discussed since Nattiez (1990) onwards, establishing similarities between music and literary discourse, and its adaption to narratology. Beyond its literary form, narration is present in everyday life and is considered as a way of thinking, interpreting and creating a sense of the world and life, within the framework of one's own culture (Bruner, 1986). Narration and music are linked in inter-subjective communication from early childhood, being together in time without the need for a grammatical discourse (Malloch & Trevarthen, 1999). However, little has been written about the performer musician as a teller or narrator, and, as a consequence little is known about the ways a musical piece expresses the narrator's point of view. In this research, semi-structured interviews with performers were conducted to obtain musicians first person perspective on their performances. Results show that although grammatical content is not altered, the performer produces a new personal interpretation, a new way to make sense of the musical work. From a constructivist perspective, the performance of the work can be understood as the performer's construction of a version or transformation of a preceding version. Since narrative is expressed by the narrator's point of view or perspective (Bruner, 2002), we can understand the narration of a musical work from the performer's point of view.

Background

Talking about music is an intricate task that theorists and analysts face when giving linguistic descriptions to non-linguistic sound phenomena. The use of analogies is a frequently used resource to describe different aspects of music. Based on simple relations of similarity, analogies can be used to apply terms and categories of a known domain to music in the form of linguistic statements to explain a phenomenon. Although analogical thinking contributes to knowledge, it is limited since it is unable to compare phenomena from different disciplines..

Many theorists have referred to music as a narrative, addressing musical texts through the implementation of Narratology categories (Maus, 1991; Newcomb, 1987). The main criticisms of these approaches focus on the impossibility of music to organize itself and to signify in the way that language does (Abbate, 1991; Nattiez, 1990). While all agree that narrative facilitates discussion about music in terms of a temporal unfolding, some authors argued that this analogy is made with the purpose of considering tonal music as a language in its own terms (Klein, 2013; McClary, 2000,).

Beyond its literary form, narration is present in everyday life and is considered as a way of thinking, interpreting and creating a sense of the world and life, within the framework of one's own culture (Bruner, 1986). Narration and music are linked in inter-subjective communication from early childhood, being together in time without the need for a grammatical discourse (Español, 2014; Malloch & Trevarthen, 2009).

For decades the musical narrative focused on the creation of a musical discourse, understanding that composers create a linguistic syntax associated to the narrator's function. A possible alternative is to turn towards musical performance and the performer's role, (Abbate, 1991; Meelberg, 2006; Rink, 1994, 1999) to think of a narration that does not end in the score, but rather takes place between the musician and the audience (Shifres, 2007). However, there is little evidence in the literature about the performer musician as a teller or narrator, and as a result little is known about how a musical piece expresses the narrator's point of view.

Aims

The current study aims to investigate performers' descriptions of their own musical performances, from a narrator's point of view.

Method

A semi-structured interview was designed and conducted with 11 performers (9 pianists, 2 guitarists). The musicians were asked about (i) the interpretation of musical works previously performed, (ii) the musical scores, (iii) the process of rehearsal, and (iv) the performance in a concert situation. To develop categories, an analytical guide for coding and create nodes to organize and visualize the information, all interviews were transcribed and analyzed with NVivo 11 software.

Results

The most frequent categories identified in coding analysis were "to narrate", "personal", and "emotions", while another relevant, although less frequent category, "dynamic", describes the performance by integrating the previous categories.

"To narrate" involves the use of linguistic analogies (to speak, to tell something) when the musician expresses him or herself through music; it also involves communication with the audience in the form of narrator-narratee relations; finally it involves descriptions of the performance as a creative activity. The activities of the performer understood as a narrative action involves the use of actions and gestures, which can also be described in terms of energy, or as an acting that exhibits them to others. The creative action of the performer has also been described as "to give life", "to form", "to project", "to build", "transform", "modify", "re-create", "invent", "play" or "to mould".

"Personal" emphasizes the role of the performer in deciding how to interpret the score, compared to the original intention of the composer. This often implies that what is done during the performance is something of their "own" creation, and implies "choosing", "deciding ", or "determining" the actions that are carried out. The expression of "personality", "imprint", "authenticity", and auto-perceived "presence" determine the sense of the "self" that performers intend to communicate through the musical work.

"Emotions" bring together feelings, sensations, and moods that are experienced by performers. In interviews, performers say that emotions are what make communication with the audience possible, they also emphasize their physical perception and the role of the body in the experience of these types of sensations.

"Dynamic" includes processes that occur over time in relation to the transformation or change of state of the work in terms of its interpretation. The dynamic aspect also implies that the performance of the work during performance is intimately linked to the emotional state that the performer presents at each moment.

Conclusion

The evidence provided by the results of this study shows the musicians first person perspective on their performance as a creative or re-creative action. Although grammatical content is not altered, the musician provides a new way to interpret the music in a personal manner, making sense of the musical work.

The narrative that can emerge in a musical performance has more to do with human activity rather than being a resource of linguistic or literary organization. Since only sentient creatures can have or experience emotions (Davies, 2011), the performer's role is required to endow the work with intentionality, so the audience can attribute or share emotions related to music. In this sense, the social context of the performance gives meaning to the piece beyond its pure form (Gomila, 2013).

Unlike the grammatical content of the work, the interpretation is constantly updated and re-narrated in a dynamic manner. This renewal may occur due to repeated performances during the rehearsal or could be the result of emotions that the performer experiences in relation to the audience. From a constructivist perspective, the performance of the work can be understood as the performer's construction or transformation of a preceding version (Bruner, 1986).

The performer's activity understood as a narrative allows us to think about his or her actions in relation to the intentionality of telling something that goes beyond the content of the score, or that arises from it when they are guided toward a communicative intention in relation to the audience. Narrative could also be considered as the rethinking of time and context, to the extent that the music being performed allows the emergence of changes in the performer's self-perception, and also accompanies the musician's self-knowledge processes while interpreting the work.

Since narrative expresses the narrator's point of view or perspective (Bruner, 2002), in the same way we can understand the narration of a musical work from the performer's point of view.

References

Abbate, C. (1991). *Unsung Voices: Opera and Musical Narrative in the Nineteenth Century*. Prínceton: Princeton University Press.

Bruner, J. (1986). *Actual Minds, Possible Worlds*. Cambridge, MA: Harvard University Press.

Bruner, J. (2002). *Making stories. Law, literature, life*. New York: Farrar, Strauss and Giroux.

Davies, S. (2011). Music and metaphor. In S. Davies (Ed.) *Musical Understandings and Other Essays on the Philosophy of Music* (pp. 11-33). Oxford: Oxford University Press.

Español, S. (Comp.) (2014). *Psicología de la Música y del Desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana*. Buenos Aires: Paidós.

Gomila, A. (2013). San Lorenzo en la ESMA: sobre la ambigüedad moral de la música. In F. Shifres, M. de la P. Jaquier, D. Gonnet , M. I. Burcet, & R. Herrera (Eds.) *Actas de ECCoM* (pp. 403-416). Buenos Aires: SACCoM.

Klein, M. L. (2013). Musical story. In M. L. Klein, & N. Reyland (Eds.) *Music and Narrative since 1900* (pp. 3-28). Bloomington: Indiana University Press.

Malloch, S., & Trevarthen, C. (Eds.) (2009). *Communicative musicality: exploring the basis of human companionship*. Oxford & New York: Oxford University Press.

Maus, F. E. (1991). Music as narrative. *Indiana Theory Review, 12*, 1–34.

McClary, S. (2000). The impromptu that trod on a loaf or how music tells storíes. In M. McQuillan (Ed.) *The Narrative Reader* (pp. 166-170). London: Routledge.

Meelberg, V. (2006). *New Sounds, New Stories. Narrativity in Contemporary Music*. Leiden: Leiden University Press.

Nattiez, J.-J. (1990). Can one Speak of Narrativity in Music?. *Journal of the Royal Musical Association, 115*, 240-257.

Newcomb, A. (1987). Schuman and late eighteenth-century narrative strategies. *19th Century Music, 11(2),* 164-174.

Rink, J. (1994). Chopin's Ballades and the Dialectic: Analysis in Historical Perspective. *Music Analysis*, *13(1),* 99-115.

Rink, J. (1999). Translating musical meaning: The nineteenth-century performer as narrator. In N. Cook, & M. Everist (Eds.), *Rethinking Music* (pp. 217-238). Oxford: Oxford University Press.

Shifres, F. (2007). Beyond Cognitivism. Alternative perspectives on the communication of musical structure through performance (Unpublished doctoral dissertation). University of Roehampton, London, UK.

Reasons for liking sad music in a population from Turkey: Relations with Music Empathy and Rumination

Hasan Gürkan Tekman[1] and Engin Can Tokgöz[2]

[1] *Psychology Department, Uludag University, Turkey*
[2] *Psychology Program, Middle East Technical University Northern Cyprus Campus, Turkey*

[1]hgtekman@gmail.com, [2]engin.tokgoz@metu.edu.tr

Abstract

Paradoxically, many people are attracted to works of art that evoke negative emotions such as sadness and fear. Research on why people like sad music has shown that certain personality traits including empathy and rumination. Music empathy, which has been defined as an empathic style of thinking specific to music, has also been found to be associated with liking sad music. There is no empirical work on liking sad music in Turkey despite existence of Turkish styles known for their specialization on sadness and distress. Our aim was to assess the relation of liking sad music with music empathy and rumination in a Turkish sample. The reliability and validity of a Turkish adaptation of a scale for music empathizing was also established for this purpose. Both the Turkish translation of the Like Sad Music Scale (LSMS) and Turkish Music Empathizing (ME) scale showed sufficient internal consistency. The significant correlation of the ME scale and the Basic Empathy Scale is evidence of its validity. ME correlated significantly with LSMS and was a significant predictor of LSMS after the variance due to age, gender, experience in performing music, and basic empathy has been controlled. Although rumination correlated with LSMS it was not a significant predictor after the other factors have been controlled. Establishing reliability and validity of the music empathy scale supports the cross-cultural significance of this concept.

Introduction

The question why people are attracted to works of art that evoke negative feelings has attracted attention since the time of Aristotle. One specific context in which this question has been posed is in relation to liking sad music (e. g., Garrido & Schubert, 2013; Huron, 2011; Kawakami, Furukawa, Katahira & Okanoya, 2013; Peltola & Eerola, 2016; Schubert, 2016). While some of the work on this issue concentrated on the paradoxical nature of seeking an experience that evokes negative emotions (Garrido & Schubert, 2015; Schubert, 2016), some researchers claimed music that expresses sadness may evoke positive emotions in the listener (Kawakami et al, 2013). Aside from the reasons for listening to music that evokes sadness, the personality characteristics of individuals who have a preference for sad music has been another focus of research. Personality traits that seem to have strong associations with liking sad music include empathy (Eerola, Vuoskoski & Kautiainen, 2016; Kawakami & Katahire, 2015; Vuoskoski, Thompson, McIlwain & Eerola, 2012), absorption (Garrido & Schubert, 2013), and rumination (Garrio, Eerola & McFerran, 2017; Garrido & Schubert, 2013, 2015).

Although a particular brand of sad music called *Arabesk* has put its stamp on the popular music scene of Turkey for a period extending from the 1970s to the 1990s (Stokes, 1992), preference for sad music, independent of its genre, has not been investigated to this time. The current project was an attempt to investigate the relationship between liking sad music and some variables that have been found important in research in other cultures. Empathy was one of these variables in consideration with its reliable association with liking sad music. However, another variable related to empathy that received emphasis in recent research in this area was also included: This was music empathizing (Garrido & Schubert, 2013; Kreutz, Schubet & Mitchell, 2008).

We took this opportunity to adapt the Music Empathizing (ME) scale for a Turkish population developed by Kreutz and others (2008) for this research project. An already available basic empathy scale (Topçu, Erdur-Baker & Çapa-Aydın, 2010) in Turkish was used in order to assess the relationship between liking sad music and empathy. This scale also served as a check for the criterion validity of the ME scale, because the measures of these conceptually related constructs were expected to be correlated.

Rumination has been found to be related to preferring sad music as well as attracting the attention of researchers in this area because of the mental health implications of using sad music as an instrument that enables pondering on sad events (Garrio, Eerola & McFerran, 2017; Garrido & Schubert, 2013, 2015). We included rumination in the study, because in addition to its theoretical and applied interest, a scale in Turkish was available for measuring this variable. Thus, current research examined the relationships between liking sad music, empathy, music empathizing, and rumination in a Turkish sample.

Method

Participants

One hundred and ninety one participants were recruited through social media to respond to an online questionnaire. One hundred and twenty eight (67%) were women. Two participants (1%) preferred not to specify their gender. One hundred and twenty eight participants (67%) reported they played a musical instrument or performed with voice. The average age of the participants was 30.96 with a standard deviation of 10.26.

Measures

Like Sad Music Scale. The Like Sad Music Scale (LSMS; Garrido & Schubert, 2013) was translated into Turkish. This scale consisted of 11 items that required responses on a 5-point scale (1 = strongly disagree to 5 = strongly agree). Four items of the scale were reverse worded. Garrido and Schubert (2013) reported Cronbach $\alpha = .72$ for the scale.

Music Empathizing Scale. The nine-item Music Empathizing (ME) scale developed by Kreutz, Schubert and Mitchell (2008) was translated into Turkish by the researchers and then it was back translated into English by a native speaker of Turkish who had lived and held academic jobs in English-speaking countries for more than ten years. Four of the nine items were reverse worded. The internal consistency reported by Kreutz and others for their scale was Cronbach α = .69. Participants responded on a four-point scale (1 = strongly disagree to 5 = strongly agree).

Basic Empathy Scale. The Turkish adaptation of the Basic Empathy Scale (BES, Jolliffe & Farrington, 2006) was developed by Topçu and others (2010). This scale consisted of 20 items that were responded to on a 5-point scale (1 = strongly disagree to 5 = strongly agree). BES had an emotional empathy subscale and a cognitive empathy subscale, each one consisting of 10 items. In two applications of the scale on different samples Topçu and others reported Cronbach α = .74 and 76 for emotional empathy and α = .79 and .80 for cognitive empathy. Three items of the emotional empathy subscale and 5 items of the cognitive empathy subscale were reverse coded.

Repetitive Thought Questionnaire. The Turkish adaptation of the Repetitive Thinking Questionnaire (RTQ, McEvoy, Mahoney, & Moulds, 2010) was done by Gülüm and Dağ (2012). In the RTQ respondents are asked to think of the last time something upset them or made them unhappy. After describing the incident they are asked to evaluate how frequently the situation described in 31 statements on a 5-point scale (1 = never to 5 = frequently). Four items were reverse coded.

Procedure

The study was approved by the relevant ethics committee of Uludag University. The study was advertised on social media and those interested were directed to the web address containing the scales. After providing informed consent, participants first responded to questions about age, gender, and musical experience. Then the instructions for each scale was given, followed by the scale content.

Table 1. Means standard deviations and scale reliabilities of LSMS, ME, BES-cognitive, BES-emotional, and RTQ.

	Mean	Standard deviation	Cronbach α
LSMS	3.74	0.69	.84
ME	3.15	0.49	.75
BES-cognitive	4.21	0.51	.83
BES-emotional	3.79	0.55	.78
RTQ	3.34	0.84	.95

Results

Scales and reliability. Means, standard deviation, and Cronbach α coefficients of the scales are given in Table 1. All scales had acceptable internal consistency. As in the work of Garrido and Schubert (2013) the tenth item of LSMS (Whether music is happy or sad has no influence on whether I enjoy listening to it or not) was not found to be informative and reduced the internal consistency of the scale. The reported α value for LSMS is based on the remaining 10 items. Importantly the reliability of the Turkish ME scale was .75.

Correlations between measures. The zero order correlations between pairs of measures are given in Table 2. ME correlated positively with both the cognitive and the emotional dimension of the BES. LSMS correlated positively with ME, both dimensions of BES, and RTQ. Age correlated negatively with LSMS, ME, and RTQ.

Table 2. Correlation coefficients and significance levels among age, LSMS, ME, BES-cognitive, BES-emotional, and RTQ. Significant correlations are given in bold print. The number at the bottom is number of participants.

	LSMS	ME	BES-cognitive	BES-emotional	RTQ
Age	**-.28** < .001 191	**-.19** .010 190	-.01 .878 190	.08 .272 190	**-.20** .006 189
LSMS		**.36** < .001 190	**.17** .020 190	**.20** .005 190	**.25** .001 189
ME			**.26** .001 190	**.29** < .001 190	**.17** .017 189
BES-cognitive				**.48** < .001 190	.12 .112 189
BES-emotional					**.30** < .001 180

The relationship between the scale measures and two categorical variables were investigated by calculating independent samples *t* tests. The two dimensions of BES were the only variables that differed according to gender (see Table 3). Women had both higher cognitive empathy, $t(186) = 3.19$, $p < .005$, and higher emotional empathy, $t(186) = 4.44$, $p < .001$, compared to men. LSMS and ME scores differed according to musical experience. Participants who had played a musical instrument or performed with voice like sad music more, $t(95.88) = 2.26$, $p < .05$, and showed higer music empathizing, $t(103.73) = 3.49$, $p < .001$, compared to those who did not have experience performing music. (Equal variances were not assumed in these last comparisons because Leven's test showed that this assumption was violated, $F(1, 188) = 11.17$, $p < .001$ for LSMS and $F(1, 188) = 5.12$, $p < .05$ for ME.)

Predictors of liking sad music. In order to find out which variables predicted liking sad music a hierarchical regression analysis was performed. LSMS was the predicted variable in this hierarchical regression. In the first step of the analysis age, gender and music experience were included. Cognitive and emotional BES scores were included in the second step. In the last step ME was included. (Entering RTQ either before or after ME did not show RTQ to be a significant predictor of LSMS.)

Table 3. Comparisons of average LSMS, ME, BES-cognitive, BES-emotional, and RTQ for female and male participants. Means that are significantly different are given in bold print.

	Gender	N	Mean	Standard deviation
LSMS	Female	128	3.72	0.72
	Male	61	3.76	0.61
ME	Female	128	3.12	0.50
	Male	60	3.19	0.44
BES-cognitive	Female	128	**4.29**	0.49
	Male	60	**4.05**	0.50
BES-emotional	Female	128	**3.90**	0.53
	Male	60	**3.54**	0.52
RTQ	Female	128	3.34	0.83
	Male	59	3.33	0.89

Table 4. Comparisons of average LSMS, ME, BES-cognitive, BES-emotional, and RTQ for participants with and without experience in playing a musical instrument or singing. Means that are significantly different are given in bold print.

	Music experience	N	Mean	Standard deviation
LSMS	No	63	**3.56**	0.82
	Yes	128	**3.83**	0.60
ME	No	63	**2.97**	0.54
	Yes	127	**3.24**	0.44
BES-cognitive	No	63	4.25	0.54
	Yes	127	4.19	0.50
BES-emotional	No	63	3.89	0.62
	Yes	127	3.74	0.51
RTQ	No	63	3.26	0.93
	Yes	126	3.37	0.79

The results of the hierarchical regression are given in Table 5. Age and musical experience were significant predictors of LSMS at every step of the analysis. Younger participants and those with musical experience liked sad music more. In the second and third steps of the analysis only the emotional dimension of BES was a significant predictor. Emotional basic empathy was positively related to liking sad music after the effects of age and musical instrument had been controlled for but cognitive basic empathy did not make an independent contribution. Participants with higher music empathizing liked sad music more significantly after the other variables have been controlled for.

Discussion

In this study we tried to produce a Turkish version of the ME scale (Kreutz et al, 2008) and to test the power of both ME and rumination as measured by the RTQ (McEvoy, Mahoney, & Moulds, 2010; adapted to Turkish by Gülüm and Dağ, 2012) for predicting liking sad music. The Turkish translation of ME did have acceptable internal consistency and correlated positively with both the cognitive and the emotional dimension of BES (Jolliffe & Farrington, 2006; adapted for Turkish Topçu, et al, 2010). Participants with experience in performing music had higher ME compared to those without musical experience as Kreuzt and others (2008). However, men and women had comparable levels ME although women scored significantly higher on both dimensions of BES. Kreutz and others had observed a gender difference. We did not find a gender difference possibly because our respondents consisted mostly of individuals with experience in music.

Table 5. Results of the hierarchic regression analysis for predicting LSMS. Significant predictors are given in bold print.

Model		Standardized Coefficients Beta	*t*	*p*
1	Age	**-.30**	**-4.27**	**< .001**
	Gender	-.02	-.024	.814
	Playing an instrument or singing	**.20**	**2.85**	**.005**
2	Age	**-.31**	**-4.61**	**< .001**
	Gender	.06	0.91	.362
	Playing an instrument or singing	**-.22**	**-3.27**	**.001**
	BES-cognitive	.08	1.03	.305
	BES-emotional	**.24**	**3.03**	**.003**
3	Age	**-.26**	**-3.90**	**< .001**
	Gender	.03	0.48	.631
	Playing an instrument or singing	**.16**	**2.28**	**.024**
	BES-cognitive	.04	0.60	.552
	BES-emotional	**.17**	**2.16**	**.032**
	ME	**.20**	**2.65**	**.009**

Apart from translating the ME scale to Turkish and establishing its psychometric soundness, we also examined whether music empathizing, basic empathy, and rumination were related to liking sad music in this sample. All of these variables correlated positively with the LSMS. This did replicate existing research (Eerola et al, 2016; Garrido & Schubert, 2011; Kawakami & Katahira, 2015; Vuoskoski & Eerola, 2012; Vuoskoski, Thompson, McIlwain, & Eerola, 2012). As a contribution, we found that of the cognitive and emotional dimensions of basic empathy only emotional empathy was a significant predictor. In addition, we also observed that music empathizing explained a unique share of variance after age, gender, music experience, and empathy have been controlled for.

A further finding in this research was that although repetitive thinking was correlated with liking sad music, the RTQ score did not explain a unique part of variance in the LSMS after age, gender, music experience, and empathy have been controlled for. Rumination has been found to be associated with liking sad music (Garrido et al, 2017; Garrido & Schubert, 2013). However, Garrido and others (2017) associated listening to sad music and rumination in the context of depression and Garrido and Schubert (2013) found rumination to be correlated with some items of the LSMS only. Therefore, it may not be surprising that rumination has a weaker connection with liking sad music.

This was, to our knowledge, the first survey study on correlates of liking sad music in Turkey. The results were mostly consistent with results obtained in other cultures and extended them in certain ways. However, the variables taken into consideration in this project were quite limited and other variables deserve attention. Another path to follow from this point would be to include situational and interactional aspects of using music in order to better understand the fascination with sad music.

Acknowledgements. We would like to thank Suzan Alptekin for translation from Turkish to English, and Deniz Bilger and Kenan Alpaslan for conducting the internet searches for the scales.

References

Eerola, T. Vouskosky, J. T., & Kautiainen, H. (2016). Being moved by unfamiliar sad music is associated with high empathy. *Frontiers in Psychology*.

Garrido, S., Eerola, T. & McFerran, K. (2017). Group rumination: Social interactions around music in people with depression. *Frontiers in Psychology, 8*. doi: 10.3389/fpsyg.2017.00490

Garrido, S. & Schubert, E. (2011). Individual differences in the enjoyment of negative emotions in music. *Music Perception, 28*, 279-295.

Garrido, S. & Schubert, E. (2013). Adaptive and maladaptive attractions to negative emotions in music. *Musicae Scientiae, 17*, 147-166.

Garrido, S. & Schubert, E. (2015). Moody melodies: Do they cheer us up? A study on the effects of sad music on mood. *Psychology of Music, 43*, 244-261.

Gülüm, I. V. & Dağ, I. (2012). The Turkish adaptation, validity and reliability study of the Repetitive Thinking Questionnaire and the Cognitive Flexibility Inventory. *Anatolian Journal of Psychiatry, 13*, 216-223.

Huron, D. (2011). Why is sad music pleasurable? A possible role for prolactin. *Musicae Scientiae, 15*, 146-158.

Kawakami, A. & Katahira, K. (2015). Influence of trait empathy on the emotion evoked by sad music and on the preference for it. *Frontiers in Psychology, 6*. doi: 10.3389/fpsyg.2015.01541

Kawakami, A., Furukawa, K., Katahira, K. & Okanoya, K. (2013). Sad music induces pleasant emotion. *Frontiers in Psychology, 4*. doi: 10.3389/fpsyg.2013.00311

Kreutz, G., Schubert, E., & Mitchell, L. A. (2008). Cognitive styles of music listening. *Music Perception, 26*, 57–73.

Peltola, H.-R. & Eerola, T. (2016). Fifty shades of blue: Classification of music-evoked sadness. *Musicae Scientiae, 20*, 84-102.

Schubert, E. (2016). Enjoying sad music: Paradox or parallel process? *Frontiers in Human Neuroscience, 10*. doi: 10.3389/fnhum.2016.00312

Stokes, M. (1992). *The Arabesk debate: Music and musicians inmodern Turkey*. Oxford: Oxford University Press.

Topcu, Ç., Erdur-Baker, Ö. & Çapa-Aydın, Y. (2010). Turkish Adaptation of Basic Empathy Scale: Validity and Reliability Study. *Türk Psikolojik Danışma ve Rehberlik Dergisi, 4*(34), 174-182.

Vuoskoski, J. K. ve Eerola, T. (2012). Can sad music really make you sad? Indirect measures of affective states induced by music and autobiographical memories. Psychology of Aesthetics, Creativity, and the Arts, 6, 204-213. DOI: 10.1037/a0026937

Vuoskoski, J. K., Thompson, W. F., McIlwain, D. ve Eerola, T. (2012). Who enjoys listening tos ad music and why? *Music Perception, 29*, 311-317. DOI:10.1525/mp.2012.29.3.311

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

The role of motivation to the quality of attention in deliberate practice

Eduardo de Carvalho Torres

Federal University of Rio de Janeiro, Brazil

contato.eduardotorres@gmail.com

Abstract

Researchers into musical performance and expertise have emphasized the importance of motivation and attention for the so-called deliberate practice, neglecting, however, the discussion about the conditions that favor that state of primacy. Understanding this to be such an important condition for the quality of deliberate practice, why not discuss it with greater depth?

In my research, I try to demonstrate the relations between the performer's motivational states and the attentional capacity. According to the current theoretical framework, the evidence suggests that the quality of attention is, at least in part, regulated by the subjects' levels of motivation. My theoretical references considered authors as Ericsson, Krampe & Tesch-Römer (1993), Reeve (2006), Csikszentmihaly (1997), Styles (2006), Eysenck & Keane (2010), Baddeley, Eysenck & Anderson (2015), Cohen (2014), among others.

To test my proposition, I developed a small experimental program executed with piano students of different levels so that I could observe, among other things, their routines of study, habits, tastes, personal and professional histories, aspirations, from where it was possible to testify some statements. The results showed that there is a clear relationship between the individuals' motivational states and their respective attentional abilities. Besides, certain levels of motivation induce a greater capacity for focus, suggesting that there is a need to broadening discussions about updating curricula of the formation of instrumentalists, particularly of pianists. It is essential to the pedagogy of the instrument to prepare the students to overcome the difficulties related to the inefficacy of attention, aiming at higher quality in the deliberate practice.

Introduction

Musical performance is a human activity known to be complex and for which years of intense and uninterrupted daily practice are required. A few years ago researchers have identified that this practice is not a purely recreational, uncompromising activity, but a practice focused on improving continuously acquired skills, such as the ability to play the piano, for example. It is interesting to note that this practice, at the most diverse levels, that cross individual capacity, resources and the interest of the individual, leads to very different results, even when performed by individuals with very similar conditions. This finding motivated my initial question about the reasons that would make an individual more apt to perform certain tasks when compared to other individuals under similar conditions. And a second question has broadened the previous one, trying to understand how difficulties and inconsistencies of productive behavior could be neutralized and overcome, in order to favor the achievement of intended practical objectives?

In my initial study, conducted in the master's program of the Federal University of Rio de Janeiro (Brazil), I assigned the individual motivation as a factor of greater relevance. My initial assumptions indicated that the individual's degree of motivation would be sufficient to explain the gap between skills acquired by different people under similar conditions, and I concluded, synthetically, that this position is valid. However, it was necessary to understand minimally the factors that interfered with this individual motivation. Verifying that the so-called deliberate practice was regulated by levels of motivation of the individual, but that motivation alone wasn't enough to explain the quality of this practice, I noted that it was a superficial approach by part of the studies about deliberate practice (at least in Brazil) in music performance in relation to the discussion about the necessity of the capacity of individual's attention, and more precisely, the individual's concentration.

From this recognition I began to construct a line of reasoning that considers attention as a primary element for the quality of realization of the deliberate practice considering that this attention is moderated by the levels of motivation of the subject, completing the framework of my hypothesis, which I present here as a result of this research work.

To investigate this hypothesis, I developed a small program of experiments with students of a piano class who volunteered to collaborate. This small program of experiments considered five instruments of data collection, namely: a) observation and audiovisual recording of sessions of a collective discipline of these participants, held in the second half of 2016; b) a semi-structured individual interview, held in the first half of 2017; c) recording audio of the daily practice of each participant over a week of studies; d) daily practice report, to be completed by the participant at the end of each study session; and e) a conclusive and reflective question, formulated to each participant at the end of the week of recorded practice about the work done during the week of audio recording. The result of this data collection showed that all the participants reported some kind of factor related to the ineffectiveness of the attention.

Musical performance and deliberate practice

According to Barry and Hallam (2002), "practice is necessary to enable musicians to acquire, develop, and maintain aspects of technique, learn new music, memorize music for performance, develop the interpretation, and prepare performance". This statement illustrates what every musician empirically knows: that practice plays an essential role in the construction of performance. In addition, the authors point out that the term deliberate practice has been adopted to "specify the type of practice associated with the development of expert skills", emphasizing that "deliberate practice is highly focused and requires great effort and concentration" (p.155).

On the deliberate practice, Ericsson, Krampe and Tesch-Römer (1993) characterize it essentially as a set of practical activities developed by teachers and coaches "that the individual can engage in between meetings with the teacher". These highly structured tasks aim to develop a specific skill, and for such characteristics, they differ from a mere practical activity. Still according to these authors, the deliberate practice "requires effort and is not inherently enjoyable" considering that "individuals are motivated to practice because practice improves performance". These authors, therefore, consider that individuals engage in deliberate practice, not for immediate results, but for long-term outcomes. In addition to the characteristics mentioned in the previous paragraph, Ericsson, Krampe and Tesch-Römer emphasize the importance of "focused attention required by deliberate practice" (p. 368). According to these authors, "it is necessary to maintain full attention during the entire period of deliberate practice" (p. 370). Some Brazilian scholars on the subject, notably Galvão (2006) and Alves and Freire (2014), adopting this same work by Ericsson, Krampe and Tesch-Römer, incorporate the considerations of this reference work in their studies on expertise, about the need for focused attention, considering that "the expert needs to develop motivational structures in order to sustain the focus on long-term deliberate practice" (p. 63), stating, in other words, that deliberate practice aiming at expertise is unachievable without motivation.

What is striking about the statements made about attention and motivation are not the statements themselves, but the superficiality with which the subject is being treated by scholars who study deliberate practice and musical performance in general in relation to the scientific basis, especially the capacity of attention. There seems to be no deepening of how this phenomenon is so necessary to deliberate practice, what are its causes and consequences, and what are the means to overcome the difficulties related to itself.

Attention and motivation

Attention has been studied objectively, especially since the mid-nineteenth century, with William James as one of his pioneers. Generally, almost all attention theorists agree that this is an individual's ability to select, respond to particular stimuli, and disregard others (Pashler, 1998; Styles, 2006; Eysenck and Keane, 2010).

According to Styles (2006), attention "is not a single concept, but an umbrella term for a variety of psychological phenomena" (p. 1), illustrating the various approaches that can be found on the subject. Agreeing with Styles, Cohen (2014) considers the attention phenomena "a cognitive experience that is self-evident, yet difficult to characterize" (p. 13).

Castle and Buckler (2009), on the other hand, consider that attention can be observed by at least three perspectives: filter, focus or source. These same authors distinguish between the concepts of attention and concentration (mental), usually treated as synonyms. For these authors, the attention can be divided into "selective" and "divided", the first being (selective) relative to "focused attention", or concentration. Thus "divided" would become attention in the broadest sense, considering that it is possible for the individual to be attentive without necessarily being fully focused on what he does, which is understandable, whereas if we were always fully aware of everything, our memory capabilities would be seriously compromised.

The need to relate memory to the individual's attentional capacity is clearly a clue to understanding some inconsistencies with the difficulties of focusing on deliberate practice. Baddeley, Eysenck and Anderson (2015) argue that "our knowledge of the world stored in long-term memory can influence our focus of attention, which will then determine what is fed into the sensory memory systems, how this is processed and how it will be evoked later" (p. 18), but rarely studies on deliberate practice, at least in the field of music, establish a relationship under this approach.

Another clue, where attention and motivation are related, is given in the same works previously quoted here in this text, when, for example, Alves and Freire (2014) argue for the expert's need to develop resources for motivation as a means of maintaining concentration on the deliberate practice, as stated by Ericsson, Krampe and Tesch-Römer.

To Csikszentmihaly (1997), how far apart the motivational and emotional affinities of an individual in relation to an object or a situation, the more painful the ability to focus the attention of the individual. Conversely, when the individual is interested in the task he is engaging in, when he is motivated to overcome the difficulties imposed by the task, the more easily concentration occurs. According to this author, when abilities and challenges are in balance, the individual experiences a state of pleasure and total concentration, a state that the author called "flow".

Recapitulation and Synthesis

Recapitulating what has been said so far, we can observe that, first, practice is indispensable for the improvement of the performance musician in its totality and that there is a modality of practice, in particular, called deliberate practice, whose purpose is the certain skills in particular. This practice is not pleasurable, and what moves individuals to their realization is an external motivation, determined by the reward that the individual achieves by accomplishing the intended goals. Furthermore, it is unanimous to consider that, in order to carry out this practice, it is necessary for the individual to be fully concentrated, fully attentive, although in general scholars do not explain exactly how this condition can be achieved, or what are the characteristics of this condition.

Attention scholars explain that this ability is not easily observable, and therefore difficult to determine its constituent parts. However, they note that attention is primarily given to the skills of selecting, attending or neglecting stimuli. In addition, the attention can be more or less concentrated, being also limited by the storage capacity guaranteed by the memory. Scholars of expertise, in turn, consider it indispensable to the ability to develop means for motivation in order to maintain attentive focus on deliberate practice. Finally, it is known that the greater the affectivity of the individual in relation to someone, to an object or situation, the greater the motivation to focus on them.

From these considerations taken from the theoretical framework briefly presented in this work, the result of a more comprehensive literature review, I was able to conclude that, at least in part, the individual's motivation moderates the subject's ability to focus his attention on certain tasks, such as those required by deliberate practice. This relation is reciprocal, that is, attention also motivates the subject, insofar as the subject perceives the observable results in the quality of realization in the deliberate practice by the obtained results. The operation of this system can be observed in figure 1.

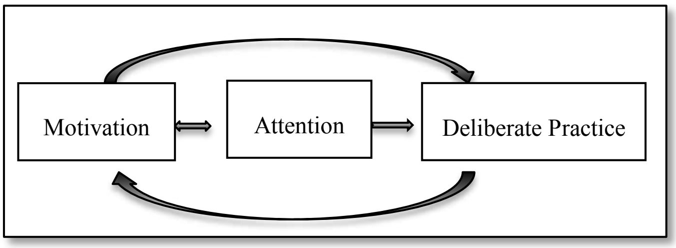


Figure 1. Relationship between motivation and attention in relation to deliberate practice

In order to consolidate this hypothesis, I developed a small experimental program, to be detailed in the next topic, which took place with four piano students, aged between 18 and 23 years, with different profiles in all aspects (psychological, gender, socioeconomic, etc.), residents of the city and state of Rio de Janeiro (Brazil) between 2016 and 2017.

Methodology

The application of this experiment had the collaboration and guidance of a professor of the piano course of the School of Music of the Federal University of Rio de Janeiro, and its student participants, whose names will remain hidden.

Before starting the data collection process, I introduced myself to the professor, clarifying my proposal. In the first part of my data collection, with the permission of the professor and his students, I accompanied some classes of a collective discipline called pianistic practice accompanied by a camcorder. This course, entirely integrated by students of the piano course, consists of playing to others part of the repertoire that is being studied so that everyone can appreciate and present suggestions of all kinds (interpretative, technical, etc). A total of six sessions of observations were carried out in classes between October and December 2016. The objective here was to observe the behavior of the students in a performance situation before an external spectator and to know the profile of the participants. The recordings always took place in the same classroom, at the same time and on the same day of the week, Wednesdays between 9 and 11 o'clock in the morning. This stage generated 9 hours, 27 minutes and 12 seconds of audiovisual records. The equipment was a Finepix S2800 HD Series, of the label FUJIFILM.

My second instrument was a semi-structured interview, held individually with each student. This interview consisted in knowing biographical data of the individuals that could contribute to their performances, motivations, tastes, practices, habits, and all the data that could aid in the later observations. These interviews were conducted in April 2017 in the same building where the audiovisual recordings took place, but in different rooms, days, times and conditions. To guide this interview, I used Irving Seidman's Interviewing as *Qualitative Research: A Guide for Researchers in Education and the Social Sciences* (2016) as a reference. This step resulted in about 180 minutes of recordings.

The third step consisted of records of the individual practice sessions. The objective was to observe facts that were stated by the participants during the interview, the study characteristics of each one, the study environment, and other data that could be considered relevant to understand the conditions in which the practices of each were given. The recordings were made by the students themselves in their study sections. The equipment used was the same used for the interviews: a ZOOM H1 digital recorder, accompanied by AA rechargeable batteries of the SONY label and its respective charger. From this stage, there were 26 sessions, totaling about 23 hours of recordings.

In addition to the tape recorder, I delivered a practice report called *relatório de prática,* (practice report), which was intended to prompt participants to fill in the situation data at the end of each practice session, avoiding problems of forgetting. The objective here was to cross what was informed by the participant with what was observed in the recordings of practices, interviews, what was observed in the room, etc.

Finally, at the end of each week of sound records, when searching for the recorder with the student, he performed a single question, also recorded in audio. The question was: "Did the work routine and recording procedures of your study activities during the week make you think better about some of the issues we discussed earlier?". These recordings, made with the same ZOOM recorder, totaled about 4 minutes in total. The objective was for the participants to report facts that they considered relevant, based on their reflections.

Results and discussion

First, I listened to the interview audiences (before and after the study week), and I wrote down the most relevant data, aiming to draw a pattern of responses, convergent data and the profile of each one. Then I analyzed the practice reports, crossing the information extracted with those observed in the audios. After that, I observed the audiovisual recordings in order to observe facts that could corroborate with the observations obtained until then.

In this work of data collection I sought to observe which behaviors and statements of the students were perceptible within the theoretical framework constituted by the theories of attention and motivation, besides the theories of memory. For practical reasons, these models and theories were not incorporated in the present paper but briefly referenced, since in this work I try to evidence the hypothesis and general results of the work, not the theories. However, for guidance, I will explain below some of the theories reviewed during the research work:

a) On the basic functioning of memory, I took as main reference the work of Baddeley, Eysenck and Anderson (2015). The types of memory studied were: sensory, short-term memory, working memory and long-term memory.

b) Regarding motivation, my basic reference was the work of Reeve (2006). The theories studied were: achievement motivation, cognitive dissonance, flow, goal setting, learned helplessness, reactance, self-efficacy, self-determination, self-regulation and value expectancy.

c) In the case of attention, it was more complex. I have taken into account several works as references, especially the publications of Pashler (1998), Styles (2006), Eysenck and Keane (2010) and Cohen (2014). Of these authors, I highlighted varieties, manifestations and limitations of attention, in addition to Cohen's "types and conditions" (p. 4), among which: selective attention, divided attention, focused attention, sustained attention, and voluntary attention. In addition, I adopted in my observations the elements associated with the ineffectiveness of attention, such as inattention, fatigue, distraction, confusion, lack of persistence, neglect or lack of control.

In concluding the reading of practice reports, observation of audiovisual sessions and listening to audios, including interviews, practice sessions, and the concluding question, I made brief individual reports on each student, seeking to find parallels with the theoretical models adopted in this work.

Regarding the first student, I found that the participant was in the main part of his practice, concentrated, even when with external interferences, like noises, for example. It was also evident that the participant is guided by a goal when trying to study every day, as he had reported in the interview. His observations about himself are consistent, which suggests that his study is self-regulated. Just once the participant did report being deconcentrated, when he reported being tired, which is considered an element related to the ineffectiveness of attention.

With the second student, it was possible to verify that the participant sought to be consistent with his declared goals in the interview. In almost all of his practice sessions, he evidenced some element related to the inefficacy of attention, such as tiredness, noises, and pains in the wrist of the arm. The study of this participant appeared to be at least partly extrinsically motivated in an introjected way. He presented demotivation for being studying the same repertoire for an extended period.

Concerning the third, I observed that the time dedicated to the practice reported by the participant was not consistent with that observed in his (practice) sessions. This participant pointed out the "lack of practice" as a difficulty because the participant traveled in the week of his records. In general, it was possible to observe that this participant adopts a position of strategic allocation of time in the base of the difficulty, since its practice is in general, fractioned throughout the day.

About the fourth participant, she studied throughout the week as she had reported in the interview, although she did not strictly follow the schedules she had previously reported. Her study instrument has mechanical problems, as she had also stated. Throughout the week she stated that she was almost always "calm," and mentioned "lack of concentration" or "concentration during only part of the time," as she had previously stated. Almost always the excess noise declared by the participant was actually recorded in his practice sections, which indicates that these interferences affect, in part, the routine of the participant, but not always. There is also a motivational factor that was present. Her motivation is mostly integrated extrinsic, as had been verified in interviews.

After these steps of collecting data and observing the results of the data obtained, I began to interpret them. From the outset, I compared the ages of the participants to the necessary ten thousand hours of study made explicit by Ericsson, Krampe, and Tesch-Römer (1993) for the acquisition of expertise. It is clear that this is a reference value, but it is already possible through this reference to have an idea of reality. In the comparison, I found that none of the participants (who were between 18 and 23 years old) fit the level of experts according to the model of the 10 thousand hours, depending on the approximate time of study that was dedicated daily added to the years of study that already had on their instrument, the piano. However, it was found that the first participant was closest to this number, coinciding with the fact that he was the one who started his studies earlier, who reported that he studied more during the week and in more favorable conditions, which studied more concentrated, which stated to have goals (motivation), to be older and to present more positive academic results. That is, there is at least evidence that the deliberate practice of itself works efficiently.

When asked about their difficulties, almost all participants reported: time, back pain, tendonitis, pain in the hands, posture, technical difficulties, tiredness, lack of concentration and environmental issues, all of these elements related to the ineffectiveness of attention. Concerning what most favored concentration, participants reported focusing better during piano lessons. In addition, all of them, except for the first participant, reported being deconcentrated by external noises in their practices, a fact confirmed by listening to the audios, which suggests that the quality of their practices is, at least in part, linked to environmental issues that go beyond the reach of the individual. Concerning their motivations to join the piano course, the participants reported the support of family, friends, and teachers, as well as the repertoire studied at the academy.

One of the interesting observations highlighted in the interviews was the fact that the participants reported not being motivated by the job market, which suggests that everyone has some degree of intrinsic motivation. However, the motivation was not generally sufficient to guarantee a good performance of the practical work, coinciding with the various elements related to the ineffectiveness of attention observed, both in the interviews and in the audits, a fact that alerts to the health conditions of the students, the psychological and material conditions for their individual practices—almost always carried out in inadequate conditions.

It was also evident that the training curricula of performance musicians, particularly of pianists, at least in Brazil, need updating, since, as is evident, attention is not purely a congenital ability or a natural consequence of repertoire study, but the result of a series of constraints that need a systematic study to identify them.

Conclusion

The results of this work were surprising, since all participants reported to a greater or lesser degree, some difficulty of concentration, for reasons such as fatigue and external factors, such as noise, for example, and the consequence of the ineffectiveness of attention in these cases was determinant for the quality of the practice of individuals, including focusing on their motivation to perform the proposed tasks. However, when highly motivated, as in the case of goals, they were able to overcome the obstacles and remain focused, as observed in the reports. It is concluded that the deliberate practice is, in an ideal situation, carried out in a quiet environment, that the practitioner is calm, and that his practices are balanced with his difficulties.

Acknowledgments. Thank you to my advisor, Prof. Marcos Nogueira, for the partnership. I also thank Prof. Midori Maeshiro for the contributions, to the students that collaborate with my research, to my companion Ana Luiza, to CAPES, and to all those who somehow collaborated in the elaboration of this work.

References

Alves, A. C. & Freire, R. D. (2014). Processos de construção da expertise na clarineta: Investigação das trajetórias de clarinetistas brasileiros. [Processes of the construction of expertise in clarinet: Investigation of the trajectories of Brazilian clarinetists.] Percepta, v. 2, n.1, 61–84.

Baddeley, A., Anderson, M. C. & Eysenck, M. W. (2015). Memory. (2nd ed.) New York: Psychology Press.

Barry, N. H. & Hallam, S. Practice. (2002). Practice. In: Parncutt, R. & Mcpherson, G. E. The science and psychology of music performance: Creative strategies for teaching and learning. (pp. 151–165). New York: Oxford University Press

Castle, P. & Buckler, S. (2009). What I was saying? Concentration and attention. In: Castle, Paul & Buckler, Scott. How to be a successful teacher: Strategies for personal and professional development. (pp. 12–32). Sage Publications.

Cohen, Ronald. (2014). The Neuropsychology of Attention (2nd ed.) New York: Springer.

Csikszentmihaly, M. (1997). Finding flow: The psychology of engagement with everyday life. New York: Basic Books.

Ericsson, K., Krampe, R. & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100 (3), 363–406.

Eysenck, M. & Keane, M. (2010). Cognitive Psychology: A student's handbook. (6th ed.). New York: Psychology Press.

Galvão, Afonso. (2006). Cognição, emoção e expertise musical. [Cognition, emotion and musical expertise] Psicologia: Teoria e Pesquisa, 22 (2), 169–174.

Pashler, H. (1998). The psychology of attention. Cambridge, MA: MIT Press.

Styles, Elizabeth. (2006). The Psychology of Attention. (2ª ed.). New York: Psychology Press.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

The Importance of Song Context and Song Order in Automated Music Playlist Generation

Andreu Vall,[1] Massimo Quadrana,[2] Markus Schedl,[1] and Gerhard Widmer[1]

[1] *Department of Computational Perception, Johannes Kepler University, Austria*
[2] *Pandora Media, Inc., CA, U.S.A.*

andreu.vall@jku.at

Abstract

The automated generation of music playlists can be naturally regarded as a sequential task, where a recommender system suggests a stream of songs that constitute a listening session. In order to predict the next song in a playlist, some of the playlist models proposed so far consider the current and previous songs in the playlist (i.e., the song context) and possibly the order of the songs in the playlist. We investigate the impact of the song context and the song order on next-song recommendations by conducting dedicated off-line experiments on two datasets of hand-curated music playlists. Firstly, we compare three playlist models, each able to consider a different song context length: a popularity-based model, a song-based Collaborative Filtering (CF) model and a Recurrent-Neural-Network-based model (RNN). We also consider a model that predicts next songs at random as a reference. Secondly, we challenge the RNN model (the only model from the first experiment able to consider the song order) by manipulating the order of songs within playlists. Our results indicate that the song context has a positive impact on the quality of next-song recommendations, even though this effect can be masked by the bias towards very popular songs. Furthermore, in our experiments the song order does not appear as a crucial variable to predict better next-song recommendations.

Introduction

Automated music playlist generation is a specific task in music recommender systems in which the user receives a list of song suggestions that constitute a listening session, usually listened to in the given order. This is in contrast to the browsing scenario, in which users receive a collection of recommendations and actively choose their preferred option.

According to interviews with practitioners and postings to a dedicated playlist-sharing website, the choice of songs and their order–or at least their relative position–have been identified as important aspects when compiling a playlist (Cunningham et al., 2006). Although some approaches to playlist generation take the previous songs in the playlist (i.e., the song context) and the song order into consideration, to the best of our knowledge, they do not explicitly analyze the importance of these variables.

Modeling Music Playlists

In this section we describe the models we use for automated music playlist generation. We adopt the following approach for every model. Two disjoint sets of playlists are available, one for training and one for test, such that all the songs in the test playlist also occur in the training playlists. Hyperparameter tuning, if necessary, is performed on a validation split that is withheld from the training set. Given one or several songs from a test playlist, a trained playlist model has to be able to rank all the candidate songs according to how likely they are to be the next song in the playlist.

Song Popularity

This model computes the frequency of each song in the training playlists. At test time, the candidate songs are ranked according to their frequency. Thus, the predictions of this model (equivalent to a unigram model–see e.g., Manning & Schütze (2000)) are independent of the current song.

Song-Based Collaborative Filtering

This is an item-based Collaborative Filtering (CF) model. A song s is represented by the binary vector $\mathbf{p}_s$ indicating the playlists to which it belongs. The similarity of each pair of songs s_i, s_j in the training set is computed as the cosine between $\mathbf{p}_{si}$ and $\mathbf{p}_{sj}$. At test time, the next-song candidates are ranked according to their similarity to the current song, but previous songs in the playlist are ignored.

Recurrent Neural Networks

Recurrent Neural Networks (RNNs) are a class of neural network models particularly suited to processing sequential data. They have a hidden state that accounts for the input at each time step while recurrently incorporating information from previous hidden states.

We adopt the approach proposed by Hidasi et al. (2016), where an RNN model with one layer of gated recurrent units is combined with a loss function designed to optimize the ranking of next-item recommendations. At test time, given the current and all the previous songs in the playlist, the RNN outputs a vector of song scores that is used to rank the next-song candidates.

Playlist Datasets

The "AotM-2011" dataset (McFee & Lanckriet 2012) is a collection of playlists derived from the playlist-sharing platform Art of the Mix (www.artofthemix.org). Each playlist is represented by song titles and artist names, linked to the corresponding identifiers of the Million Song Dataset (MSD) (Bertin-Mahieux et al. 2011), where available.

The "8tracks" dataset is a private collection of playlists derived from 8tracks (https://8tracks.com), an on-line platform where users can share playlists and listen to playlists other users prepared. Each playlist is represented by song titles and artist names. Since there are many different spellings for the same song-artist pairs, we mimic the AotM-2011 dataset and use fuzzy string matching to resolve the song titles and artist names against the MSD.

A considerable number of playlists in the AotM-2011 contain songs by one or very few artists. In order to study more diverse playlists (which we assume to correspond to a more careful compilation process), we keep only the playlists with at least 3 unique artists and with a maximum of 2 songs per artist. Although the 8tracks dataset is not affected by this issue (the terms of use of the 8tracks platform require that no more than 2 songs from the same artist or album may be included in a playlist), we apply the same filters for the sake of consistency. Furthermore, we keep only the playlists with at least 5 songs. This ensures a minimum playlist length, that is required to study the effect of the song position on model performance. Finally, songs occurring in less than 10 playlists are removed to ensure that the models have sufficient observations for each song.

We randomly assign 80% of the playlists to the training set and the remaining 20% to the test set. Note that full playlists are assigned to either split. At test time, the model deals with playlists that were never seen before. As in any recommendation task blind to item content, the songs that only occur in test playlists need to be removed because they can not be modeled at training time.

The filtered AotM-2011 dataset includes 17,178 playlists with 7,032 songs by 2,208 artists. The filtered 8tracks dataset has 76,759 playlists with 15,649 songs by 4,290 artists.

Evaluation of Playlist Models

A trained playlist model is evaluated by repeating the following procedure over all the test playlists. We show the model the first song in a playlist. It then ranks all the candidate songs according to their likelihood to be the second song in that playlist. We keep track of the rank assigned to the actual second song and of the fact that this was a prediction for a song in second position. We then show the model the first and the second actual songs. The model has to rank all the candidate songs for the third position, having now more context. In this way, we progress until the end of the playlist, always keeping track of the rank assigned to the actual next song and the position in the playlist for which the prediction is made.

A perfect model would always rank the actual next song in the first position. A random model would, on average, rank the actual next song approximately in the middle of the list of song candidates. An extremely poor model would rank the actual next song in the last position. Note that the actual rank values depend on the number of candidate songs available.

Previous research has often summarized the ranking results in terms of recall at *K*, where *K* is the length of the list of top next recommendations (see e.g., Hariri et al. (2012), Bonnin & Jannach (2014), Hidasi et al. (2016)). However, the proposed evaluation setting may be too pessimistic in the music domain (Platt et al. 2002, McFee & Lanckriet 2011), where songs other than the actual one may serve as valid playlist continuations. As a consequence, long lists of next-song candidates are needed to observe the model behavior. In order to better observe the performance of each model, we opt for analyzing the full distribution of predicted ranks, summarized by the first quartile, the median and the third quartile rank values (Figures 1 and 2). This approach also facilitates the comparison of different models.

Song Context

Figure 1 displays the ranks attained by the actual next songs in the test playlists, given the predictions of the considered playlist models. The distributions of attained ranks are split by the position in the playlist for which the next-song prediction is made. The popularity-based model and the song-based CF model, which have no context and a context of 1 song, respectively, do not improve their predictions as they progress through the playlists. The RNN model, which is aware of the full song context, improves its performance as it progresses through the playlist.

Regarding the absolute model performance, it is worth noting that the popularity-based model and the RNN model show comparable overall performances, despite the fact that the RNN model is much more complex. We could explain this apparent contradiction in our previous work (Vall et al. 2017) as an effect of the bias towards popular songs, common in the music domain. We found that the popularity-based model performs outstandingly well on the most popular songs, but performs poorly on the infrequent songs. On the other hand, the performance of the RNN model is unaffected by the song popularity.

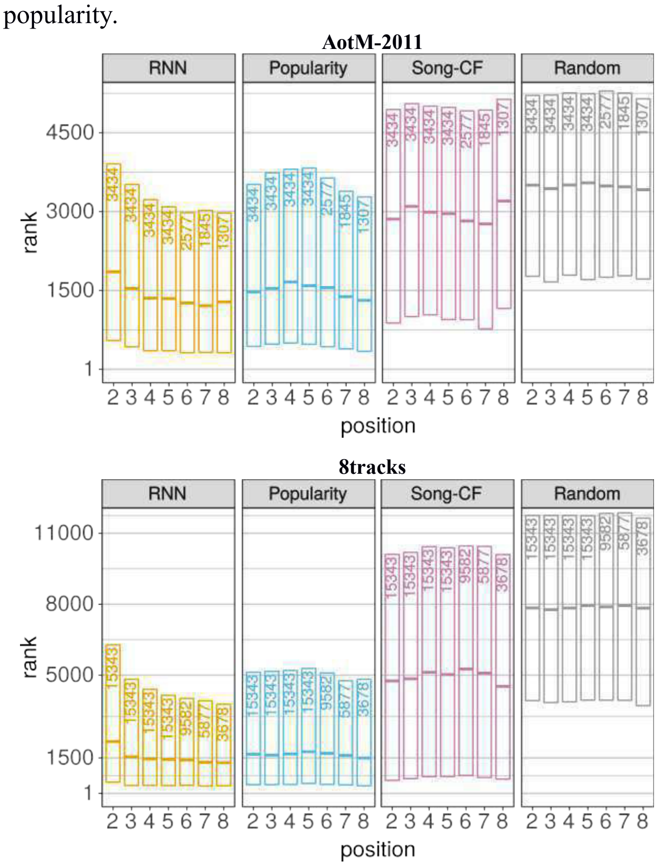


Figure 1. Distribution of ranks attained by the actual next songs in the test playlists (lower is better) for the AotM-2011 (top) and the 8tracks (bottom) datasets. Each panel corresponds to a playlist continuation model. The *x* axis indicates the position in the playlist for which a prediction is made. The *y* axis indicates the attained ranks and its scale relates to the number of songs in each dataset. The boxplots summarize the distribution of attained ranks by their first quartile, median and third quartile values. The number of rank values at every position is annotated.

Song Order

We consider three song order manipulation experiments. For the first experiment we train the RNN model on original playlists, but we evaluate it on shuffled playlists (we refer to this setting as "shuffled test"). For the second experiment we train the RNN model on shuffled playlists and evaluate it on original playlists (we refer to this setting as "shuffled training"). Finally, we train and evaluate the RNN model on shuffled playlists (we refer to this setting as "shuffled training and test"). Figure 2 displays the ranks attained by the actual next songs in the test playlists, given the predictions of the RNN model under the different song order randomization experiments. The distributions of attained ranks are split by the position in the playlist for which the next-song prediction is made. The performance of the RNN model trained and evaluated on original playlists is kept as a reference.

The performance of the RNN model is comparable for all the song order randomization experiments, regardless of whether the song order is maintained, broken at test time or broken at training time. This result suggests that the song order may not be a crucial variable for automated music playlist generation. Even though we considered a competitive RNN model, further investigation on order-aware models is still required.

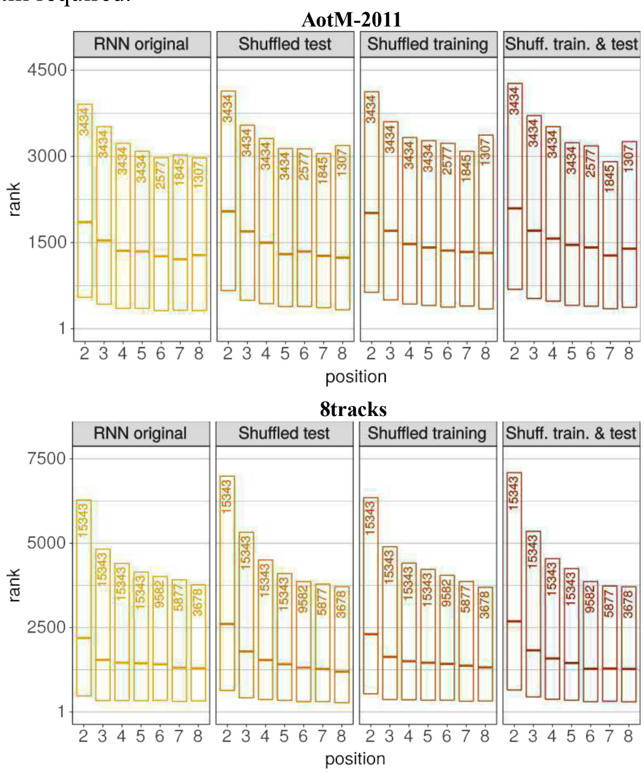


Figure 2. Distribution of ranks attained by the actual next songs in the test playlists (lower is better) for the AotM-2011 (top) and the 8tracks (bottom) datasets. The panels include the predictions of the RNN on the original playlists and on the song order randomization experiments. The *x* axis indicates the position in the playlist for which a prediction is made. The *y* axis indicates the attained ranks and its scale relates to the number of songs in each dataset. The boxplots summarize the distribution of attained ranks by their first quartile, median and third quartile values. The number of rank values at every position is annotated.

Conclusion

In this work we explicitly analyzed the importance of considering the song context and the song order for automated music playlist generation. We conducted off-line experiments in two datasets of hand-curated music playlists, where we compared different playlist models with different capabilities. Our results indicate that the song context has a positive impact on next-song recommendations. Still, as we observed in previous works, the bias towards populars songs can mask the importance of considering the song context. On the other hand, the song order did not appear as a relevant variable to predict better next-song recommendations.

Acknowledgements. The authors want to thank Matthias Dorfer, Bruce Ferwerda, Rainer Kelz, Filip Korzeniowski, Rocío del Río and David Sears for helpful discussions. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 670035 (Con Espressione).

References

Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The million song dataset. In Proceedings ISMIR.

Bonnin, G., & Jannach, D. (2014). Automated generation of music playlists: Survey and experiments. ACM Computing Surveys, 47(2), 1–35.

Cunningham, S. J., Bainbridge, D., & Falconer, A. (2006). "More of an art than a science": Supporting the creation of playlists and mixes. In Proceedings ISMIR.

Hariri, N., Mobasher, B., & Burke, R. (2012). Context-aware music recommendation based on latent topic sequential patterns. In Proceedings RecSys.

Hidasi, B., Karatzoglou, A., Baltrunas, L., & Tikk, D. (2016). Session-based recommendations with recurrent neural networks. In Proceedings ICLR.

Manning, C. D., & Schütze, H. (2000). Foundations of statistical natural language processing. MIT Press.

McFee, B., & Lanckriet, G. R. (2011). The natural language of playlists. In Proceedings ISMIR.

McFee, B., & Lanckriet, G. R. (2012). Hypergraph models of playlist dialects. In Proceedings ISMIR.

Platt, J. C., Burges, C. J., Swenson, S., Weare, C., & Zheng, A. (2002). Learning a Gaussian process prior for automatically generating music playlists. In Proceedings NIPS.

Vall, A., Schedl, M., Widmer, G., Quadrana, M., & Cremonesi, P. (2017). The importance of song context in music playlists. In RecSys 2017 Poster Proceedings.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Relation between melodic characteristics and tempo determination

Leigh VanHandel,[1] Audrey Drotos,[2] and J. Devin McAuley[2,3]

[1] *College of Music, Michigan State University, USA*
[2] *Neuroscience Program, Michigan State University, USA*
[3] *Department of Psychology, Michigan State University, USA*

[1]lvh@msu.edu, [2]drotosau@msu.edu, [3]dmcauley@msu.edu

Abstract

Research on tempo has typically focused on listeners' ability to discriminate tempo, the relative vs. absolute nature of tempo memory, and individual differences in preferred tempo. Much less research has investigated tempo determination, or the process of actively determining an appropriate tempo for an unfamiliar musical excerpt. Here, two experiments investigated whether melodic characteristics provide contextual cues for determining the appropriate tempo of an unfamiliar melody. In both experiments, participants listened to continuous (looped) presentations of unfamiliar melodies and adjusted the tempo in real time to decide on an appropriate final tempo. Experiments 1 and 2 differed only in that in addition to recording final tempo, Experiment 2 recorded continuous tempo tracks generated by each participant for each melody. Final selected tempos were generally slower for melodies with a larger pitch range, more non-zero pitch intervals, more contour changes, and those with larger (and greater numbers of) pitch skips/leaps. Examination of continuous tempo tracks from Experiment 2 revealed a range of different strategies that participants used to determine the appropriate final tempo. Moreover, it took longer (more beats and more melody repetitions listened to) for participants to decide on an appropriate final tempo when the melody contained fewer pitch changes. Results support the view that melodic characteristics provide non-temporal information to participants about what an appropriate tempo for a melody should be.

Introduction

In the field of music cognition, research on tempo has primarily focused on the concepts of preferred tempo (i.e., spontaneous tapping rate), tempo memory (tempo memory and imagery for familiar songs, or group memory for tempo), and tempo identification or discrimination (the ability to identify or discriminate different tempos). There has been little research on tempo determination, or the process of actively determining an appropriate tempo for an unfamiliar musical excerpt. In this study, two experiments investigated the role that melodic characteristics play (if any) in providing contextual cues for an appropriate tempo for an unfamiliar melody.

Prior research has shown that musical characteristics including melodic, rhythmic, and harmonic cues can have an influence on composed, perceived or preferred tempo. Melodic characteristics can include such things as pitch height (Broze and Huron 2013), amount of ornamentation (Kuhn 1987), or the number of contour changes (Boltz 1998). Rhythmic cues can include the attack rate and rhythmic complexity, the amount of syncopation, and the presence or absence of hypermetrical levels (London 2004/2012). Harmonic cues can include the rate of harmonic change, the specific harmonic progressions being used (Lerdahl & Krumhansl 2007), and the amount of repetition in harmonic structure (Margulis 2013).

Boltz (1998) investigated the role of contour changes and pitch skips on tempo perception using a paired comparison task where participants were asked to rate the tempo of a comparison melody relative to a standard melody. Boltz found that comparison melodies that had more contour changes and larger pitch skips than the standard melody were judged to be slower in tempo than the standard melody – even when the standard and comparison melodies had the same tempo. Quinn and Watt (2006) studied appropriateness of tempo using a two-alternative forced choice task in which participants were asked to rate a piece of music as 'too fast' or 'too slow,' using the point at which the probability of the two responses were equal as the optimal tempo. They found that participants were able to make consistent judgments regarding the appropriateness of the tempo of their stimuli, which were MIDI versions of Scottish fiddle music, and that the musical contents of the extract (melody and rhythm) contributed to the perception of the appropriate tempo, though they did not find a relationship between contour changes or interval size with perceived appropriate tempo.

The current study extends prior research on appropriate tempo using a novel experimental design in which participants were presented with a continuously repeating (looped) melodic stimulus and asked to actively determine what they believe to be the 'right,' or most appropriate, tempo for the melody by manipulating the tempo in real time via a spin wheel controller. Two experiments were conducted. The experiments differed in that Experiment 1 only recorded the final 'right' tempo determined by participants, while Experiment 2 additionally recorded the continuous tempo tracks generated by each participant for each melody that led up to their final tempo determination. Of primary interest was whether, similar to Boltz (1998), the musical characteristics of each melody (including pitch range, number of contour changes, number and size of skips and leaps) would affect the appropriate tempo selected by participants for each melody.

Experiment 1

Methods

Participants. Fifty-eight undergraduate students (49 female, M = 20 years of age, SD = 1.6) from the Department of Psychology subject pool participated in the experiment in for course credit. Participants self-reported normal hearing and varied in number of years of formal music training from zero to ten or more years.

Stimuli. Stimuli were thirty-one isochronous melodies (14–34 notes, M = 21.3, SD = 5.1). Melodies were short unfamiliar excerpts adapted from Haydn Baryton Trios and other similar style classical pieces. Each melody was recorded with a starting tempo of 100-120 beats per minute, using a grand piano timbre from the EastWest sound library. All notes had the same duration and the same amplitude to minimize metrical cues.

Melodies were coded for multiple melodic characteristics, including pitch range, number of non-unison intervals, number of contour changes, number of skips and leaps, number of simultaneities, mean interval size, mean skip or leap size, and melodic density (see Table 1). The pitch range was calculated in semitones as the distance between the highest and lowest notes in the melody. The number of non-unison intervals was defined as the number of non-zero intervals – that is, the number of intervals between non-repeating successive notes. Contour change was defined as a change from an upward (ascending) to downward (descending) pitch trajectory or from a downward to upward pitch trajectory. A skip/leap was defined as an interval greater than or equal to 3 semitones between successive notes. Simultaneities were defined as a skip or leap at the same time as a contour change. The average pitch interval size was the sum of all intervals (in semitones) divided by the number of notes. The average skip/leap size was the average of all intervals greater than or equal to three semitones. Melodic density was calculated as the number of contour changes divided by the number of notes, giving a proportion of the number of intervals in a melody that involved a contour change.

Table 1. Descriptive summary of the melodic characteristics of the thirty-one melodies. Standard deviations are shown in parentheses.

Melodic Characteristic	Minimum	Maximum	Mean
Pitch Range	7	20	12.5 (3.1)
# of non-unison Intervals	11	32	19.1 (5.0)
# of Contour Changes	3	18	8.90 (3.6)
# of Skips/Leaps	1	17	7.32 (4.3)
Mean Interval Size	1.3	4.9	2.70 (0.9)
Simultaneities	0	12	4.90 (3.3)
Mean Skip/Leap Size	3	8.4	4.80 (1.1)
Melodic Density	0.19	0.71	0.42 (0.16)

Procedure. The experiment was presented on an iMac computer using a custom experimental program written in PsiExp (Smith 1995). After a practice melody, the thirty-one melodies were presented randomly to participants. Participants heard each melody in a continuous loop and used an Arduino spin wheel to manipulate the tempo of each melody in real time. The spin wheel controlled a phase vocoder, written in Miller Puckette's open source programming environment *Pure Data*, that manipulated the tempo without changing pitch or introducing other artifacts. The starting tempo of each melody was 100 bpm, corresponding to a 600-ms inter-onset interval (IOI) between notes. To begin the presentation of each melody, participants had to turn the spin wheel, which added a randomized jitter of plus or minus six percent to the starting tempo of each melody. Participants were instructed to "use the spin wheel to control the speed of the melody until you have found the 'right' speed for the melody." Participants were allowed to listen to the melodic stimulus for as long as they wished to complete the task. When they had determined the 'right' tempo for a melody, they indicated their choice by using the mouse to press a button in the experimental program, which recorded their response for that melody and advanced to the next randomly selected melody. Following completion of the tempo determination portion of the experiment, participants completed a short demographic survey, the Goldsmith's Musical Sophistication Index (Müllensiefen et al., 2013) and a daily strategies survey, which asked questions about their perception of the experiment.

Data analysis. The primary dependent variable was the final selected tempo recorded as the final inter-onset-interval (IOI) between notes (in milliseconds). Trials where the selected tempo was greater than two standard deviations from the mean IOI were excluded from the final analysis; in total, there were nine exclusions from a total of 1,798 data points (<1% of the data).

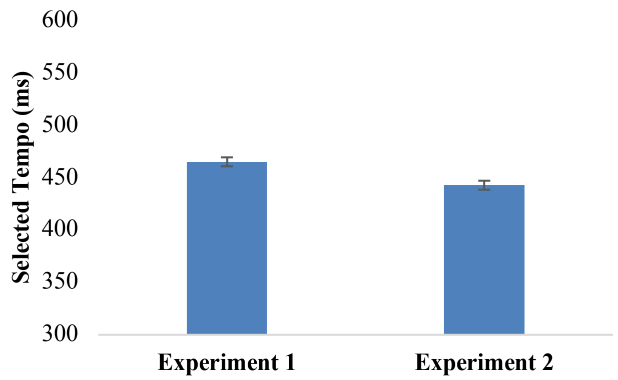


Figure 1. Average selected tempo across all melodies in Experiment 1 (left) and Experiment 2 (right)

Results and Discussion. Figure 1 (left bar) shows the average final selected tempo selected for the 31 melodies. In general, although the starting tempo was 100 bpm, corresponding to 600 ms between note onsets, individuals tended to set final tempos that were faster than the starting tempo. The final selected tempos ranged from 423 ms to 502 ms (M = 462, SD = 22.5). To consider whether different melodic characteristics predicted the tempo selected for each melody, Pearson correlation coefficients were calculated between each melodic characteristic and final selected tempo (averaged across participants). The results are shown in Table 2 and Figure 2. Greater pitch range, increasing number of non-unison intervals, number of contour changes, skips/leaps, as well as simultaneous skips/leaps and mean skip/leap size, were associated with slower final selected tempos (range, $r = 0.43$, $p < 0.05$; number of non-unison intervals, $r = 0.48$, $p < 0.01$; number of contour changes, $r = 0.57$, $p < 0.01$; number of skips/leaps, $r = 0.37$, $p < 0.05$; simultaneous skips/leaps, $r = 0.42$, $p < 0.05$, average skip/leap size, $r = 0.41$, $p < 0.05$). A step-wise regression revealed that the number of contour changes accounted for 33% of the variance in final selected tempos, with no other melodic characteristic accounting for additional variance. In sum, results show that participants use non-temporal melodic characteristics to determine the 'right' tempo for an unfamiliar melody. Moreover, melodies with more melodic change lead to slower final selected tempos.

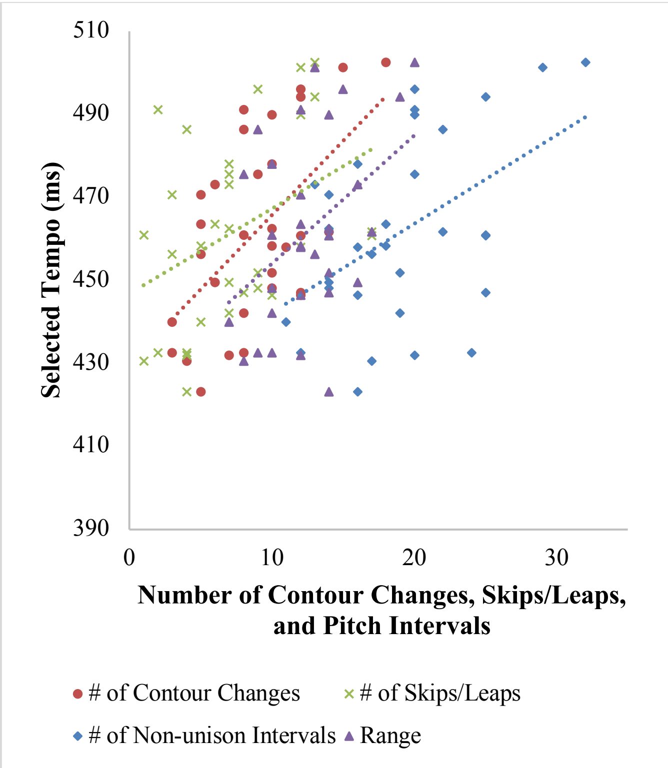


Figure 2. Relation between number of contour changes, number of skips/leaps, and number of non-unison intervals with selected tempo for Experiment 1.

Table 2. Correlation between melodic characteristics and final selected tempos for Experiment 1 and Experiment 2.

Melodic Characteristic	Exp. 1	Exp. 2	Both
Pitch Range	0.43 *	0.25	0.37 *
# of non-unison intervals	0.48 *	0.38 *	0.48 **
# of Contour Changes	0.57 **	0.33	0.49 **
# of Skips/Leaps	0.39 *	0.11	0.27
Mean Interval Size	0.35	0.11	0.25
Simultaneities	0.42 *	0.06	0.25
Mean Skip/Leap Size	0.41 *	0.25	0.36 *
Melodic Density	0.30	0.07	0.20

** Correlation is significant at the 0.01 level.
* Correlation is significant at the 0.05 level.

Experiment 2

Methods

Participants. Fifty-nine undergraduate students (45 female, M = 19 years, SD = 1.3) from the Department of Psychology subject pool participated in the experiment in return for course credit. Participants self-reported normal hearing and varied in number of years of formal music training from zero to ten or more years.

Stimuli. Stimulus melodies in Experiment 2 were the same as those in Experiment 1.

Procedure. The procedure for Experiment 2 was identical to that in Experiment 1, except that continuous tempo-tracking data was obtained for each participant and melody to investigate how people arrived at a particular tempo determination. The Arduino spin wheel apparatus is capable of outputting data at a rate of up to 960 samples/second, allowing the rate of data recording to exceed the speed of participants' responses.

Data analysis. Trials where the final selected tempo was greater than two standard deviations from the mean or where a participants' individual real-time tempo adjustments went outside a 24 bpm – 1200 bpm tempo range were excluded from the final analysis. Excluded data accounted for only thirty-nine of 1,829 total trials (~2% of the total data points).

Results and Discussion. Figure 1 (right bar) shows the final selected tempos for the thirty-one melodies. Similar to Experiment 1, although the starting tempo was 100 bpm (or 600 ms between note onsets), individuals tended to set final tempos that were faster than the starting tempo. The final selected tempos ranged from 402 ms to 494 ms (M = 441, SD = 24.0). The correlation between final selected tempos in Experiment 1 and Experiment 2 was $r = 0.63$, $p < 0.01$.

Pearson correlation coefficients were calculated between each melodic characteristic and final selected tempo (averaged across participants). The results are shown in Table 2 and Figure 3. Increasing number of notes was associated with slower final selected tempos ($r = 0.38$, $p < 0.05$). In contrast to Experiment 1, the number of skips/leaps in a melody was not correlated with the final selected tempo ($r = 0.11$, $p = 0.54$). Similar to Experiment 1, participants set a slower tempo for melodies which had more contour changes ($r = 0.33$, $p = 0.07$), but this relation was only marginally significant. A step-wise regression revealed that the number of non-unison intervals accounted for 15% of the variance in final selected tempos, with no other melodic characteristic accounting for additional variance. A step-wise regression combining data from both experiments revealed that number of contour changes accounted for 24% of the variance.

Analysis of the continuous tempo tracks revealed that participants used a range of strategies to determine a final selected tempo for each melody. Some participants tended to listen to each melody for an extended time period before adjusting the tempo, while others made immediate tempo adjustments. Some participants used a strategy of alternating between fast and slow tempos before arriving at a final selected tempo, while other participants tended to make only small adjustments to the tempo before arriving at a final selected tempo. The average number of beats listened to each melody was negatively correlated with the number of skips/leaps in each melody, the number of contour changes, the number of simultaneous skips/leaps and contour changes, the mean interval size, melodic density, and the number of non-unison intervals (skips/leaps, $r = -0.42$, $p < 0.05$; contour changes, $r = -0.59$, $p < 0.01$; simultaneities, $r = -0.36$, $p < 0.01$; mean interval size, $r = -0.43$, $p < 0.05$; density, $r = -0.47$, $p < 0.01$, non-unison intervals, $r = -0.37$, $p < 0.01$). Participants tended to arrive at a final selected tempo more quickly when there was more melodic change in a melody (e.g., more contour changes, a large pitch range, etc).

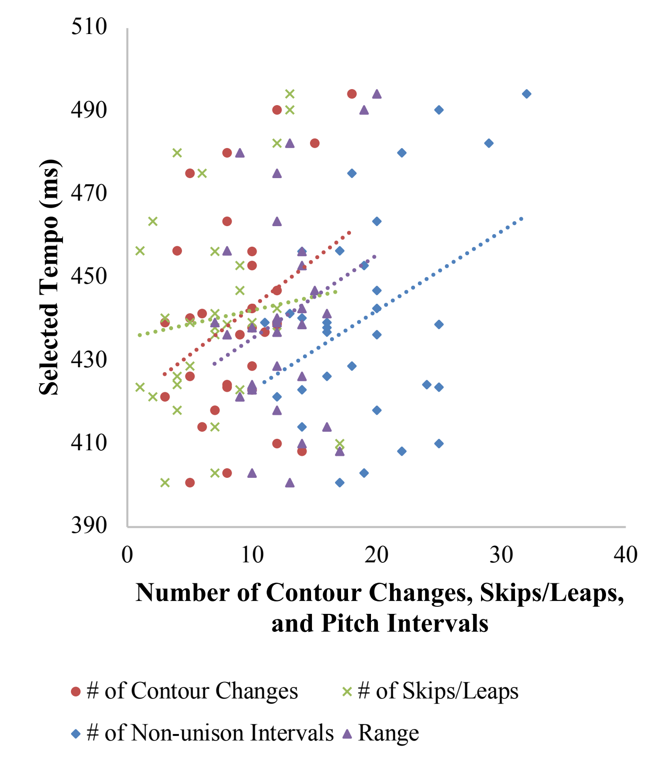


Figure 3. Relation between number of contour changes, number of skips/leaps, and number of non-unison intervals with selected tempo for Experiment 2.

Conclusion

Prior research by Boltz (1998) and Quinn and Watt (2006) demonstrated that melodic characteristics influenced the perception of tempo in a paired comparison or forced-choice task. Our results indicate that this influence of melodic characteristics carries over to our tempo determination task, where participants have complete control over the selection of an appropriate or 'right' tempo for a melody. Melodies with more contour changes and more skips and leaps are more likely to be set at slower tempos. In addition, participants were able to determine the tempo for melodies with more melodic activity more quickly than for melodies with less melodic activity, indicating that the presence of melodic activity is a determining factor for tempo.

These experiments show that even when given complete freedom to determine tempo, rather than using a paired comparison or forced-choice task as in previous studies, subjects use melodic characteristics such as contour change and the number and size of intervals as cues for appropriate tempo. Similar to Boltz, we found that increased number of contour changes was correlated with slower determined tempos; however, we found inconsistent results for the number of skips and leaps and number of simultaneities across the two experiments. Table 2 includes a column of correlations for the combined data for the two experiments, showing a strong correlation overall for pitch range, number of non-unison intervals, contour changes, and mean skip and leap size. An increase in each of these characteristics is correlated with a slower determined tempo. The continuous data recorded in Experiment 2 indicates that participants took less time to complete the task if melodies had more contour changes or skips and leaps; this indicates that an increased number of contour changes or skips and leaps made it easier for participants to determine the 'right' tempo because of the increased melodic activity. The continuous data also indicates that participants used a variety of strategies to complete the task; a further investigation of these strategies may help interpret these results.

Acknowledgements. The authors would like to thank Stephen McAdams and Bennett Smith and the MPCL lab at McGill for their assistance with early stages of this project, Lynnsey Lambrecht (Michigan State) for assistance with stimuli creation, and Daniel Alt for coding assistance. The authors would also like to thank Paul Clancy, Isabella Green, Alison Eberle, Becca Vroegop, Gabby Kindig, and the members of the MSU TAP Lab.

References

Boltz, M. G. (1998). Tempo discrimination of musical patterns: Effects due to pitch and rhythmic structure. *Perception & Psychophysics*, *60*(8), 1357-1373.

Broze, Y., & Huron, D. (2013). Is higher music faster? Pitch–speed relationships in Western compositions. *Music Perception: An Interdisciplinary Journal*, *31*(1), 19-31.

Kuhn, T. L. (1987). The effect of tempo, meter, and melodic complexity on the perception of tempo. *Applications of research in music behavior*, 165-174.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception: An Interdisciplinary Journal*, *24*(4), 329-366.

London, J. (2004/2012). *Hearing in time: Psychological aspects of musical meter*. Oxford University Press.

Margulis, E. H. (2013). *On repeat*. Oxford University Press.

Müllensiefen, D., Gingras, B., Stewart, L., & Musil, J. J. (2013). Goldsmiths Musical Sophistication Index (Gold-MSI) v1. 0: Technical Report and Documentation Revision 0.3. London: Goldsmiths, University of London.

Quinn, S., & Watt, R. (2006). The perception of tempo in music. *Perception*, *35*(2), 267-280.

Smith, B. (1995). PsiExp: An environment for psychoacoustic experimentation using the IRCAM musical workstation. *Society for Music Perception and Cognition Conference '95*. Berkeley, CA: University of California, Berkeley.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Behavioral Studies on the Role of Melodic Contours in Linguistic Processing in Chinese Musicians and Non-musicians

Yun Wang,[1] Udo Will[2]

[1,2]*School of Music, Ohio State University, USA*

[1]yunwang012221@gmail.com, [2]will.51@osu.edu

Abstract

The role of melodic contour has been an appealing perspective in understanding the relationship between musical and linguistic information processing. Our previous work provides evidence that tonal information may facilitate vocal responses, and vocal/non-vocal contour information may elicit different influence on linguistic processing (Poss&Will, 2007; Will&Poss, 2008; Will et al., submitted). That tonal information may have an influence on lexical processing could also been seen in many musical speech surrogates. In practical applications, musical speech surrogates map speech onto instruments or whistling by maintaining either pitch information or formant contour (Sebeok&Umiker-Sebeok, 1976), suggesting that tonal information may facilitate lexical processing. However, it is unclear what the influence of contour information on linguistic processing might be. In this study, we aim to explore whether and how melodic contours affect linguistic processing, and what the role of musical training and experience in the process might be. We conducted a lexical decision task and a syllable repetition task with Chinese musicians and non-musicians. A priming paradigm was applied in both experiments. Our experiments show that melodic contour primes facilitate certain types of linguistic processing in tone-language speakers, and the facilitatatory effect appears to be both lexical and extra-lexical. Musicians responded consistently faster than non-musicians. Our results also demonstrate that vocal contours produced faster responses than instrumental contours and unpitched white noise.

Introduction

Melodic contour, as Edworthy (1983) defined, refers to “the sequence of ups and downs independent of the precise pitch relationships” (p.263). It is considered as one of the basic aspects of music and easier perceptible than intervals (e.g. Edworthy, 1985; Patel, 2008). Edworthy (1985) conducted several experiments, where participants were required to detect violations in interval and contour tasks. She pointed out that “contour information is immediately available to the listener regardless of novelty, familiarity, transposition, or nontransposition” (p.172), while interval processing is more dependent on tonality. Different from many other musical abilities, the ability to process contour is widespread in not only musicians but also non-musicians, not only in adults but also infants (e.g. Plantinga&Trainor, 2005; Dowling, 1999; Trehub et al., 1984). Currently, researches concerning how contour information affects linguistic processing gradually set up a bridge connecting linguistics and musicology, which may further offer new evidence in explaining musical phenomenon, such as musical speech surrogates (e.g. Poss, Hung, & Will, 2008; Will &Poss, 2008; Poss, 2012; Cheong et al., 2017).

Previous work regarding the role of contour in linguistic processing adopted different tasks and arrived at conflicting results. However, many interpretations are complicated by the fact that the stimuli used also contain segmental information. We have previously introduced the use of melodic contour prime stimuli without segmental information, which allows to specifically focus on the effect of tonal information. Results have shown that pitch contour primes facilitated responses to both words and pseudo-words compared to non-pitched noise (e.g. Will &Poss, 2008; Poss, 2012). In addition, many factors were found to be important in how melodic contours influence the responses to target words. For example, effects of timbre and lexical category were reported (e.g. Will et al., submitted). Moreover, cultural factors, such as musical training and language background were found to be important in the process as well (Will et al., submitted).

However, a more refined lexical category classification needs to be made, as our previous studies did not distinguish between pseudo-words and non-words and did not analyze specific effects on syllables with different tonal information. What is more, the interesting factor gender was ignored in previous studies. Do effects in males differ significantly from those in females? How does this factor interact with other factors? These questions will be addressed in the current study.

We performed a lexical decision task and a syllable repetition task to test at which level of cognitive processing there is an interaction between the prime and target contours. In addition, conducting both a shadowing and a lexical decision task allows for a comparison of the error responses in the two tasks, which might yield additional insights in the underlying processes.

Experiment 1: Lexical Decision Task

Participants

We recruited all participants in China. Since we are interested in effects of contour primes on tone language speakers, we chose Chinese participants and Chinese stimuli in this study. To qualify as Chinese musicians, subjects needed to satisfy all the following criteria: 1) at least 5 years of formal musical training, 2) ongoing active musical engagement at the time of participation, 3) self-identification as a musician, and 4) native Chinese speaker.

Those participants who are native Chinese speakers but did not meet all the other three criteria were categorized as Chinese non-musicians. For the musicians, there was no specific requirement for their major, since the influence of their specific musical expertise (singing, instruments) is not a factor for this study (though we still collected this information in the questionnaire). All musicians had passed a competitive entrance examination before they entered their program, and were required to take ear training and sight singing class in the first two years. Participants in this lexical decision task consisted of 40 adults (mean age: 20± 4 years for non-musicians, 21± 3 years for musicians, 20 musicians, 20 non-musician) with gender-matched, i.e. 10 males and 10 females in each group. Musicians at least had 5-years formal music training, and for non-musicians, none had any formal music training. All of the participants were right-handed based on an adapted version of Stanley Coren's handedness questionnaire (Anonymous, n.d.). None of the participants reported abnormal hearing or motor disability.

All participants gave their informed consent before participating in this study in accordance with the Ohio State University Institutional Review Board regulations, and all of them completed the task.

Stimuli

In this study, we used a priming paradigm, where stimuli were presented in pairs: each trial consisted of a prime and a target syllable. We used two types of primes: one was vocal, and the other instrumental. All primes had a melodic contour corresponding to one of the four Mandarin speech tones. Vocal contour primes were created by extracting the contour information from spoken syllables recorded by a native female speaker and were then resynthesized with a formant spectrum corresponding to the resonance of the human vocal tract. Thus, there was no segmental information in vocal primes. The instrumental contour primes were produced by playing a traditional Chinese instrument, Xiao, a vertical aerophone. Besides these two types of primes with timbral contrast, a control prime was constructed by convolving a spoken syllable's amplitude envelop with white noise. Thus, the control prime did not offer any segmental or tonal information. All the primes (instrumental, vocal and control) were set to have the same duration of 420 msec, and RMS power (both average and total) was similar around -23 dB.

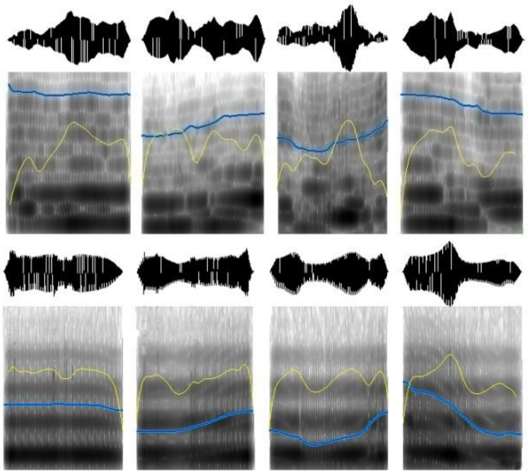


Figure 1. Instrumental (top) and vocal (bottom) primes in the four standard Mandarin tones (1, 2, 3, and 4; from left to right). Top: waveforms; bottom: spectrograms with pitch (blue) and intensity (yellow) curves.

In this study, we took Chinese, a tonal language, where lexical tones offering pitch contrast are used to mark word meanings, as targets. There are four tones, marking high-level pitch, rising pitch, low-dipping pitch and falling pitch, respectively. In the both experiments, 112 spoken syllables were equally distributed in four tones. All the target syllables were recorded by a native female speaker at a sampling rate of 44.1 kHz.

In addition, targets were also divided into three categories, as words (n=48), non-words (n=32) and pronounceable pseudo words (n=32). Non-words are pronounceable phoneme combinations (syllables) not used in Mandarin. Pseudo-words are phoneme combinations that are used as words in Mandarin, but not in combination with the specific tone(s).

There were three types of prime conditions (the relationship between prime and target pitch contour), match, non-match and control. If the prime shares the same contour with the target, it is in the match condition, and if prime differs from the target in contour, it is in the non-match condition. For the control condition, it means the prime involved is a control prime (white noise) with no tonal contour.

Experimental Design and Procedure

The stimulus onset asynchrony (SOA, the time interval between the end of the prime and the start of the target syllable) was set to 250 ms.

For each subject, trials were divided into 3 sections, thus participants would have some rest time during the experiment. The total number of prime & target pairs presented in each run was 336. Each target syllables appeared three times in each run with either match/non-match/control condition (with 112 in match condition, 112 in non-match condition and 112 in control condition). Each target syllable would not appear more than once in one section. In addition, every syllable appears with control prime, but half of the syllables appear with vocal, the other half with instrumental primes. Target syllables with vocal and instrumental primes were counterbalanced across subjects.

In this experiment, participants were asked to classify the auditory target items as either words or non-words as quickly and accurately as possible. The lexical decision experiment was conducted on a Sony Vaio laptop, and all stimuli were controlled and presented by DMDX software (Forster & Forster, 2003). All stimuli were played through a headphone connected to the laptop. Response time and accuracy was recorded for each target syllable. The response time was automatically calculated from the beginning of the targets, however, as target syllables do not have identical length, we set the end of the targets as the reaction time start point. One point that needs to be clarified is that subjects sometimes responded before the end of the target items (in this case, the reaction times would be shown as negatives). However, this does not mean the response is invalid. Such a response was not necessarily taken as anticipation, because identification of lexical entries can happen before a syllable is fully presented. Thus, the negative reaction times only demonstrated that participants successfully used their background or other sources to make the lexical decision before the target offset.

Results

We removed responses made in less than 200 ms from the onset of target (0.02%), and those responses made in more than 2000 ms from the onset of target (1.56%), because responses beyond these thresholds may likely be a guess or anticipation. Response times were analyzed based on correct responses, so we removed those incorrect answers, which counts for 11.3% of the items. We applied SAS software with MIXED (effect) procedure for reaction time analysis. Due to the complexity of experimental design, using a couple of nesting options in order to avoid too long run times of the experimental sessions, we did not apply mixed random effects modeling (i.e. taking both subjects and target syllables as random factors in one model). Rather, mixed effect models that included random effects for subjects with intercepts (F1) and for target syllables with intercepts (F2) were constructed for reaction time as dependent variable.

Our results showed that the effect of lexical category reached significance in both by-subject (F1) and by-item (F2) analysis (F1(2, 39) =373.4, p<0.0001, F2(2, 111) =18.08, p<0.0001). Non-words were responded to most quickly. Words were in the middle, and pseudo-words required longest time to make a button decision.

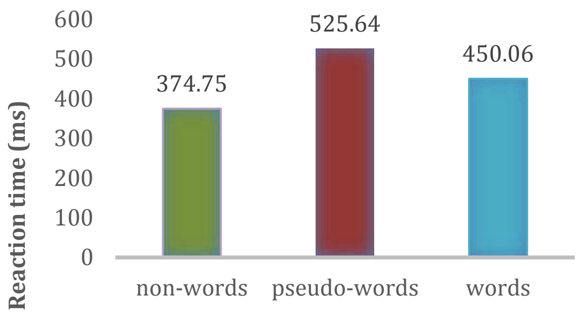


Figure 2. The main effect LEXICAL CATEGORY in terms of reaction time (expt 1)

The main effect of gender reached significance in both by-subject (F1) and by-item (F2) analysis (F1(1,39)=10.53, p=0.0025, F2(1,111)=623.63, p<0.0001). Female participants on the average took a shorter time to give a feedback than male participants (-112.06 ms). There was a main effect of music training only in by-item analysis (F2(1,111)=187.52, p<0.0001). Musicians and non-musicians showed mean reaction times of 416.26 ms and 478.85 ms, respectively. There was also a 2-way interaction lexical category: prime condition in both by-subject (F1) and by-item (F2) analysis (F1(4,39)=2.65, p=0.0145, F2(4,111)=2.37, p=0.0275). For words, there is an increased facilitation (-20.58 ms) in the match condition but not in non-match condition (-6.19 ms). In contrast, non-words show an augmented facilitation in non-match condition.

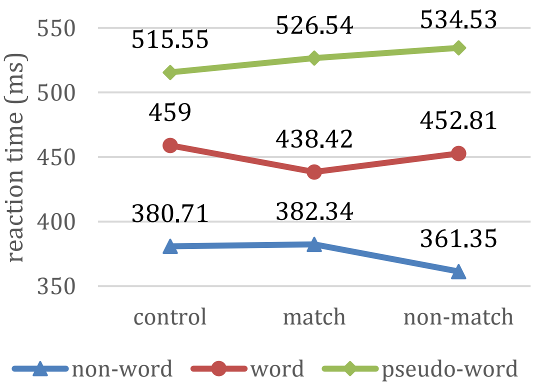


Figure 3. The LEXSTATUS*PRIME CONDITION interaction in terms of reaction time (expt 1)

A 3-way interaction music training: lexical category: gender reached significance in both by-subject (F1) and by-item (F2) analysis (F1(2,39)=7.83, p<0.0001, F2(2,111)=10.23, p<0.0001).

We applied PROC GLIMMIX with binomial distribution and logit function to analyze accuracy. The results showed that the main effect of lexical category was significant in both by-subject (F1) and by-item (F2) analysis (F1(2,39)=69.43, p<0.0001, F2(2,111)=13.49, p<0.0001). On the average, participants reached 95.35% correct rate for non-words, 88.27% for pseudo-words, and 87.86% for words. The effect of music training reached significance only in by-subject analysis (F1(1,39) =5.89, p=0.0198). Non-musicians (91.82%) had a higher correct rate than musicians (88.44%). The 2-way interaction music training: lexical category reached significance in both by-subject (F1) and by-item (F2) analysis (F1(2,39)=11.91, p<0.0001, F2(2,111)=22.32, p<0.0001). The 3-way interaction music training: lexical category: gender was significant in both by-subject (F1) and by-item (F2) analysis (F1(2,39)=10.73, p<0.0001, F2(2,111)=13.64, p<0.0001).

Discussion

The results show that non-words were responded to much faster than words (-75.31 ms) and pseudo-words (-150.89 ms). This indicates that in the lexical decision task it is faster to make a decision when a syllable is not a word (as there is no entry in the mental lexicon exists). For pseudo-words, a prolonged search was required because they were very similar to real words, and making a decision therefore took more time.

Our results concerning reaction time of pseudo-words and non-words, support the information load hypothesis, in which tones carry the least information compared with consonants and vowels. Thus, it would be easier for native Chinese speakers to reject non-words with illegal phoneme combination because they carry a large amount of information that can constrain lexical access. Interestingly, our study suggests that non-word processing maybe primarily based on phonological discrimination, but tonal information may also exert an influence. This is supported by the finding that the decision process was speeded up when melodic contour primes and non-word syllables differ in tones. In addition, under match condition, the facilitatory effect was only found in words. This helps us understand how musical speech surrogates work, because many musical speech surrogates map lexical tones onto whistling/instrument pitches: as tonal information is a part of lexicon, melodic pitch contour primed the pre-activation of the lexicon, resulting in the larger facilitation size.

In the current study, subjects with music training background responded 51.63 ms faster than non-musicians. Our experiments are in line with previous studies, indicating that musical training may facilitate participants' responses. However, in terms of accuracy, non-musicians were only 2.7 pp more correct than musicians, and this probably is because musicians adopted a different listening strategy, where they weighted speed over accuracy even though they were instructed to respond as accurately and quickly as possible.

Meanwhile, gender as a factor affected participants' response significantly, and females had a faster reaction time than males in this experiment. As Roebuck and Wilding (1993) suggested, female listeners and male listeners performed better when they heard a voice from the same sex. That could be the reason why female subjects performed significantly better since we used a female voice for the recording of the stimuli in our experiment. Though this is not to say that females are generally better at auditory processing, this experiment suggest that sex differences play a role in pitch contour processing because the voice contains information about the gender of the speaker.

Experiment2: Syllable Repetition Task

Participants

In the syllable reproduction experiment, we again recruited Chinese musicians and Chinese non-musicians as our subjects. The criteria for Chinese musicians and non-musicians were same as in our previous lexical decision experiment.

All musicians had passed a competitive entrance examination before they entered their program, and were required to take ear training and sight singing class in the first two years. Among the 28 participants there were 14 musicians (mean age=25.6± 8 years,11females), and 14non-musicians (mean age= 22± 4years, 7 females). Musicians on the average had 17-years formal music training, and no non-musician had any formal music training. All of the participants were right-handed, established by an adapted version of Stanley Coren's handedness questionnaire (Anonymous, n.d.). None of the participants reported abnormal hearing or motor disability. All participants gave their informed consent before participating in this study in accordance with the Ohio State University Institutional Review Board regulations.

Procedure

In this syllable reproduction task, we again used a priming paradigm with a stimulus onset asynchrony of 250 ms between primes and targets.

For the syllable reproduction task, we used the same prime-target pairs as in the first experiment.

In this experiment, participants were required to orally repeat the syllables they heard as quickly and accurately as possible. The experiment was conducted on a Sony Vaio laptop, and all stimuli were controlled and presented by DMDX software (Forster & Forster, 2003). There was a headphone with mic attached to computer and participants' vocal responses were recorded for each target item. After the experiment, response times of all the target items were checked and corrected with CheckVocal (Protopapas, 2007). Accuracy was checked by controlling for correct reproduction of both the phoneme sequence and the speech tone of each response.

Results

We removed the responses made in less than 200 ms from the onset of target (0.17%), and those responses made in more than 2000 ms from the onset of target (0.04%), because responses beyond these thresholds may likely be a guess or anticipation.

Response times were analyzed based on correct responses, so we removed incorrect answers, which counts for 7.03%. We applied SAS software with MIXED (effect) procedure for reaction time analysis. The results showed that the effect of timbre was significant in both by-subject (F1) and by-item (F2) analysis ($F1(2,27)=6.26$, $p=0.0019$, $F2(2,111)=7.76$, $p=0.0004$). Results showed that both instrumental and vocal primes facilitated participants' reaction time. Post-hoc test (lsmeans) showed that comparison between control and vocal ($p=0.0003$), instrumental and vocal ($p=0.0425$) reached significance (in by-item analysis).

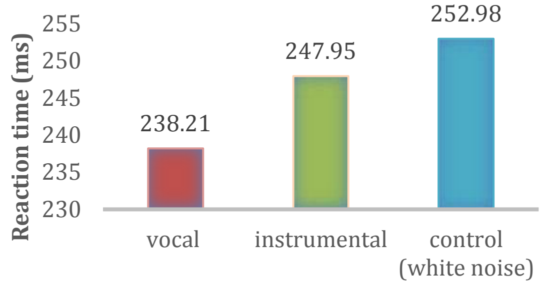


Figure 4. The main effect TIMBRE in terms of reaction time (expt 2)

There was a main effect of lexical category only in by-subject analysis ($F1(2,27)=64.05$, $p<0.0001$). Post-hoc test (lsmeans) showed that comparison between words (267.54 ms) and pseudo-words (228.63 ms), words and non-words (231.65 ms) reached significance (all p-values are less than 0.0001). The effect of status reached significance in both by-subject and by-item analysis ($F1(1,111)=4.62$, $p=0.041$; $F2(1,111)=343.3$, $p<0.0001$). Musicians on the average took a shorter time to repeat syllables than non-musicians.

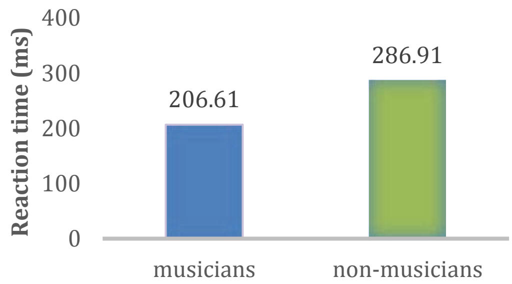


Figure 5. The main effect STATUS in terms of reaction time (expt 2)

We applied PROC GLIMMIX with binomial distribution and logit function to analyze accuracy. The results showed that the main effect of lexical category was significant in both by-subject (F1) and by-item (F2) analysis ($F1(2,27)=61.05$, $p<0.0001$, $F2(2,111)=3.28$, $p=0.0417$). Post-hoc test (lsmeans) showed that comparisons between non-words (87.64%) and words (94.23%), and between pseudo-words (95.62%) and non-words reached significance.

Discussion

In this experiment, both instrumental and vocal primes have a facilitating effect, indicating that melodic contours may affect linguistic processing. Interestingly, vocal stimuli elicited larger priming than instrumental ones, suggesting that our brain may deal with timbres differently. As Fitch (2006) proposed, researchers may investigate the origin of music by exploring two different paths determined by timbre, namely, vocal and non-vocal. Our results support the idea that vocal and non-vocal sounds are processed differently, in line with previous studies arguing for a clear distinction between timbre (Poss, 2012; Hung, 2011).

In line with previous studies and our lexical decision task, the results showed that musicianship indeed played an important role. In this syllable reproduction task, musicians responded 80.3 ms faster than non-musicians. In addition, we noticed a difference between musicians and non-musicians in wrong tone production. Results showed that for error items, wrong tones only occurred 4 times in musicians (0.61% of all errors), however, they occurred 25 times in non-musicians (3.79%). There was also a large difference in phoneme error rate in musicians (255 errors, 38.64%) and non-musicians (376 errors, 56.97%). This indicates that musicians are not only more sensitive to tonal information but also show higher phonological awareness than non-musicians.

General Discussion

Our experiments show that melodic contour information can influence in tone-language speakers. In the lexical decision task matching contour primes facilitated decisions on words but not on non- and pseudo-words: information delivered by the primes speed-up lexical processing. We also found facilitatory effects that seem to be pre- or non-lexical. 1) In the lexical decision task decision on non-words, which do not have lexical entries, are speeded-up if the contour primes do not match the contour of the target syllables. 2) In the syllable repetition task vocal and instrumental contour primes caused facilitation, but there was no significant difference between the match and non-match condition, suggesting that this effect is not caused along the lexical pathway. We suggest both these facilitatory effects are part of the mechanisms underlying communication through speech surrogates.

In addition, the effect of timbre was significant only in the syllable repetition task, suggesting the priming effect may not be connected with higher lexical decision processes. The syllable repetition task was in line with previous studies, indicating the "humanness bias" (Lévêque&Schön, 2013; Hung, 2011; Klyn et al., 2016), that is, vocal sounds enjoy superiority over non-vocal sounds. Since humans are social animals living in association with others mainly through speech, vocal sounds acquire a biological and social significance in human's life during the long evolutionary history. Human have different physiological reactions when being exposed to vocal sounds than to non-vocal sounds (Belin et al., 2002; Loui et al., 2013), supporting that we have greater excitability during listening to a specific timbre. In the speech shadowing task, vocal primes speeded up responses more than instrumental primes, supporting that the special role played by the voice in human communication contributed to an increased sensibility and attention to human voice. In addition, speech motor theory suggests that we identify articulatory gestures when hearing speech, and vocal tract muscles, pre-motor as wells as motor cortexes will be activated. Though there is no vocal articulatory gesture in our contour stimuli, vocal contours did have melodic gesture, which is clearly a part of vocal gestures when we produce speech. Thus, we argue that vocal contours may activate the speech motor system, which is engaged in the vocal tract movement planning, leading to faster responses.

That social-cultural differences lead to different behavioral performance has been confirmed in many studies. In both experiments, musicians outperformed non-musicians in reaction time. Our results together with previous studies show that musical training may enhance participants' performance due to their extended experience with pitch and contour processing. We propose that musicians may have advanced melodic abilities, which help them process the contour information more efficiently.

Conclusion

Our experiments suggest a rather complex relationship between musical and linguistic information processing. In tone-language speakers melodic contour information was found to have lexical as well as extra-lexical effects on syllable processing. The effect size increases with musical training. The different effects of vocal and instrumental melodic contours could be due to specific cognitive adaptations.

References

Anonymous. (n.d.).Handedness Questionnaire. Retrieved from https://faculty.washington.edu/chudler/rltablen.html

Belin, P., Zatorre, R. J., &Ahad, P. (2002).Human temporal-lobe response to vocal sounds. Cognitive Brain Research, 13(1), 17-26.

Cheong, Y.J., Will, U., and Lin, Y.Y (2017) Do vocal and instrumental primes affect word processing differently? An fMRI study on the influence of melodic primes on word processing in Chinese musicians and non-musicians. Proceedings of the 25th Anniversary Conference of the European Society for the Cognitive Sciences of Music, Ghent, Belgium, 35-39.

Dowling, W. J. (1999). The development of music perception and cognition. In: Diana Deutsch (Ed.), Psychology of Music (603-625). San Diego: Academic Press.

Edworthy, J. (1983).Towards a contour-pitch continuum theory of memory for melodies. In D. Rodgers, & J. Sloboda (Eds.), The Acquisition of Symbolic Skills (263-271). New York: Plenum.

Edworthy, J. (1985). Interval and contour in melody processing. Music Perception: An Interdisciplinary Journal, 2(3), 375-388.

Fitch, W.T. (2006). The biology and evolution of music: a comparative perspective. Cognition, 100(1), 173-215.

Forster, K. I., & J. C. Forster.(2003). DMDX.University of Arizona.

Hung, T.-H.(2011). One Music?Two Musics? How Many Musics? Cognitive Ethnomusicological, Behavioral, and fMRI Study on Vocal and Instrumental Rhythm Processing (Unpublished doctoral dissertation). The Ohio State University, Columbus, USA.

Klyn, N. A., Will, U., Cheong, Y. J., & Allen, E. T. (2016). Differential short-term memorisation for vocal and instrumental rhythms. *Memory*, *24*(6), 766-791.

Lévêque, Y., &Schön, D. (2013). Listening to the human voice alters sensorimotor brain rhythms. PloS One, 8(11), 1-9.

Loui, P., Bachorik, J. P., Li, H. C., &Schlaug, G. (2013). Effects of voice on emotional arousal. Frontiers in Psychology, 4, 675.

Meyer, J. (2015). Whistledlanguages. Heidelberg: Springer Verlag.

Patel, A. D. (2008). Music, Language, and the Brain. New York: Oxford University Press.

Plantinga, J., &Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. Cognition, 98(1), 1–11.

Poss, N., & Will, U. 2007. The role of pitch in the processing of tonal languages: new evidence on the connection between music and language. Paper presentation at the SEM symposium 'New Directions in Cognitive Ethnomusicology', Columbus, OH.

Poss, N., Hung, T.H., Will, U. (2008). The Effects of Tonal Information on Lexical Activation in Mandarin

Speakers.Proceedings of the 20th North American Conference on Chinese linguistics, Columbus, OH, 205-211.

Poss, N. F. (2012). Hmong music and language cognition: An interdisciplinary investigation (Unpublished doctoral dissertation). The Ohio State University, Columbus, USA.

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. Behavior Research Methods, 39(4), 859-862.

Roebuck, R., & Wilding, J. (1993). Effects of vowel variety and sample length on identification of a speaker in a line-up. Applied Cognitive Psychology, 7, 475 – 481.

Sebeok, T. A., &Umiker-Sebeok, D. J. (1976). Speech surrogates: drum and whistle systems. The Hague: Mouton.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. Child development, 55, 821-830.

Will, U., &Poss, N. (2008). The role of pitch contours in tonal languages processing. Proceedings of The Fourth International Conference on Speech Prosody, Campinas, Brazil, 309-312.

Will, U., Poss, N., & Hung, T.H. (submitted). Melodic contours prime speech processing in tone-language and non-tone-language speakers.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Acetaminophen Blunts Emotional Responses to Music

Lindsay Warrenburg[1] and Baldwin Way[2]

[1] *School of Music, Ohio State University, USA*
[2] *Department of Psychology, Ohio State University, USA*

[1]warrenburg.3@osu.edu, [2]way.37@osu.edu

Abstract

The capacity of listeners to perceive or experience emotions in response to music depends on many factors including dispositional traits, empathy, and musical enculturation. Emotional responses are also known to be mediated by pharmacological factors, including both legal and illegal drugs. Existing research has established that acetaminophen, a common over-the-counter pain medication, blunts emotional responses (e.g., Durso, Luttrell, & Way, 2015). The current study extends this research by examining possible effects of acetaminophen on both perceived and felt responses to emotionally-charged sound stimuli. Additionally, it tests whether acetaminophen effects are specific for particular emotions (e.g. sadness, fear) or whether acetaminophen blunts emotional responses in general. The experiment employs a randomized, double-blind, parallel-group, placebo-controlled design. Participants are randomly assigned to ingest acetaminophen or a placebo. Then, they are asked to complete two experimental blocks regarding musical and non-musical sounds. The first block asks participants to judge the extent to which a sound *conveys* a certain affect (on a Likert scale). The second block aims to examine a listener's *emotional responses* to sound stimuli. The study is currently in progress; here, preliminary results are reported for 19 participants of a planned 200 cohort. In light of the fact that some 50 million Americans take acetaminophen each week, if the final results prove consistent with existing research on the emotional blunting of acetaminophen, this suggests that future studies in music and emotion might consider controlling for the pharmacological state of participants.

Introduction

Acetaminophen (paracetamol) is the active ingredient in several popular over-the-counter analgesic medications. It is estimated that over 20% of all adults in the United States consume acetaminophen at least once a week (Kaufman et *al*., 2002). Although it is typically used to reduce physical pain, acetaminophen has also been shown to reduce social pain and empathy for others (Durso, Luttrell, & Way, 2015; Mischkowski, Crocker, & Way, 2016). One of the possible reasons for these emotional effects is that there are speculated common neurochemical pathways for physical and social pain (Panksepp, 1998). If acetaminophen affects these shared neural pathways, an intended reduction in pain for the physical domain could inadvertently cause a reduction in the social domain. Consistent with this theory, recent research has shown that acetaminophen reduces neural activity in two cortical areas responsible for social pain: the anterior cingulate cortex (ACC) and the agranular insular area (AI) (DeWall et *al*., 2010).

Since acetaminophen is taken by approximately 50 million Americans every week (Kaufman et *al*., 2002), the possibility that it reduces social pain and empathy has wide-reaching implications. It is possible that people who regularly take acetaminophen are unintentionally living in a world that is less emotionally charged and is more isolated than for people who do not take the drug. In other words, those who take acetaminophen could be experiencing blunted emotional responses to affectively-charged events compared to their normal state. A recent study (Durso, Luttrell, & Way, 2015) is consistent with this theory. In their study, participants who ingested acetaminophen made attenuated emotional judgments of pleasant and unpleasant pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008), compared to those who took a placebo. Of particular importance for the present study is that participants in the acetaminophen condition experienced blunted reactions to both *perceived* and *induced* emotion in response to the pictures. That is, participants on the drug not only rated pleasant and unpleasant stimuli as *displaying* less emotion than did those who received a placebo, but they also *experienced* less emotion in response to pleasant and unpleasant pictures, compared to those in the placebo group.

In the current study, we aim to extend the findings of Durso et al. (2015) by replicating their methodology using sound stimuli instead of visual stimuli. Specifically, we aim to test whether the emotional blunting effects of acetaminophen extend to the auditory and musical domains.

In auditory research, three broad categories of sound can be distinguished: (1) natural sounds, (2) speech, and (3) music. All three categories are used in this study.

(1) *Natural Sounds*: These sounds include the sounds of natural and common manufactured objects, such as a bubbling brook or a door slamming, as well as non-speech vocalizations, such as the sounds of sneezing, coughing, crying, and screaming.

(2) *Speech*: In the case of speech, affect can be communicated via prosodic features and by semantic content. Speech prosody includes features such as precision of articulation, microstructural irregularity, and the contour of the spoken utterances. Semantic content can be roughly defined as the understood meaning of words spoken in a sentence.

(3) *Music*: Although music contains its own syntax, grammar, and formal structures, it is also known to emulate speech characteristics (Juslin & Laukka, 2003). For example, sad music contains many features of sad speech. Sad speech exhibits a quieter-than-normal voice, a slower speaking rate, low pitch, a monotone voice, mumbling, and a dark timbre (Kraepelin, 1921). Sad-music characteristics parallel these sad-speech characteristics: sad music is quieter, slower, lower in pitch, has smaller pitch movements, is legato (smooth-sounding), and uses darker timbres (Huron, 2008; Schutz, Huron, Keeton, & Loewer, 2008; Turner & Huron, 2008; Post & Huron, 2009; Yim, Huron, & Chordia, MS). In the current

study, we use all three categories of sound: natural sounds, speech, and music.

Many empirical studies have examined how people respond to semantic and prosodic elements of speech. Speech studies have included how people respond to emotionally-charged speech (e.g. Mitchell et *al.*, 2003), how language and prosody affects recognition of emotion in speech (e.g. Ververidis & Kotropoulos, 2006), how meaning can be created in language (e.g. Ricoeur, 2003), and how basic demographics affect emotional speech perception (e.g. Paulmann et *al.,* 2008; Schirmer et *al.,* 2004). By contrast, there has been comparatively little empirical research examining why emotional reactions to music vary across situations and people. In the music and emotion literature, studies tend to be divided into two categories. First, researchers study the kinds of emotions that listeners believe the music is displaying. This is often referred to in the literature as the study of *perceived emotion* – the study of how listeners recognize emotions in music. This type of research is often focused on the structural aspects of the music itself. Musical works that are thought to display fear, for example, have been examined in terms of its compositional features, such as the use of wide leaps, the creation of scream-like sounds, and instruments that are used in unusual ways (e.g. prepared piano). This area of research has given rise to a wealth of knowledge about compositional practice, performance decisions, and how music can help amplify emotional effects in movies and other soundtracks.

Second, music and emotion researchers have been interested in what kinds of emotion music is able to induce in its listeners. *Induced* or *experienced emotion* studies examine how listeners can feel an emotional response to (typically) instrumental music. Researchers want to know, for example, if people actually experience sadness when listening to sad music, and if so, why they enjoy the experience of being in a negatively-valenced state. Experts on musically-induced emotions use different methodological techniques than do experts on perceived emotion. Studies of music-induced emotions employ fMRI, EEG, and prosodic analysis of musical excerpts. The driving theories often rely on evolutionary psychology, cognitive psychology, and even comparative psychology.

In the current study, the effect of acetaminophen on emotional responses to sound will consider both *perceived* and *experienced* emotion.

Hypotheses

H1: Compared to those in the placebo condition, participants who ingest acetaminophen will *perceive* positive sounds as less positively valenced and negative sounds as less negatively valenced.

H2: Compared to those in the placebo condition, participants in the acetaminophen condition will *experience* less valenced emotion when listening to positive and negative sounds.

H3. The blunting effects of acetaminophen will differ among emotion categories. That is, some perceived and induced emotions will be blunted more than other emotions.

Methods

As noted, the goal of the study is to test the effect of acetaminophen on perceived and experienced emotion in response to emotionally-charged auditory stimuli. The experiment follows the design used by Durso et *al.* (2015) and employs a randomized, double-blind, parallel-group, placebo-controlled design.

Pharmacological Procedure

Participants signed up for the experiment online. An email was sent to the participants detailing the risk factors associated with acetaminophen. They were also instructed to avoid eating any food for three hours prior to the experiment. Participants arrived at the lab and were told that they would be given a pill containing either 1000 mg of acetaminophen or a placebo. Participants were randomly assigned to the drug or placebo condition. At the time of the study, participants and the experimenters were blind to the drug condition. After consuming the pill, participants were asked to wait 50 minutes for the drug to take effect (Durso et *al*., 2015). During this waiting period, participants completed questionnaires described below. After the study was over, participants were asked to guess whether they had been given the drug or the placebo. They were then told they should avoid consuming acetaminophen and drinking alcohol for the next 15 hours.

Stimuli and Experimental Conditions

The experiment consisted of two blocks of trial, where the first block was aimed at testing hypotheses 1 and 3 and the second block was aimed at testing hypotheses 2 and 3. The order of these blocks was counter-balanced across participants, and the sound examples were randomized within each condition. For each block, participants listened to selected musical passages (described below), speech samples from the Crowd-Sourced Emotional Multimodal Actors Dataset (CREMA-D) (Cao et *al.*, 2014), and natural sound stimuli taken from the International Affective Digital Sounds (IADS-2) (Bradley & Lang, 2007). Participants listened to the stimuli via headphones. Prior to the two blocks–while waiting for the drug to take effect–participants provided a blood sample to measure inflammation and were asked to complete measures of current affect (Positive and Negative Affectivity Scale (PANAS); Watson, Clark, & Tellegen, 1988), musical sophistication (Ollen Musical Sophistication Index (OMSI); Ollen, 2006), musical preferences (Short Test of Musical Preferences-Revised (STOMP-R); Rentfow, Goldberg, & Levitin, 2011), personality (Big 5 Personality Questionnaire; John & Srivastava, 1999), empathy (Interpersonal Reactivity Index (IRI); Davis, 1980), absorption in music (Absorption in Music Scale (AIMS); Sandstrom & Russo, 2011), and basic demographic questions.

Block 1: Perceived Emotion

The goal of the first block was to examine perceived emotion in sound stimuli. Namely, participants were asked to judge the extent to which a sound conveyed a certain emotion. The sound stimuli consisted of emotionally-charged musical excerpts, speech excerpts, and non-musical sounds.

Natural Stimuli. The natural stimuli were drawn from the International Affective Digitized Sounds library (IADS-2; Bradley & Lang, 2007). This set of stimuli consists of 167 sound stimuli that have been coded for valence, arousal, and dominance, and have been normalized in sound properties such as loudness and duration (all samples are 6 seconds in length). Sounds include non-speech vocalizations (e.g., coughing, laughing), non-human sounds (e.g., alarm clock, breaking glass), and musical excerpts (e.g., choir singing, bagpipes, whistling). No musical excerpts from the IADS library were used. The sounds in the other two categories were sorted by the rated valence (which ranged from 1.57-7.78) in the Bradley & Lang (2007) study. It was established *a priori* that audio files with valence scores less than 3.5 would be considered "low valence," audio files with valence scores between 3.5 and 5.5 would be considered "neutral valence," and audio files with valence scores higher than 5.5 would be considered "high valence." Once the audio files were sorted by valence, the sound clips were sorted into five sub-categories based on how the participants of the Bradley & Lang (2007) rated arousal levels: 1) Low-Valence, High-Arousal (LVHA); 2) Low-Valence, Low-Arousal (LVLA); 3) Neutral-Valence, Middle-Arousal (NVMA); 4) High-Valence, High-Arousal (HVHA); and 5) High-Valence, Low-Arousal (HVLA).

Speech stimuli. Speech samples were taken from the CREMA-D (Crowd-sourced Emotional Multimodal Actors Dataset; Cao, et *al.*, 2014). This dataset consists of audio (spoken), visual (video), and multimodal (audio-visual) performances by 91 professional actors (ages 5-74). The actors were instructed to portray twelve "neutral" sentences with six emotions: *anger, disgust, fear, neutral, happy,* and *sad.* The sentences used in the database were the following: (1) Don't forget a jacket. (2) It's eleven o'clock. (3) I'm on my way to the meeting. (4) I think I have a doctor's appointment. (5) I think I've seen this before. (6) I would like a new alarm clock. (7) I wonder what this is about. (8) Maybe tomorrow it will be cold. (9) The airplane is almost full. (10) That is exactly what happened. (11) The surface is slick. (12) We'll stop in a couple of minutes.

For the purposes of this study, 12 sentence performances were chosen as the stimuli. Only four of the six emotional categories of speech were used, selected to represent a continuum of both arousal and valence scores: namely, fear (low valence, high arousal), happy (high valence, high arousal), sad (low valence, low arousal), and neutral (neutral valence, medium arousal). The sentences with the highest CREMA-D agreement scores for each emotion category resulted in the final stimuli list of 12 stimuli (3 'fear', 3 'happiness', 3 'sad', and 3 'neutral' performances).

Musical stimuli. The musical stimuli were drawn from excerpts of film soundtracks (curated by Eerola & Vuoskoski, 2011). The selected excerpts have been empirically shown to be unfamiliar to Western-enculturated participants of a similar age to those participating in the current experiment, and so minimizes confounds with possible episodic memories. Each excerpt is between 10 and 15 seconds in duration and represents one of five discrete emotions: *fear, anger, sadness, tenderness,* and *happiness.* Participants in Eerola & Vuoskoski's (2011) study showed high agreement in choosing the categorical emotion for each of these excerpts. Four of these categories were chosen in order to represent the four arousal-valence quadrants of the Russell et *al*. (1989) affect grid: fear (low valence, high arousal), happiness (high valence, high arousal), sadness (low valence, low arousal), and tenderness (high valence, low arousal). In this study, three excerpts were chosen per categorical emotion (3 'fear', 3 'sadness', and 3 'happiness' passages). However, six tenderness passages were used because there were no high valence, low arousal speech samples.

Experimental Procedure. Participants were reminded about the difference between perceived and felt emotion and then were asked three questions. The first two of these three questions aim to probe the perceived valence of the sound; the third question aims to identify the perceived emotional arousal of the sound. Perceived valence was separated into two unipolar scales because previous research has shown that positivity and negativity are separable in emotion experiences (Larsen & McGraw, 2011). The first question was, "*To what extent does this audio file sound positive?*" (11-point Likert scale from 0 'not at all positive' to +10 'extremely positive'); the second question was, "*To what extent does this audio file sound negative?*" (11-point Likert scale from 0 'not at all negative to +10 'extremely negative'; the third question was, "*To what extent does this audio file sound energetic/arousing?*" (11-point Likert scale from 0 'this sound represents no energy/arousal' to +10 'this sound represents an extreme amount of energy/arousal').

Next, participants were asked to "*Identify which emotion(s) the audio file represents by checking the appropriate emotion(s) from the following list. You may select as few or as many as you like.*" Once participants finished selecting the emotion terms, the list of terms they chose reappeared in isolation on the screen. They were then given the following instructions: "*Given this list of emotion terms you chose, which one(s), if any, strongly apply?*" By asking participants to choose emotion terms that strongly apply, a three-level response gradient is available for analysis (i.e. does not apply, applies, strongly applies).

The questions about emotion categories are exploratory questions, whose aim is two-fold: first, to identify possible affective confusions that might arise due to the effect of the acetaminophen; second, to test whether acetaminophen effects are different for separate emotions (hypothesis 3).

The emotion choices listed in the fourth question were inspired by the music cognition literature and the speech literature (both perceived and induced emotion). First, emotions were listed that were the intended emotion of the musical and speech stimuli (Cao et *al.*, 2014; Eerola & Vuoskoski, 2011): *angry, disgusted, fearful, happy, sad, tender*. The term *grieved* was added, as grief is thought to be related—but distinct—from musical sadness (Huron, 2015; Warrenburg & Léveillé Gauvin, 2017). The emotion terms *bored* and *relaxed* were added in order to investigate whether sad or tender sounds could be confused with boring or relaxing sounds (Huron, Kinney, & Precoda, 2006). Finally, the term *surprised* was used, as music and speech are known to be surprising when they do not conform to our enculturated expectations (Huron, 2006). Although these emotions span the four quadrants of a typical circumplex model, there is a bias towards high arousal-negative valence emotions. The terms *excited* and *invigorated* were therefore added to balance the number of terms in each quadrant (King & Meiselman, 2010). The final list of 14 emotions was the following: *angry, bored, disgusted, excited, fearful, grieved, happy, invigorated,* *relaxed, sad, surprised, tender, neutral/no emotion, other emotion(s).*

Finally, participants were asked to indicate their degree of familiarity with the music files using a three-point scale (0 = not familiar, 1 = somewhat familiar, 2 = very familiar). After completing the first block, participants were given a short break.

Block 2: Induced Emotion

The purpose of the second block was to examine the *induced* emotion from 18 sound stimuli, with a focus on music-induced emotions. Specifically, we measured the magnitude of listeners' emotional responses to the audio stimuli. The musical and natural sound stimuli in this block were different from the musical stimuli in the first block. The natural sounds were included because it was thought that they might induce stronger emotions than the musical samples (e.g., a person is likely more likely to respond to an audio file of domestic abuse than to musical passages). The speech samples were not included in this block, as the files were intended to examine perceived emotion, rather than to arouse emotion in listeners (Cao et *al.*, 2014; Keutmann et *al.,* 2015).

Stimuli. The primary aim of the second block was to examine whether acetaminophen can reduce music-induced emotions. Music is known to be able to induce strong emotion in listeners. Choosing passages of music that will reliably induce a specific emotion in listeners, however, is difficult. In the first instance, it is difficult to identify passages that might be considered to evoke a single affect. Even works that are typically thought to be affectively homogenous can represent and evoke more than one emotion. For example, Samuel Barber's *Adagio for Strings* is widely considered to be a quintessentially "sad" work. When listening to this piece, however, one can hear clear shifts in the affective mood. One may call some passages *sad*, *melancholic*, and *despondent*, but describe other passages as *grief-like* or *full of despair* (Huron & Warrenburg, 2017). The audience may experience alternating emotions of *sadness*, *despair*, and *compassion* throughout the work.

A second consideration relates to musical preferences. If a listener dislikes a particular musical style, the dominant emotional experience may simply be one of *boredom*. Both of these potential confounds can be minimized through the use of shorter musical passages. It is easier to find a shorter passage of music that is affectively homogenous or that will evoke a single affect than it is to find a longer passage that accomplishes the same thing. Fortunately, previous studies have tested many passages of music and identified various passages that are effective in displaying and inducing certain emotions in listeners (e.g., Eerola & Vuoskoski, 2011).

A further problem confounding research on music-induced emotion is high between-listener variability in induced affect. Differences in emotional responses to music can be due to many factors, including episodic memory, musical preferences, current mood, trait empathy, familiarity, and age. There is no guarantee that a participant will even react emotionally to a given musical passage.

A final caveat is that music is thought to commonly induce mixed-emotions in listeners (Juslin, 2013a; Juslin, 2013b). For example, when listening to nominally sad music, many people may experience a mixture of positive and negative valence. Listeners may label these affective states differently, although there is some consensus that people may claim to feel sadness and tenderness, sadness and compassion, or sadness and beauty (Peltola & Eerola, 2016; Tarrufi & Koelsch, 2014). Music has also been known to induce nostalgia, which is characterized by a bittersweet emotion (Barrett et *al.*, 2010). The experience of mixed emotions (presumably) consists of various numbers of emotion categories. These categories may differ from person to person. An additional source of variation is that each component emotion may be experienced at different intensities. This gives rise to a large number of possibilities of emotion combinations.

Although listeners are relatively quick to identify the displayed or represented emotion in an auditory stimulus, it takes somewhat longer for an emotion to be induced in a listener. A meta-analysis of music and emotion research conducted by Eerola and Vuoskoski (2013) estimates that stimulus durations of 30-60 seconds may be necessary in order to induce an emotion in a listener. At the same time, longer musical passages are more likely to exhibit subtle or substantial shifts in affective content. Consequently, we made use of one-minute musical clips as stimuli in our induced-emotion block. Specifically, we employed a subset of stimuli used by Eerola and Vuoskoski (2013) representing scary, happy, sad, and tender feelings. These four emotion categories parallel the four emotion categories in the perceived emotion block. In addition to the music stimuli, participants listened to 10 natural stimuli, independent from the ones used in the first block. These stimuli were chosen in the same way as in the perceived emotion block.

Experimental Procedure. The participants heard the sounds in a random order. After listening to each stimulus, participants were asked the following two questions: (1) "*To what extent does this audio file make you feel a positive emotional reaction?*" using an 11-point Likert scale (from 0 'I feel little or no positive emotion' to 10 'I feel an extreme amount of positive emotion'); (2) "*To what extent does this audio file make you feel a negative emotional reaction?*" using an 11-point Likert scale (from 0 'I feel little or no negative emotion' to 10 'I feel an extreme amount of negative emotion'). Participants were also asked to identify which emotion(s) they felt by checking the appropriate emotion(s) from a list of emotions. The emotion choices included all of the 14 terms from the first block (*angry, bored, disgusted, excited, fearful, grieved, happy, invigorated, relaxed, sad, surprised, tender, neutral/no emotion, other emotion(s)*). Also included were emotions that are commonly induced by music (Zentner, Grandjean, & Scherer, 2008): *wonder, transcendent, nostalgic, peaceful, power, joyful, tension*. The term *anxious* was also added. One additional term was used in order to directly investigate the amount of empathy (or compassion) a person might feel towards the sounds (Greitemeyer, 2009): *sympathetic.* This resulted in a final list of 24 items: *angry, anxious, bored, disgusted, excited, fearful, grieved, happy, invigorated, joyful, nostalgic, peaceful, power, relaxed, sad, soft-hearted, surprised, sympathetic, tender, transcendent, tension, wonder, neutral/no emotion, other emotion(s).*

Similarly to Block 1, after participants checked which emotions they felt from the list of 24-terms, they were presented with a list of the terms they checked. From this list, they were asked to respond to the question "*From this list of* *emotions that you chose, which one(s), if any, strongly apply?"* Finally, participants were asked to indicate their degree of familiarity with the musical passages on a three-point scale (0 = not familiar, 1 = somewhat familiar, 2 = very familiar).

Preliminary Results and Conclusion

The study in currently in progress. Nineteen participants have completed the task (10 placebo condition, 9 drug condition). No statistical tests regarding a difference between the two drug conditions have been carried out, as a power analysis indicates that approximately 200 participants would be needed in order to establish a reliable effect.

For the initial 19 participants, descriptive statistics of the stimuli have been conducted by using t-tests. In the perceived emotion task, the stimuli correspond to their hypothetical quadrants on the Russell et *al*. (1989) Affect Grid. That is, stimuli that were *a priori* considered to be low-valence, high-arousal (including fearful music and speech) were rated as equally negative, but more arousing than stimuli considered to be low-valence, low-arousal (including sad music and speech) (LVHA mean negativity = 6.91, LVLA mean negativity = 6.13, $p > 0.05$; LVHA mean arousal = 5.91; LVLA mean arousal = 3.63, $p < 0.05$). Stimuli considered to be high-valence, high-arousal (including happy speech and music) were rated as significantly more positive and arousing than stimuli considered to be high-valence, low-arousal (including tender music) (HVHA mean positivity = 7.06, HVLA mean positivity = 5.59, $p < 0.05$; HVHA mean arousal = 6.72, HVLA mean arousal = 4.05, $p < 0.05$).

For the initial 19 participants, in the induced emotion task, there was no difference in the experienced negativity of the low-valence, high-arousal stimuli and the low-valence, low-arousal stimuli (LVHA mean negativity = 5.69, LVLA mean negativity 5.04, $p > 0.05$). There was a difference in the experienced positivity of the high-valence, high-arousal stimuli and the high-valence, low-arousal stimuli, with the high-arousing stimuli contributing to greater experienced positivity (HVHA mean positivity = 6.59, HVLA mean positivity = 5.22, $p < 0.05$). Complete results for the planned cohort of approximately 200 participants are pending.

References

Barrett, F. S., Grimm, K. J., Robins, R. W., Wildschut, T., Sedikides, C., & Janata, P. (2010). Music-evoked nostalgia: affect, memory, and personality. *Emotion*, *10*(3), 390-403.

Bradley, M. M. & Lang, P. J. (2007). The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. Technical report B-3. University of Florida, Gainesville, Fl.

Cao, H., Cooper, D.G., Keurmann, M.K., Gur, R.C., Nenkova, A., & Verma, R. (2014). CREMA-D: Crowd-sourced emotional multimodal actors dataset. IEEE Transactions on Affective Computing 5(4), 377-390.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10,* 85.

DeWall, C.N., MacDonald, G., Webster, G.D., Masten, C.L., Baumeister, R.F., Powell, C., & Eisenberger, N.I. (2010). Acetaminophen reduces social pain: Behavioral and neural evidence. *Psychological Science, 21,* 931–937.

Durso, G. R., Luttrell, A., & Way, B. M. (2015). Over-the-counter relief from pains and pleasures alike acetaminophen blunts evaluation sensitivity to both negative and positive stimuli. *Psychological Science, 26*(6), 750-758.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18-49.

Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal*, *30*(3), 307-340.

Greitemeyer T. (2009). Effects of songs with prosocial lyrics on prosocial thoughts, affect, and behavior. *Journal of Experimental Social Psychology, 45*(1), 186–190.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Huron, D. (2008). A comparison of average pitch height and interval size in major-and minor-key themes: Evidence consistent with affect-related pitch prosody. *Empirical Musicology Review, 3*(2), 59-63.

Huron, D. (2015) Affect induction through musical sounds: an ethological perspective. *Phil. Trans. R. Soc. B 370*: 20140098.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review 1*(3), 170-177.

Huron D. & Warrenburg, L.A. (2017). Sadness versus Grief: Has research on musical 'sadness' conflated two different affective states? Paper presented at the biannual national conference of the Society for Music Perception and Cognition, San Diego, CA.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, *2*(1999), 102-138.

Juslin, P. N. (2013a). From everyday emotions to aesthetic emotions: Towards a unified theory of musical emotions. *Physics of Life Reviews, 10*(3), 235-266.

Juslin, P. N. (2013b). What does music express? Basic emotions and beyond. *Frontiers in Psychology*, *4*(596).

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, *129*(5), 770-814.

Kaufman, D.W., Kelly, J.P., Rosenberg, L., Anderson, T.E., Mitchell, A.A. (2002). Recent patterns of medication use in the ambulatory adult population of the United States: The Slone Survey. *Jama, 287*(3), 337–344.

Keutmann, M.K, Moore, S.L., Savitt, A., & Gur, R.C. (2015). Generating an item pool for translational social cognition research: Methodology and initial validation. *Behavior Research Methods 47,* 228-234.

King, S. C., & Meiselman, H. L. (2010). Development of a method to measure consumer emotions associated with foods. *Food Quality and Perference 21*, 168-177.

Kraepelin, E. (1921). *Manic depressive insanity and paranoia.* Trans. R. Mary Barclay. Edinburgh: E. S. Livingstone.

Larsen, J. T., & McGraw, A. P. (2011). Further evidence for mixed emotions. *Journal of Personality and Social Psychology, 100*(6), 1095-1110.

Mischkowski, D., Crocker, J., & Way, B. M. (2016). From painkiller to empathy killer: acetaminophen (paracetamol) reduces empathy for pain. *Social cognitive and affective neuroscience, 11*(9), 1345-1353.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*(10), 1410-1421.

Ollen, J. E. (2006). *A criterion-related validity test of selected indicators of musical sophistication using expert ratings* (Doctoral dissertation, The Ohio State University).

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions.* New York, NY: Oxford University Press.

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and language*, *104*(3), 262-269.

Peltola, H. R., & Eerola, T. (2016). Fifty shades of blue: classification of music-evoked sadness. *Musicae Scientiae*, *20*(1), 84-102.

Post, O., & Huron, D. (2009). Western classical music in the minor mode is slower (except in the Romantic period). *Empirical Musicology Review, 4*(1), 2-10.

Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology, 100*, 1139-1157.

Ricoeur, P. (2003). *The rule of metaphor: The creation of meaning in language*. Psychology Press.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, *57*(3), 493-502.

Sandstrom, G. M., & Russo, F. A. (2013). Absorption in music: Development of a scale to identify individuals with strong emotional responses to music. *Psychology of Music*, *41*(2), 216-228.

Schirmer, A., Zysset, S., Kotz, S. A., & von Cramon, D. Y. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, *21*(3), 1114-1123.

Schutz, M., Huron, D., Keeton, K., & Loewer, G. (2008). The happy xylophone: acoustics affordances restrict an emotional palate. *Empirical Musicology Review, 3*(3), 126-135.

Taruffi, L., & Koelsch, S. (2014). The paradox of music-evoked sadness: An online survey. *PLoS One*, *9*(10), e110490.

Turner, B., & Huron, D. (2008). A comparison of dynamics in major-and minor-key works. *Empirical Musicology Review, 3*(2), 64-68.

Ververidis, D., & Kotropoulos, C. (2006). Emotional speech recognition: Resources, features, and methods. *Speech communication*, *48*(9), 1162-1181.

Warrenburg, L.A. & Léveillé Gauvin, H. *Assembling a Database of Validated Audio Stimuli: Evaluating Valence in Musical and Non-Music Sounds.* Music Informatics Interest Group of the Society for Music Theory Annual conference, Arlington, VA, November, 2017. Poster.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063-1070.

Yim, Huron, & Chordia. (MS). The effect of 'lower than normal' pitch collections on sadness judgments of unconventionally-tuned melodies.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion, 8(4), 494-521.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Some Influences of Chord Progressions On Accent

Christopher Wm. White,[1] Dakota Gill,[2] and Nicholas Shea[3]

[1,2] *Department of Music and Dance, University of Massachusetts Amherst, USA*
[3] *School of Music, The Ohio State University, USA*

[1]cwmwhite@umass.edu, [2]dagill@umass.edu, [3]shea.185@buckeyemail.osu.edu

Abstract

This paper outlines four experiments that investigate the ways that chords and chords progressions might cause listeners to hear a pulse as accented. In each experiment, listeners hear two alternating pulses and are asked to tap on the pulse they perceive as more accented. In Experiment 1, the initial pulse (what we call CASE A) sounded a major or minor triad that retained all but one of the pitch classes of the previous pulse (CASE B), while CASE B sounded a triad with two p.c.'s different from CASE A. Participants rated CASE B as accented significantly more frequently than CASE A, indicating that pitch-class change encouraged a feeling of accent. Triads have three ways they can transition to one another while changing only one pitch-class, and Experiment 2 tested whether these different progressions would be heard as differently accented. While we found no effect of transformation type, participants identified pulses hosting more major triads as the more accented. In order to determine whether particular root motions in of themselves might influence accent perception, Experiment 3 constructed a format such that CASE A would sound a major triad, and CASE B would sound a major triad rooted a third lower. Major versus minor thirds were used in different trials. No effect root motion was found. Finally, Experiment 4 alternated major triads on CASE A and minor triads on CASE B. Participants chose major triads as more accented significantly more than minor triads. These findings suggest that pitch-class change influences whether listeners perceive a chord event as relatively accented, that the chord's mode affect accents accent, but the root motion and distance by which p.c.'s change do not have a significant effect. These findings have implications on the overlap between harmony and meter, and the construction of triadic chord grammars.

.

Introduction

Do particular chord progressions encourage listeners to hear pulses as accented? Corpus analyses of Western European common-practice tonal music have shown that certain chords tend to occur on strong beats (Prince and Schmuckler 2014), while music theory literature has focused on ways that harmonic events align with and support metric accents (Yeston 1975, Lerdahl and Jackendoff 1983, Mirka 2009, Aldwell and Schachter, 2011). For over a century music psychologists have studied how pitch events like contour and grouping contribute to accent (Woodrow 1911, Huron & Royal, 1996, Acevedo et al., 2014), but the specific interactions of meter, harmony, and accent have begun to receive more attention in recent years. Prince, Thompson, and Schmuckler (2009), for instance, found that key orientation affected how participants metrically classified a stimulus, and Rosenthal and Hannon (2016) similarly found that metrical position affects perceptions of tonal stability in both long- and short-term memory. In terms of chordal harmony, London, Himberg and Cross (2009) found that a melody's implied harmonic prolongation influences a participant's metric interpretation of a passage; White (2017) suggests that metric accent affects how listeners prefer to harmonize a melody.

While this literature indicates that harmonic change can contribute to feelings of accent, little research has been done to pinpoint what particular attributes of harmony change are felt as accented, and whether different kinds of changes yield varying strengths of accents. This paper isolates specific aspects of harmonic change – namely chord quality, directed root motion, and the amount of pitch-class change – and tests their contribution to perceived accent by presenting listeners with alternating pulses that vary how these harmonic parameters are instantiated.

General Method

This study's four experiments follow the same overall design and participants are tasked with the same basic tapping activity. In each experiment, listeners hear two alternating pulses and are asked to tap on the pulse they perceive as more accented. Each experiment, then, varies some harmonic component and tests its effect on participant responses.

Participants

All participants were music majors, music minors, music graduate students, or otherwise self-identified as musicians at the University of Massachusetts Amherst. Participants indicated the number of years they had studied their instrument (1-5, 5-10, 10-15, 15+ years), and their gender identity. We did not exclude participants in one experiment from participating in later experiments.

Stimuli

The stimuli for each experiment were comprised of a computer-generated string of sounds that would alternate between two triads, designed to be heard as alternating strong and weak pulses. For consistency, we will call these CASE A and CASE B. All auditory stimuli were presented using in-house software written in MaxMSP (Zicarelli, 1998), using a Macbook Air, 2015 (OS X 10.10.5).

To remove any order effects, the program randomized which case was to be heard first, and the stimulus was faded in from silence into its final level of loudness over the first 10 pulses until reaching the MIDI velocity value of 100, at which it remained for the duration of the test. Participants were allowed to adjust the loudness to their comfort level during practice trials. The key orientation (or pitch-class center) of the stimulus was randomized by adding a constant integer between 0 and 11 to the MIDI notes of a particular iteration of a trial, and MIDI patch 1 (concert piano) was used. Tempo was randomized between 120 and 140 beats per minute, a tempo chosen to encourage duple groupings (London, 2004). Each note's duration was half of the inter-onset-interval (IOI).

The absolute-pitch level of the notes comprising each pulse (i.e., the octave of each pitch class within some chord) was randomized between MIDI pitch 36 (C3) and 108 (C9). Each pitch's octave was randomized independent of the other, therefore making chord inversion and voiceleading random.

Statistics

Throughout, the dependent variable was the ratio of the number of times participants chose CASE A to the number of choices made overall. ANOVAs were used to determine whether this ratio significantly changed between experimental and control/randomized conditions. *Eta*^{2} was used to estimate effect size.

Procedure

Participants were instructed to listen to the sequence and decide which pulse (CASE A or CASE B) they heard as more accented. When they had come to a decision, they were to tap along with the accent pulse on a keyboard's space bar to indicate their choice. The stimulus ended when the participant tapped 20 times.

Participants responded using an external Apple Pro keyboard and listened with noise isolation headphones (Bose QuietComfort 35 wireless). Each participant practiced the test once before trials were recorded; numbers of trials varied between experiments. Each person participated individually, and altogether the trials took between three and five minutes in total. The MaxMSP software determined whether a keystroke occurs during the IOI between when a chord begins and the initiation of the next chord, and the software recorded how often participants tapped during the IOI of each alternating pulse. A participant was recorded to have selected the pulse on which they tapped more than 50% of the time.

For comparison, each experiment also included a randomized baseline task. For statistical analysis, the initial pulse generated by the computer in these randomized tasks was tracked and heuristically equated with CASE A.

Experiment 1

Method

This experiment investigated whether participants more often tapped on pulses that sound a major or minor triad and that either introduce relatively more new pitch classes compared to the previous pulse or introduce relatively fewer new pitch classes. Participants' performance on this task was compared to a baseline/control condition in which major and minor triads were presented in a random order.

Participants. 14 male, 15 female, and 1 GenderQueer student participated in return for a snack of their choice. Age ranged from 18 to 33 (median: 21). Each participant engaged in six trials after one practice trial, with three trials in the experimental condition and three in the control condition.

Stimuli. In the experimental condition, one pulse (CASE A) sounds a triad that shares two pitch classes with that of the previous pulse, while the alternating triad (CASE B) shares one or zero pitch classes with that of the previous triad. Major and minor triads were used, but the universe of possible triads included only those within a particular major key. Key orientation and octave level for each pitch class were randomized. Figure 1 shows a sample progression: the C-major triad first shares two pitch classes with the E-minor triad (E and G), which itself shares no pitch classes with the following D-minor triad. Given this series of alternating cases, participants would either tap on those pulses that shared more or fewer pitch classes with the surrounding triads.

The control condition used the same set of major and minor triads as used in the experimental condition but ordered them randomly, prohibiting repeats. The initial pulse generated by the computer was tracked and heuristically equated with the pulse changing more pitch classes, CASE A.

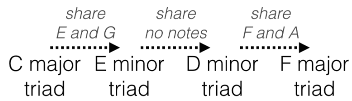


Figure 1. A sample pattern from Experiment 1's stimulus.

Results

Figure 2 shows the average participant responses divided by whether they heard the experimental or randomized condition. A one-way ANOVA using tapping preference as the dependent variable, experimental condition as the independent variable, and participant as a random factor was significant: $F(1, 27)=1.1$, $p<.01$, $eta^2 = .31$. A three-way ANOVA that includes randomized/experiment condition, gender identity and training years returns the experimental condition as a significant factor: $F(1, 45) = 11.1$, $p<.05$, $eta^2 = .20$.

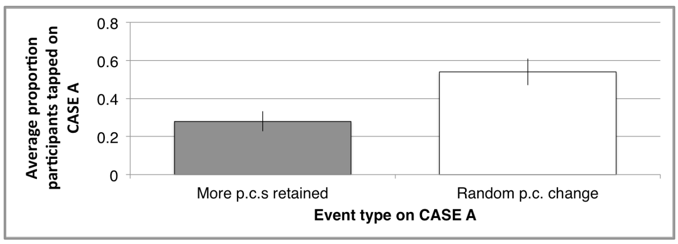


Figure 2. Responses to CASE A given the experimental and control conditions in Experiment 1.

Discussion

These results indicate that participants tap on pulses on which more pitch classes change from the previous pulse rather than on pulses on which more pitch classes are retained, with a relatively large effect size. These findings suggest that moving to a harmony that involves a change of relatively more pitch classes indicates a stronger emphasis.

Experiment 2

Method

When tracking the pitch-class change between two chords, we could not only ask *whether* a pitch class changes, but also *by how much* the change occurs. This experiment tested whether progressions that change the same number of pitch-classes but change those pitch classes by different intervals would be heard as differently accented.

Triads have three ways they can transition to one another while changing only one pitch class: moving the root up or down a third, and by switching the modal third of the triad. In some of these changes, there is a change of a whole step (i.e., moving the root of a major triad down a minor third to a minor triad and *vice versa*), while the other options change only a half step (see Cohn 1998). This experiment tested whether larger (whole-step) or smaller (half-step) changes had a stronger effect on participants' assessments of strong pulses.

Participants. 8 Females, 10 males, and 2 GenderQueer individuals participated in return for a snack of their choice. Age ranged from 19 to 28 (median: 20). Each participant engaged in ten trials after one practice trial, with five trials in the experimental condition and five in the control condition. Ordering of trials was randomized. (The increase in trial number was a result of a decreased hypothesized effect size.)

Stimuli. Participants heard alternating pulses each sounding major or minor triads. One pulse was arrived at by transposing a single p.c. of the previous chord by whole-step (this was considered CASE A), while the other pulse was arrived at by a p.c. changing by half step (this was considered CASE B). This pattern of whole-step/half-step changes results in alternating major and minor triads with roots related by alternating ascending and descending thirds; whether the pattern began with an ascending or descending third and whether the pattern began on a major or minor triad was randomized. Figure 3 shows a sample pattern. The control condition presented random major and minor triads drawing from all 24 triads within the chromatic universe.

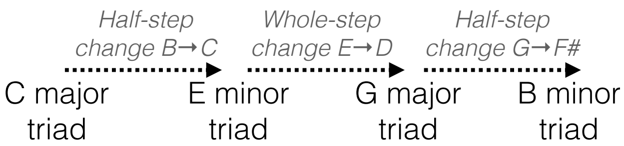


Figure 3. A sample pattern from Experiment 2's stimulus.

Results

Figure 4 shows the average participant responses divided by experimental and control conditions. A one-way ANOVA using tapping preference for CASE A as the dependent variable, experimental condition as the independent variable, and participant identity as a random factor was not significant (eta^2 = .02). A three-way ANOVA that includes randomized/experiment condition, gender identity and training years had no significant factors. However, as noted above, the pulses alternated between major and minor triads; Figure 3 also shows that participants more often selected the pulse on which the major triad occurred. (For this post-hoc analysis, the major-triad pulse was considered CASE A.) A one-way ANOVA using tapping preference for major triads as the dependent variable, experimental condition as the independent variable, and participant identity as a random factor was significant $F(1, 17) = 6.866$, $p < .05$, $eta^2 = .266$. The effect of experimental condition is also significant in a three-way ANOVA that tested experimental condition, gender identity, and training years against participant responses: $F(1, 20) = 3.83$, $p < .05$, $eta^2 = .161$. Also as noted above, pulses alternated ascending and descending root motion; however a one-way ANOVA using tapping preference for ascending thirds as the dependent variable, experimental condition as the independent variable, and participant identity as a random factor was not significant, nor was a three-way ANOVA including gender and training years.

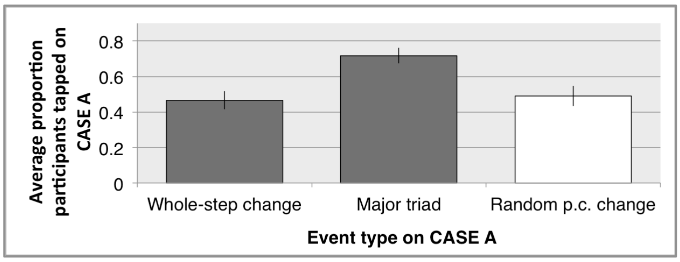


Figure 4. Responses To CASE A in the experimental and control conditions in Experiment 2.

Discussion

These results indicate that participants do not appear to tap on pulses on which more pitch classes change by whole steps versus half steps. Instead, the results appear to be explained by participants' aligning major triads with relative accent. The direction of root motion was not found to significantly affect feelings of accent. However, this last finding requires more investigation: various interlocking factors could have confounded a root motion effect. In particular, the intervals separating the chords' roots were both major and minor thirds, and both major and minor triads were used in the pattern. The next experiment, then, isolates these factors to test whether the direction of root motion by third does indeed affect the perception of accent.

Experiment 3

Method

As discussed above, triads with roots related by major and minor third not only share pitch classes, but the pitch classes that change do so by half or whole step. This experiment isolated the effects of these third relations on feelings of accent by testing whether participants would tap more consistently on a pulse arrived at by a root motion descending or ascending a major or minor third.

Participants. 17 Females, 22 males, and 1 GenderQueer respondent participated in return for a snack of their choice. Age ranged from 18 to 37 (median: 20). Each participant engaged in ten trials after one practice trial, with five trials in the experimental condition and five in the control condition. Ordering of trials was randomized. 20 participants were each randomly allocated to the major- and minor-third conditions.

Stimuli. Participants heard alternating pulses such that one pulse sounded a major triad (CASE A), while the other pulse sounded a major triad rooted a third above (CASE B). In the experimental condition, participants either heard the chord roots separated by a major or minor third (i.e., third quality was constant throughout a single participant's trials). Whether the pattern began with an ascending or descending third was randomized. Half of the participants heard triads rooted a major third apart; the other half heard triads rooted a minor third apart. The control condition randomly presented major triads from the chromatic universe.

Results

Figure 5 shows the average participant responses divided by experimental and control conditions, first separating by major and minor thirds (in grey) and then pooling the experimental condition into a single category to compare with the randomized condition (in white). A one-way ANOVA using tapping preference for CASE A as the dependent variable, interval type (major/minor/random) as the independent variable, and participant identity as a random factor was not significant (eta^2 = .024). A similar ANOVA that used the experimental/random conditions as the independent variable was also not significant (eta^2 = .018). A three-way ANOVA that tested experimental condition, gender identity, and training years against participant responses returned no significant factors.

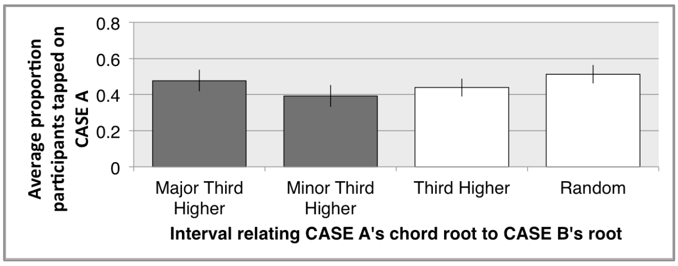


Figure 5. Responses To CASE A in the experimental and control conditions.

Discussion

These results indicate that pulses alternating two major triads rooted a major or minor third apart do not consistently encourage a feeling of relative accent. This finding suggests that the root direction of third-based chordal relationships has little to do with listeners' assessments of accent. It should be noted, as shown in Figure 5, that participants did tap less on the triad rooted a third lower (particularly rooted a minor third lower), just not significantly so. It is possible, then, that this root motion does encourage a feeling of accent, but with an effect size smaller and more fragile than could be detected in this experiment.

Experiment 4

Method

This experiment isolated the modal component of triadic progressions by testing whether participants would more consistently tap on pulses sounding major triads or pulses sounding minor triads.

Participants. 9 Females and 11 males participated in return for a snack of their choice. Age ranged from 18 to 34 (median: 21). Each participant engaged in ten trials after one practice trial, with five trials in the experimental condition and five in the control condition. Ordering of trials was randomized.

Stimuli. Participants heard alternating pulses such that one pulse sounded a major triad (CASE A), while the other pulse sounded a minor triad (CASE B). The control condition randomly presented major and minor triads from the chromatic universe.

Results and Discussion

Figure 6 shows the average participant responses within the experimental and randomized conditions. A one-way ANOVA using tapping preference for major triads (CASE A) as the dependent variable, experimental condition as the independent variable, and participant identity as a random factor was significant: $F(1, 19) = 6.16$, $p < .05$, $eta^2 = .245$. A three-way ANOVA that tested experimental condition, gender identity, and training years against participant responses returned no significant factors. These results indicate that participants prefer to align accented pulses with major triads relative to minor triads. Notably, though this experiment's major/minor alternation mimics that of Experiment 2, the effect size is smaller and the significance is more fragile.

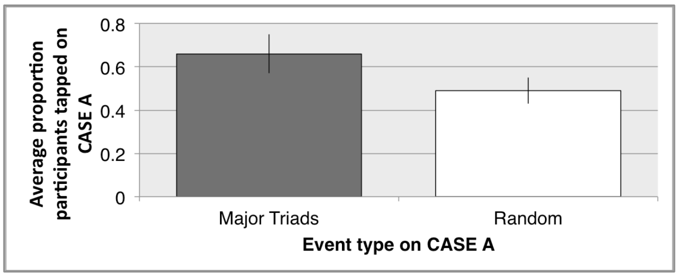


Figure 6. Responses to CASE A in Experiment 4's experimental and random conditions.

Conclusion

This series of experiments investigated several parameters of chord progressions that might contribute to a relative feeling of accent. Experiment 1 showed that triads that change relatively more pitch classes are assessed as relatively more accented, while Experiment 2 found that the distance of these pitch-class changes (specifically, half-step versus whole-step changes) did not consistently encourage a feeling of accent. The experiment also raised the question of whether chord qualities – specifically major versus minor triads – and directions of root motion – specifically, third relations – might contribute to feelings of accent. Experiment 3 therefore tested whether ascending or descending major or minor thirds contributed to feelings of accent: no effects were found. Experiment 4 then tested whether pulses sounding major or minor triads were more frequently chosen as strong beats: major triads were more often chosen.

These findings show ways that particular harmonic qualities and particular harmonic changes contribute to accent. In aggregate, this study suggests that participants prefer to align accent with pulses that both a) change more pitch classes from adjacent pulses and b) sound major triads more often than minor triads.

Some more subtle interpretations can be read from our data as well. For instance, comparing the results of Experiment 2 with Experiment 4, the preference for accented major triads appears to somewhat diminish in a less structured context. This suggests that the preference for aligning accent with major triads might be more strongly activated by chord progressions that are more similar to (and ecological approximations of) music with which our participants are familiar. Additionally, the strongest effect size was seen in Experiment 1: this suggests that the most robust conveyor of harmonic accent may be pitch-class change.

In Experiment 3, we also saw a smaller and insignificant effect of root motion on accent assessment. However, Figure 5 visually suggests that when triadic roots are related by minor third, the higher of the two is felt as accented: if the same arose with a doubled participant size, the mean would be considered significantly different. It is possible, then, that triads rooted a minor third higher than the surrounding triad are heard as very mildly accented.

Provocatively, both meter and chord progressions are linear events, and this study suggests that they might interact to determine the kinds of pitch events that occur adjacent to one another. It is possible that preferences to align certain harmonic events with accented pulses and other events with unaccented pulses could contribute to what chords follow other chords in tonal-music grammars. In other words, it is possible that conventional tonal harmonic grammars could be at least partially predicted by alternations of strong/weak pulses that change relatively more pitch classes on – and align more major triads with – accented pulses. This kind of speculation remains for future research.

While this study pinpoints certain harmonic parameters associated with strong beats, it does not determine *why* those associations exist. It is possible that humans are more predisposed to hear certain structures as aligning with strong beats (Large et al. 2016), but it is also possible that we learn expectations about these harmonic/accent expectations by being exposed to repertoires with these associations embedded within them (Saffran 1999, Prince, Thompson, & Schmuckler 2009, Louie et al. 2010, Louie 2012).

Finally, it should be acknowledged that meter is a complex phenomenon that has various interlocking levels (Cohn 2001, London 2004) and various interacting and competing parameters (Lerdahl and Jackendoff 1983, Krebs 1999). While this paper studies consistently and regularly accented pulses, it remains for future studies to investigate the complexities of meter, accent, and their relationships. However, we believe our current design highlights several ways in which harmonic events can express consistent and discernable accents.

Acknowledgements. The authors wish to thank Christopher Pandiscio and Jonathon Williams for developing earlier versions of these experiments.

References

Acevedo, S., Temperley, D., & Pfordresher, P.Q. (2014). Effects of metrical encoding on melody recognition. *Music Perception, 31(4),* 372- 386.

Aldwell, E., & Schachter, C. (2011). *Harmony and voiceleading* (4th ed.). Boston: Schirmer.

Cohn, R. (1998). An Introduction to Neo-Riemannian Theory: A Survey and Historical Perspective, *Journal of Music Theory*, *42/2*, 167–180.

Cohn, R. (2001). Complex hemiolas, ski-hill graphs and metric spaces. *Music Analysis, 20(3*), 295-326.

Huron, D. & Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Perception, 13,* 489-516.

Krebs, H. (1999). *Fantasy pieces: metrical dissonance in the music of Robert Schumann.* New York: Oxford University Press.

Large, E. W., Kim, J. C., Flaig, N. Brahucha, J. & Krumhansl, C.L. (2016) A neurodynamitc account of musical tonality. *Music Perception. 33(3),* 319-331.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: The MIT Press.

London, J., Himberg, T., & Cross, I. (2009). The effect of structural and performance factors in the perception of anacruses. *Music Perception*, *27(2*), 103–120.

Loui, P**.**, Wessel, D. L., & Hudson Kam, C. L. (2010). Humans Rapidly Learn Grammatical Structure in a New Musical Scale. *Music Perception*, *27(5)*, 377-388

Loui, P**.** (2012). Learning and liking of melody and harmony: Further studies in artificial grammar learning. *Topics in Cognitive Science, 4,* 1–14.

Mirka, D. (2009). *Metric Manipulations in Haydn and Mozart: Chamber Music for Strings, 1787–1791,* New York: Oxford University Press.

Prince, J.B., Thompson, W.F., & Schmuckler, M.A. (2009). Pitch and time, tonality and meter: how do musical dimensions combine? *Journal of experimental psychology*, *35(5)*, 1598-161.

Prince, J.B. & Schmuckler, M.A. (2014). The tonal-metric hierarchy: a corpus analysis. *Music Perception*, *31(3)*, 254-270.

Prince, J.B., W.F. Thompson, & M.A. Schmuckler. (2009). Pitch and time, tonality and meter: how do musical dimensions combine? Journal of experimental psychology, *35(5)*, 1598-161.

Rosenthal, M. A. and Hannon, E. H. (2016). Cues to perceiving tonal stability: The role of temporal structure. *Music Perception, 33,* 601-612.

Saffran. J.R, Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). "Statistical learning of tone sequences by human infants and adults," *Cognition, 70*, 27–52.

Woodrow, H. (1911). The role of pitch in rhythm. *Psychological Review, 18,* 54-77.

White, C. (2017). Relationships Between Tonal Stability and Metrical Accent in Monophonic Contexts. *Empirical Musicology Review, 12/1-2*, 19-37.

Yeston, M. (1976). *The Stratification of Musical Rhythm*. New Haven: Yale University Press.

Zicarelli, D. (1998). An extensible real-time signal processing environment for Max. Paper presented at the International Computer Music Conference, University of Michigan, Ann Arbor, MI.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Limitations of the Study of Harmony as Tension

Mauro Orsini Windholz[1]

[1]*Escola de Música, Universidade Federal do Rio de Janeiro, Brazil*

[1]maurowindholz@gmail.com

Abstract

In this paper, I point out some limitations of one of the main current approaches to investigating how music and harmony communicate what might be called "expressive meaning" (Larson, 2012) in the field of music cognition, namely, the approach of musical expectations and tension. First, drawing from concepts stemming from the embodied cognitive sciences, I argue that this approach is based on a conceptual metaphor of harmony as tension. Then, I present an overview of the presence of this metaphor in recent research in music cognition, and attempt to point out a few shortcomings this literature has had in accounting for the expressive meanings of musical and harmonic materials. Then, I bring forth evidence from the field of music and emotion of the limitations of this approach, and that present aspects of the harmonic experience that seem not to be fully explained by the metaphor of tension, such as the emotional qualities conveyed by single chords. Lastly, I suggest alternative metaphors for the understanding of harmony that might guide future research in accounting for aspects of the harmonic experience that seem not to be in the scope of the metaphor of tension. The alternative metaphors I discuss are: forces, color and gesture.

Introduction

One of the main approaches to studying expression or arousal of affect and/or meaning by music in recent work in the field of music cognition is through the study of musical expectations and tension. The ideia that "The rising and falling of tension is fundamental to the experience of music and may contribute to the emotional response" (Krumhansl & Cuddy, 2010, p. 76) is very common in research in the field, and has prompted many investigations into how musical, and specifically harmonic materials, evoke tension. However, albeit its significant contributions, it is not entirely clear from this literature how exactly do expectations and tension contribute to what might be called the many "expressive meanings" (according to Larson, 2012, p. 4: "that quality we experience in music that allows it to suggest (for example) feelings, actions and movements") of music and specific musical and harmonic materials.

In this article, I intend to point out some limitations of this approach in accounting for the expressive meanings of musical-harmonic materials (such as chords, chord progressions, key-areas, etc.). First, drawing from concepts stemming from the embodied cognitive sciences, I point out that this approach might be understood as being based on the conceptual metaphor HARMONY IS TENSION AND RELAXATION. Then, I discuss how this metaphor acquired great importance in the field of music cognition, starting with the seminal work of Leonard B. Meyer (1956), through more recent empirical studies on musical tension and ending with the work of Huron (2006), and attempt to show a few shortcomings this body of work has had in enlightening how it is that music and harmony can generate expressive meanings. Then, I bring forth evidence from research in the field of music and emotion that reinforces my claim about the limitations of this approach, and that reveal some aspects of the harmonic experience that seem not to be accounted for by this metaphor. Lastly, I suggest that future research on the subject investigate alternative metaphors for the understanding of harmony, in order to account for aspects of this experience that seem not to be accounted for by the metaphor of tension.

Harmony as Tension

It is common-place among musicians and music scholars to refer to musical-harmonic materials such as chords and chord progressions as having varying degrees of "tension". It is often said that certain chords instill "tension" that is resolved or relaxed by subsequent chords. According to the embodied cognitive sciences, such linguistic expressions can serve as clues as to how a determined experience is being understood and structured, and more often than not, they are being structured in the terms of an experience of a different kind. In this case, we are understanding chords and chord progressions in the terms of our experience of the tensing and relaxing of muscles, among other possible source-domains that involve tension. When such an understanding is employed, it is referred to as a "conceptual metaphor", which is a very pervasive and basic device in human cognition for the understanding of reality and of many human experiences (Lakoff & Johnson, 1980, 1999). Therefore, it could be said that in our musical culture, we understand harmony according to the conceptual metaphor HARMONY IS TENSION AND RELAXATION.

This metaphor has been present in the modern tradition of Western music theory at least since Rameau, as when, according to Ferris (1959, p. 243) he describes the progression between the chords of the dominant-tonic and tonic as proceeding "from tension to repose". The metaphor is also prominently present in the works of influent theorists such as Riemann (1893/1896) and Schenker (1935/1979), and many others. More recently, the metaphor has structured elaborate theories and empirical research in the field of music cognition, and has aquired great importance in the study of musical emotions and musical generation of meaning and affect. In the next section, I discuss the presence of the metaphor in the field of music cognition, and attempt to show that it presents a few shortcomings in contributing to the investigation of the expressive meaning of musical and harmonic materials.

The Metaphor of Tension in Music Cognition

The importance the metaphor of tension has acquired in research in the field of music cognition grew out in part of the influence of the work of Leonard B. Meyer. In his book *Emotion and Meaning in Music* (1956), Meyer argued that emotion and meaning are perceived in response to a musical stimulus as it plays with the expectations of listeners that are familiar with the particular musical style. The author states that:

> If (...) the sound succession fails to follow its costumary course, or if it involves obscurity or ambiguity, then it can be assumed that the listener's tendencies would be inhibited or otherwise upset and that the tensions arising in this process would be experienced as affect (Meyer, 1956, p. 31).

Therefore, in Meyer's theory there is an important link between expectations, tension and affect. In the folowing passage, this link is also evident, and meaning is added to the equation:

> Thus, in a very general way, expectation is always ahead of the music, creating a background of diffuse tension against which particular delays articulate the affective curve and create meaning (Meyer, 1956, p. 59).

After Meyer's work, the notion that emotional responses to music and understanding of meaning in musical stimulus are intimately linked with expectations and tension became considerably widespread in the field of music cognition. This prompted a number of empirical studies that attempted to measure listeners experience of tension in music hearing tasks, such as Nielsen (1983), Krumhansl (1996), Bigand, Parncutt & Lerdahl (1996), Bigand & Parncutt (1999), Lerdahl & Krumhansl (2007) and Farbood (2016), most of which focussed on the tension caused by harmonic materials. However, it is not clear from this literature how the different levels of tension verified in response to certain musical materials contribute to the specific affects and meanings that might be experienced/understood in response to these stimuli. The results these studies have presented relate mostly to the memory listeners have for tonal centers and to how well do some recent theoretical models predict the levels of tension measured in the experiments. Therefore, even though it is widely accepted that expectations and tension are fundamental to the musical experience, we still don't understand very well how exactly, and to what extent, do these factors contribute to the specific expressive meanings of specific harmonic materials, which is something that could be expected to be accounted for by the theory proposed by Meyer.

David Huron attempted to address this problem in his book *Sweet Anticipation: Music and the Psychology of Expectation* (2006), in which he sought, among other objectives, to expand the work of Meyer in light of new evidence and developments supplied by the psychological sciences. The author attempted, for instance, to account for what he called the specific "emotional *qualia*" of each scale degree of the Western chromatic scale with his take on the theory of musical expectations. Huron's account, however, could be said to be too reliant on cultural convention and statistical learning, falling prey to what Larson (2012) dubbed the "single mechanism fallacy". Although learning and culture certainly play an important role in musical experience, by assuming that most associations between musical material and expressive meaning are the product of repeated exposure of listeners to arbitrary musical conventions, Huron dismisses more in depth discussions about how do the particular characteristics of musical materials contribute to such associations. Therefore, once again, the theory of musical expectations and, by extension, the metaphor of tension, seem to fall short of enlightening how it is that specific expressive meanings come to be associated with specific musical and harmonic materials. In the next section, I bring forth evidence from research in the field of music and emotion that reinforces the limitations the metaphor has had in this investigation, and that present aspects of the harmonic experience that seem to not be completely accounted for by the metaphor.

Evidence

Juslin & Västfäll (2008), researching emotional responses to music, identified six mechanisms of emotion induction in listeners by music: brain stem reflexes, evaluative conditioning, emotional contagion, visual imagery, episodic memory and musical expectancy (Juslin & Sloboda, 2013, also include the mechanism of rhythmic entrainment). The authors contend that all of these mechanisms have an important role in the emotional responses listeners have to music, and caution that music theorists and other researchers might have been ascribing too much importance to the mechanism of musical expectancy. According to the authors:

> empirical evidence from a number of studies indicates that those emotions that we would expect from the musical expectancy mechanism, based on previous and curent theoretical models (Meyer, 1956: apprehension/anxiety, p. 27, hope, p. 29, diseppointment, p. 182; Huron 2006: surprise, anticipation, awe, boredom, p. 356) occur only rarely in listeners' emotions to music (Juslin & Västfäll, 2008, p. 604).

This is evidence that, although expectations and tension do seem to play a part in musical expression, they might not have the all-encompassing role authors such as Meyer, Huron, Krumhansl and others have been attributing to it. Juslin and Västfäll further argue that this prominence given to the mechanism of musical expectancy might in part be due to a biased view of musicologists and music theorists that "spend much of their time analysing musical scores" (Juslin & Västfäll, 2008, p. 604). As I noted earlier, the idea that tension and expectations are related to certain harmonies and chords, and also other musical elements, has a long tradition in music theory, and the focus of more recent authors in this notion might reflect their training in music theory and analysis. As Juslin and Västfäll state: "We get the impression that they sometimes forget that not all listeners hear music the same way as music theorists do" (Juslin & Västfläll, 2008, p. 604).

Another evidence of the limitation of the metaphor of tension in accounting for the expressive meaning of musical and harmonic materials is supplied by an empirical study by Lahdelma and Eerola (2016). In this study, the authors found that single chords heard in isoloation consistently convey distinct emotional qualities to listeners both nonmusicians and with several degrees of musical training. For example, major triads were consistently associated with happiness, minor triads with sadness, diminished and augmented triads were associated with tension, and major seventh chords with nostalgia.

The literature on musical expectatios rarely addresses the expressive qualities of single chords, probably because it has a hard time dealing with musical phenomena that doesn't necessarily involves any expectations for continuation or closure, such as chords heard in isolation. Nevertheless, as this study argues, such musical materials can come to have meaning and affect even when removed from a larger musical context, and the metaphor of tension doesn't seem to account for this (with the expeption of the diminished and augmented triads, which were associated with tension and might have evoked expectations for resolution even when heard in isolation).

To be sure, Meyer did try to provide an explanation of the widespread association between major harmonies and happiness and minor harmonies and sadness by arguing that positively valenced emotions are experienced more often and thus become associated with the more often heard major harmonies, and that the opposite association would occur in the case of minor (Meyer, 1956, p. 227). This explanation is very much in line with Huron's approach involving statistical learning, and might thus once again be subject to the "single mechanism fallacy", attributing the association between harmonic material and meaning/affect solely do culture and convention. Additionaly, if that were the case, we would have to admit the possibility that the reverse association might just as well have occurred: if the minor triads had been more frequent in our musical culture, they instead would have been associated with positively valenced emotions. As did Larson (2012, p. 16) in discussing the similarly contrasting expressive meanings of the lamento bass and what he called the "hallelujah figure" (which I discuss in the next section), I do not think that this is likely. Once again, this explanation dismisses any investigation into how the musical materials themselves "afford" the possibilities for the different expressive meanings they come to be associated with.

Thus, we see that the theory of expectations and the metaphor of tension doesn't seem to completely account for the emotional qualities of major and minor harmonies and of several types of single chords heard in isolation. It is therefore evident that, depite it being one of the main approaches for investigating the relation between emotion, meaning and music, this metaphor does not explain many associations between musical material and expressive meaning. It is perhaps necessary to adopt different approaches to this investigation that understand harmony in different terms, in order to account for other of its aspects and advance our knowledge of this process. In the final section, I discuss some suggestions of different metaphors that might guide future research on the subject.

Alternative Metaphors for Harmonic Understanding

According to the embodied cognitive sciences, every conceptual metaphor necessarily highlights certain aspects of the experienc it attempts to structure, while at the same time "hiding" others (Lakoff & Johnson, 1980). The metaphor of tension for the understanding of harmony could be said to be highlighting mainly its syntactycal aspects. It rules over how harmonic materials can be organized into a coherent and inteligible discourse, which materials are to be expected (relaxed) and which ones are deviations from the norm (tense). It also establishes form and segmentation, as when the relaxations are understood as closure, a moment after which there is no expectation of continuation. This could explain why the empirical studies that employed tension measures presented results mostly related to the memory listeners had for tonal centers, which would relate to how listeners are perceiving tonal form. However, the metaphor does not seem sufficient to establish a *semantic* theory of tonal harmony, not being able to fully explain many associations between harmonic material and expressive meaning, even one as pervasisve and fundamental to our musical culture as the one between major harmonies and happiness and minor harmonies and sadness (Parncutt, 2014). It is perhaps necessary to seek out new metaphors for harmonic understanding in order to account for the more semantic aspects of this experience, and deepen our knowledge of its inner workings.

One possible such metaphor would be the one developed by Larson in his book *Musical Forces: Motion, Metaphor and Meaning in Music* (2012). In this work, Larson argued that humans hear and understand music metaphorically as movements that are constrained by forces, the character of which contribute to the many expressive meanings that come to be associated with different musical materials. In this manner, Larson was able to account for some expressive meanings of melodic figures such as the lamento bass and the "hallelujah figure".

The first one, a chromatically descending bass line in minor mode, slow tempo and triple meter (Figure 1), would thus be understood as "being pulled slowly and inevitably downward" (Larson, 2012, p. 84) by the forces of "melodic gravity" and "inertia", which could relate to feelings of sadness, that are themselves usually related metaphorically to a downwards orientation (as in expressions such as "feeling *low*" or "he *fell* into a deep deppression"). This could in part explain the widespread association in the baroque period of the lamento bass to texts that express mourning, sadness and death. In contrast, the "hallelujah figure", which consists of a leap from the first degree of a major scale to the fifth degree, followed by a stepwise motion to the sixth degree then back to the fifth, would be more commonly associated to lighter and gayer emotions. Larson presents an account of the beginning of the melody of "Twinkle, Twinkle, Little Star" (Figure 2), which contains a hallelujah figure:

> this leap suggests a quality of ease because it leaps from the most stable platform (the tonic) to the next-most-stable degree of the scale (the fifth scale degree). That ease, combined with the energy associated with an ascending leap of this size, suggests a kind of athletic quality that is effortless and secure. (...) And the simple repetition of each note in the melody gives the line a kind of simpleminded momentum that I associate with skipping motions expressive of unconcerned contentment (Larson, 2012, p. 83-84).

Therefore, the different melodic configurations of these two figures suggest different movements that are constrained by metaphorical musical forces in different ways, which in turn suggest contrasting expressive meanings.

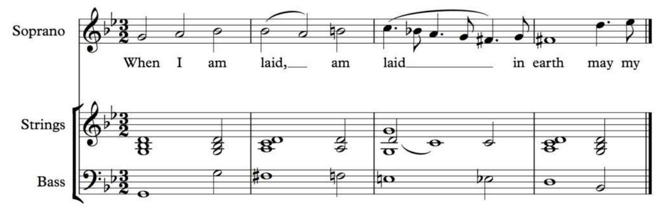


Figure 1. Begining of Purcell's "Dido's Lament", which contains a lamento bass.

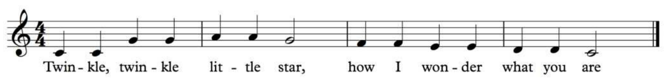


Figure 2. Begining of the melody of "Twinkle, Twinkle, Little Star". The first two measures consist of a "hallelujah figure".

Thus, with a metaphor of musical forces, Larson was able to investigate how the particular characteristics of certain melodic materials contribute to the expressive meanings they come to be associated with. Larson applies his theory mostly to the analysis of melodies and melodic contours. Perhaps one interesting avenue of investigation for future research would be applying the metaphor of musical forces to harmonic materials such as chords and chord progressions, in order to account for, for example, the different emotional qualities of single chords identified in Lahdelma and Eerola's study.

Another possible metaphor for the understanding of harmony would be *color*. Musicians often say that harmonic materials have different "colors", as when jazz musicians refer to intervalic additions to a structural chord as "coloring effects". Associations between musical passages and colors also seem to have some relation to affective response (Isbilen & Krumhansl, 2016). As LaRue notes:

> Harmonic color is in many styles the most instantaneous affective resource of music. We are immediately conscious of changes from major to minor forms of chords; open fourths or fifths vs. triads, sevenths, or higher complexities of vertical structure, open vc. close position (...); various doublings; and 'dark' (flat) or 'light' key relationships (LaRue, 1992, p. 41).

Therefore, it could be enlightening to investigate the expressive meanings of these several coloring variations of "vertical structure".

One last possible metaphor for harmonic understandig that I suggest is that of *gesture*. According to Godøy (2011), the perception of what he called coarticulated gestures in musical stimuli is fundamental to the musical experience, as they "work as schemas in the perception of several important features of musical sound", such as:

> *Tone semantic features*, meaning the emergent, contextual effects of sequentially presented tones fusing into a holistically perceived chunk having distinct modal flavours (Godøy, 2011, p. 77).

Therefore investigations on the expressive meaning of musical gestures might help enlighten the semantic qualities of harmonic materials such as modes.

All of this is not to say that tension and relaxation and expectations do not play an important role in perception and understanding of harmony, there is plenty of evidence that they do. What I point out, however, is that some recent research in music cognition might have over-emphasized this role, and that there seems to be aspects of the harmonic experience that cannot be fully explained in terms of tension and relaxation and that are not being as thouroughly investigated. With this, I hope to have suggested interesting ideas that motivate future investigations on the expressive meaning of harmony, and that may help us understand more about this very powerful, affective and meaningful experience.

References

Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research, 62*, 237-254.

Bigand, E., Parncutt, R., Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics, 58 (1)*, 125-141.

Farbood, M. M. (2016). Memory of a Tonal Center After Modulation. *Music Perception: An Interdisciplinary Journal, 34 (5)*, 71-93.

Godøy, R. I. (2011). Coarticulated Gestural-sonic Objects in Music. In E. King & A. Gritten (Eds.), *New Perspectives on Music and Gesture*, (pp. 67-82). Farnham and Burlington: Ashgate.

Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge: MIT Press.

Isbilen, E. S. & Krumhansl, C. L. (2016). The Color of Music: Emotion-Mediated Associations to Bach's Well-Tempered Clavier. *Psychomusicology: Music, Mind, and Brain, 26 (2)*, 149-161.

Juslin, P. Västfäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and brain sciences, 31 (5)*, 559-621.

JUSLIN, Patrick & SLOBODA, John. "Music and Emotion". In DEUTSCH, Diana (Ed.), *The Psychology of Music* (pp. 583-645). San Diego: Academic Press, 2013.

Krumhansl, C.L. (1996). A Perceptual Analysis of Mozart's Piao Sonata K. 282: Segmentation, Tension and Musical Ideas. *Music Percepetion: An Interdisciplinary Journal, 13 (3)*, 401-432.

Krumhansl, C. L., Cuddy, L. L. (2010). A Theory of Tonal Hierarchies in Music. In M. R. Jones, R. R. Fay, A. N. Popper (Eds.), *Music Perception* (pp. 51-87). New York: Springer.

Lahdelma, I., & Eerola, T. (2016). Single chords convey distinct emotional qualities to both naïve and expert listeners. *Psychology of Music, 44 (1)*, 37-54.

Lakoff, G., Johnson, M. (1980). *Metaphors we Live By*. Chicago and London: University of Chicago Press.

Lakoff, G., Johnson, M. (1999). *Philosophy in the Flesh*. New York: Basic Books.

Larson, S. (2012). *Musical Forces: Motion, Metaphor and Meaning in Music*. Bloomington and Indianapolis: Indiana University Press.

LaRue, J. (1992). *Guidelines for Style Analysis* (2nd ed.). Warren: Harmonie Park Press.

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling Tonal Tension. *Music Perception: An Interdisciplinary Journal, 24 (4)*, 329-366.

Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Nielsen, F. V. (1983). *Oplevelse afmusikalsk spcending*. Copenhagen: Akademisk Forlag.

Parncutt, R. (2014). The emotional connotations of major versus minor tonality: One or more origins? *Musicae Scientiae, 18 (3)* 324-353.

Rameau, J. P. (1971). *Treatise of Harmony*. Trans. P. Gosset. New York: Dover.

Riemann, H. (1896). *Harmony Simplified*. Trans. H. Bewerung. London: Augener.

Schenker, H. (1979). *Free Composition*. Trans. E. Oster. New York: Longman.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz.

Effects of background music on the mental health and task performance of office workers

Teruo Yamasaki,[1] Shigeo Matsumoto,[2] Takuya Ikegami,[2] and Kanako Morizumi[2]

[1] *Department of Psychology, Osaka Shoin University, Japan*
[2] *USEN Corporation, Japan*

[1]yamasaki.teruo@osaka-shoin.ac.jp

Abstract

Many researchers have found that music influences the mental health and task performance of the listeners. Expecting these effects, background music is frequently used in workplaces. However, such effects are not fully examined in real workplace settings. This study examined whether background music influences the mental health and task performance of office workers. Fifty office workers in an IT security company participated in this study. During four weeks, no music (first week), a standard music program for office work (second week), a sedative music program (third week), and a lively music program (fourth week) were played in the office all day long. Participants worked for a week with/without a background music program and filled out a questionnaire on Friday. The questionnaire consisted of items judging the impression of the background music played, items from the Brief Job Stress Questionnaire (BJSQ), items on the Positive Health from the Psychological Lively Scale (ver. 2), and items concerning work performance in office. The BJSQ includes five subscales; fatigue, anxiety, depressed mood, anorexia, and insomnia. The Psychological Lively Scale (ver. 2) has three subscales; emotional stability, sense of fulfilment, and vitality for living. The questions concerning work performance include four subscales; communication improvement, task performance, negative effect on task performance, and sleepiness. The Pearson's correlation coefficients between these sub-scales and the evaluation of background music were calculated. The scores on "unpleasant-pleasant" in the impression judgment of background music were used as the evaluation of background music. All subscale scores of the BJSQ had significantly negative correlations with the evaluation of background music. The scores on the task performance and communication improvement and those on the emotional stability and sense of fulfilment had significantly positive correlations with the evaluation of background music.

Introduction

Music not only attracts and moves us but also influences our state of mind and behavior. For example, music listening increases self-rated relaxation and improves subjective feelings of fatigue (Kume, et al., 2018). Listening to music leads to lower subjective stress (Miluk-Kolasa, Matejek, & Stupnicki, 1996; Linnemann, Strahler, & Urs, 2016) and decreases state anxiety (Song, Li, Zhang, Yan, Chu, Sun, & Xu, 2018). These beneficial effects suggest that music listening in the workplace can promote the mental health of employees. However, many studies have found that music listening has a negative impact on task performance. Cassidy and MacDonald (2007) showed that listening to music harmed the performance of memory task and a meta-analysis by Kämpfe, Sedlmeier, and Renkewitz (2010) revealed that background music disturbed the reading process. These negative effects may lead to the conclusion that music listening in the workplace is detrimental to productivity. Haake (2011) surveyed individual music listening in the office in the UK and found that employees listened to music for a third of their working week, seeking an improvement in their mood and focus, relaxation, a suitable atmosphere, an impetus to their creativity, and so on. Given both the beneficial and detrimental effects of listening to music in the workplace, it is important to validate the positive functions of music listening mentioned in Haake's study. Participants in Haake's study listened to music individually. However, individual music listening could cause stress to other colleagues who do not want it, disturb communication in the workplace, and be considered rude by clients. In order to avoid these disadvantages, it may be useful to play background music deliberately in the workplace. Recently in Japan, an increasing number of companies have adopted background music in the office. This study aims to investigate the effects of background music on employees' mental health and task performance in the office.

Method

Participants

Fifty participants—workers in an IT security company who worked in an open-plan office—took part in this study. The age and gender of participants were as follows: five females and nine males in their 20s, three females and twelve males in their 30s, six females and ten males in their 40s, and five males of 50 years of age or more.

Period of experiment

This study was conducted for four weeks over June and July 2017.

Stimuli

Three music programs were created from the material provided by a supplier of background music: a standard program for standard office work, a sedative program that consisted of music pieces with low arousal, and a lively program that included music pieces with high arousal. The length of each program was over a day and there were no repetitions of pieces in one day.

Questionnaire

The questionnaire consisted of four parts as follows. (1) The impression of background music was rated by 16 bipolar seven-point adjective pairs like quiet-noisy, interesting-uninteresting, warm-cold, and so on. (2) Participants' stress levels were measured by 11 items extracted from the Brief Job Stress Questionnaire (BJSQ) (Japanese Ministry of Health, Labour and Welfare, 2016). The BJSQ includes five subscales (fatigue, anxiety, depressed mood, anorexia, and insomnia) and the levels of these states of mind were rated from 1 (not at all) to 4 (nearly always). (3) Participants' mental health was measured by 14 items from the Positive Health scale of the Psychological Lively Scale (ver. 2) (Yamada, Mnematsu, & Hiyakawa, 1996). The Positive Health scale consists of three subscales (emotional stability, sense of fulfilment, and vitality for living) and the levels of these states were rated from 1 (not at all) to 4 (very). (4) Participants' work performance was rated by 8 items and measured in terms of communication improvement (e.g., I talked with my colleagues in a lively manner this week), task performance (e.g., I could concentrate on my tasks this week), negative effects on task performance (e.g., I made many mistakes in my tasks this week), and sleepiness (I was frequently struck by sleepiness this week). These items were rated from 1 (definitely not) to 6 (definitely).

In addition to these scales, empathy of participants and the perceived characteristics of the office were measured for other purposes, but those data were not included in this study.

Procedure

Participants worked as usual during experimental periods and filled in the questionnaire every Friday. The experiment lasted four weeks. No music was played in the first week, the standard program was played in the second week, the sedative program was played in the third week, and the lively program was played in the fourth week.

Results

We examined the effects of background music on all subscales. These analyses were conducted for participants who answered the questionnaire in all four weeks. Fifteen participants (eight females and seven males) met this criterion.

The impression of background music

Factor analysis was conducted to examine the structure of participants' impressions of background music. Promax rotation with the maximum likelihood method generated three factors accounting for 72.7% of the variance in ratings. Factor 1 was interpreted as reflecting the activation level of the music programs. Factor 2 was interpreted as the valence of the music programs. Factor 3 was interpreted as reflecting the brightness of the music programs (Table 1).

Table 2 shows the mean factor scores of each factor for the three music programs. In general, the perceived characteristics of the standard and sedative programs were similar, while the lively program was perceived as brighter, more active, and less pleasant than the other two programs.

The effects of background music on stress levels

In order to examine the effects of background music on stress levels, repeated measures analysis of variance (ANOVAs) for the scores of the five BJSQ subscales with the background music condition as the within-subjects variable were conducted. However, no statistically significant effect was found for any of the subscales of the BJSQ ($F(3, 42)=1.172$, n.s., $\eta p2=0.077$ for fatigue; $F(3, 42)=0.650$, n.s., $\eta p2=0.044$ for anxiety; $F(3, 42)=0.687$, n.s., $\eta p2=0.047$ for depressed mood; $F(3, 42)=0.634$, n.s., $\eta p2=0.043$ for anorexia; and $F(3, 42)=0.099$, n.s., $\eta p2=0.007$ for insomnia).

Table 1. Factor loadings from a factor analysis of adjective ratings of three background music programs

Adjective pairs	Activation	Valence	Brightness
Quiet-Noisy	-.851	.336	-.068
Fast-Slow	.848	.285	.231
Energetic-Relaxed	.821	.250	.349
Powerless-Powerful	-.793	.151	.053
Weak-Strong	-.782	-.011	.240
Lively-Deserted	.772	-.072	-.408
Unimpressive-Impressive	-.629	.171	.194
Interesting-Uninteresting	.231	.891	.091
Unpleasant-Pleasant	.181	-.835	.124
Cheerful-Gloomy	.096	.816	-.106
Dislike-Like	.154	-.761	.249
Heavy-Light	.278	-.636	.138
Warm-Cold	-.134	.242	-.680
Dark-Bright	-.265	-.362	.616
Delightful-Disagreeable	.171	.398	-.573
Eigenvalue	5.65	4.93	1.18
Percent of variance explained	36.0	30.8	6.0

Table 2. Mean factor scores of each factor for three background music programs

Factor	Standard program	Sedative program	Lively program
Activation	0.148	0.094	-0.435
Valence	-0.009	-0.060	0.124
Brightness	0.047	0.071	-0.213

The effects of background music on the Positive Health scale

The Positive Health scale of the Psychological Lively Scale (ver. 2) consists of three subscales: emotional stability, sense of fulfilment, and vitality for living. The repeated measures ANOVAs for the scores of these three subscales with the background music condition as the within-subjects variable yielded no statistically significant effects ($F(3, 42)=0.061$, n.s., $\eta p2=0.004$ for emotional stability; $F(3, 42)=1.043$, n.s., $\eta p2=0.069$ for sense of fulfilment; and $F(3, 42)=0.513$, n.s., $\eta p2=0.035$ for vitality for living).

The effects of background music on work performance

Work performance was measured by four subscales: communication improvement, task performance, negative effects on task performance, and sleepiness. The repeated measures ANOVAs for the scores of these four subscales with the background music condition as the within-subjects variable showed no statistically significant effects for communication improvement ($F(3, 42)=1.839$, n.s., $\eta p2=0.116$), negative effect on task performance ($F(3, 42)=1.608$, n.s., $\eta p2=0103$), and sleepiness ($F(3, 42)=1.135$, n.s., $\eta p2=0.075$), whereas statistically significant main effects of music were found in task performance ($F(3, 42)=3.091$, $p<0.05$, $\eta p2=0.181$). However, post-hoc multiple comparisons using the Bonferroni method revealed no significant differences between any music program pairs.

Analysis of correlation between the evaluation of background music and each subscale score

The ANOVAs showed that background music did not influence stress levels, Positive Health, and task performance of participants in this study. This null effect might be because participants' sensitivity toward or evaluation of background music varied. Therefore, next, the correlation between the evaluation of background music and the scores of the various subscales was examined. In this analysis, the scores of the unpleasant-pleasant scale were used as the indicator of participants' evaluation of background music. Pearson's product-moment correlation coefficients between the evaluation scores of background music and the value obtained by distracting the scores of 12 subscales in the first week from the scores in the background music week were calculated for participants who answered the questionnaire in the first week (no background music) and at least one of the three background music weeks. Thirty-nine participants (12 females and 27 males) met this criterion.

Table 3 shows Pearson's product-moment correlation coefficients between the evaluation scores of background music and the scores of 12 subscales and their significance levels. Here we focus on the results merging the three music programs ("total" column). For the BSJQ, the correlation between the scores of all subscales except anorexia and the evaluation scores of background music were negative and statistically significant. For the Positive Health scale, the scores of emotional stability and sense of fulfilment correlated significantly positively with the evaluation of background music. For work performance, communication improvement and task performance correlated significantly positively with the evaluation of background music and sleepiness correlated significantly negatively.

Table 3 Pearson's product-moment correlation coefficients between the evaluation scores of background music and 12 subscale scores

Subscales	Standard program	Sedative program	Lively program	Total
BJSQ				
Fatigue	-0.125	-0.476**	-0.585***	-0.363***
Anxiety	-0.252	-0.410*	-0.433*	-0.344***
Depressed mood	-0.177	-0.295+	-0.702***	-0.370***
Anorexia	-0.261	-0.126	-0.296	-0.219*
Insomnia	-0.168	-0.492***	-0.523*	-0.361***
Positive Health				
Emotional stability	0.283	0.371*	0.284	0.316***
Sense of fulfilment	0.576***	0.116	0.601***	0.464***
Vitality for living	0.228	-0.016	0.297	0.172
Work performance				
Communication improvement	0.652***	0.248	0.490*	0.485***
Task performance	0.650***	0.505***	0.395+	0.537***
Negative effect on task performance	-0.194	-0.029	-0.445*	-0.202+
sleepiness	-0.269	-0.442*	-0.558**	-0.407***

Note: *** means $p<0.001$, ** means $p<0.01$, * means $p<0.05$, and + means $p<0.1$.

Discussion and conclusion

According to the ANOVAs, in general, background music did not influence stress levels, Positive Health, and the work performance of participants. Factor analysis revealed that the impression of the standard and sedative programs differed from that of the lively program. Differences in music programs did not influence the effects of background music. However, the evaluation of background music correlated significantly with almost all subscales. Considering the signs of the correlation coefficients, participants who preferred the background music tended to show low levels of stress and high levels of Positive Health during the period when it was played. Similarly, participants who preferred the background music tended to feel that their communication and task performance in the workplace had been going well. Haake (2011) described self-reported positive functions of individual music listening in the workplace. Haake's participants seemed to have listened to their favorite music, and these results were consistent with Haake's findings.

However, that the background music could not significantly influence any subscales suggests that considerable number of participants who did not prefer the background music were influenced negatively by the background music. This means the use of background music works both positively and negatively depending on the conditions.

There are many shortcomings in this study. Firstly, all findings here were derived from self-reports. Therefore, more objective evidence is desirable in future works. Secondly, which variables make background music beneficial or detrimental must be revealed. Although this study predicted that the listener's music preference is an important variable, many types of variables remain to be examined, for example, musical features like loudness and tempo, the listener's personality traits like extraversion and neuroticism, various task features, and so on. Furthermore, it is necessary to investigate the process through which background music influences the listener.

In spite of these shortcomings, the study has important meanings for both employers and employees in that it found that background music can lower stress levels, improve mental health, and increase communication quality and task performance. At the same time, this study shows that background music must be used deliberately.

References

Cassidy, G. and MacDonald, R. A. R. (2007). The effect of background music and background noise on the task performance of introverts and extraverts. Psychology of Music, 35, 517-537.

Haake, A. B. (2011). Individual music listening in workplace settings: An exploratory survey of offices in the UK. Musicae Scientiae, 15, 107-129. doi: 10.1177/1029864911398065

Japanese Ministry of Health, Labour and Welfare (2016). Implementation manual of stress check system based on the industrial safety and health law. Retrieved from http://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei12/pdf/150507-1.pdf (in Japanese).

Kume, S., Nishimura, Y., Mizuno, K., Sakimoto, N., Hori, H., Tamura, Y., Yamao, M., Mitsuhashi, R., Akiba, K., Koizumi, J., Watanabe, Y., and Kataoka, Y. (2017). Music improves subjective feelings leading to cardiac autonomic nervous modulation: A pilot study. Frontiers in Neuroscience, 11, 108. doi: 10.3389/fnins.2018.00108

Linnemann, A., Strahler, J., and Nater, U. M. (2016). The stress-reducing effect of music listening varies depending on the social context. Psychoneuroendocrinology, 72, 97-105. doi: 10.1016/j.psyneuen.2016.06.003

Miluk-Kolasa, B., Matejek, M., and Stupnicki, R. (1996). The effects of music listening on changes in selected physiological parameters in adult pre-surgical patients. Journal of Music Therapy, 23, 208-218.

Song, M., Li, N., Zhang, X., Shang, Y., Yan, L., Chu, J., Sim, R., and Xu, Y. (2018). Music for reducing the anxiety and pain of patients undergoing a biopsy: A meta-analysis. Journal of Advanced Nursing, 74, 1016-1029. doi: 10.1111/jan.13509

Yamada, H., Minematsu, O., and Hiyakawa, A. (1996). A comparative study of positive health on normal subjects and patients with mental disorders. The Japanese Journal of Health Psychology, 9, 21-33. (in Japanese).

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Using the Three-Component Model of the Musician Definition, A Musician is Someone Who Has Six Years of Musical Expertise

J. Diana Zhang,[*1] Marco Susino,[*2] Gary E. McPherson,[#3] and Emery Schubert[*4]

[*]*School of the Arts and Media, University of New South Wales, Australia*
[#]*Melbourne Conservatorium of Music, University of Melbourne, Australia*

[1]diana.zhang1@student.unsw.edu.au, [2]m.susino@unsw.edu.au, [3]g.mcpherson@unimelb.edu.au, and [4]e.schubert@unsw.edu.au

Abstract

Researchers in behavioral, psychological and cognitive science investigate the apparent differences (usually 'advantages') musicians have over non-musicians. However, how to operationalize 'the musician' is unclear given the numerous and viscous definitions that are available. To investigate this anomaly, a range of literature covering many definitions of 'musician' was consulted. From this, the definitions were collated and used to develop the 'three-component model of the musician definition' (3CMMD). The model proposes three broad, overlapping, quantifiable ways a musician can be defined, namely (1) expertise (*e.g.*, years of formal music training), (2) self-identity (*e.g.*, I am a musician), and (3) aptitude (*e.g.*, musical potential). Using this model, a selection of music psychology literature was studied to empirically identify if any consensus exists. Thirty-nine journal articles were identified as comparing musicians and non-musicians. The results demonstrated a dominance of the 'expertise' component in the 3CMMD for determining the category of musician. Overall, a general consensus regarding the musician definition was observed as someone with at least 6 years of musical expertise. The results were consistent enough to suggest an emergent '6-year' rule used by music psychologists.

Introduction

The term 'musician', among many others, can be viewed based on having an innate musical talent (Rickard & Chin, 2017; Tan, McPherson, Peretz, Berkovic, & Wilson, 2014) or having satisfied the '10, 000 hour' practice rule (Lehmann & Ericsson, 1997). Certainly, the interest in comparing musicians with non-musicians and understanding their differential cognitive capacities has been considerable and ever growing across multiple disciplines, such as in sociology, medicine and music education (Hargreaves, Miell, & Macdonald, 2002). But interestingly, no attempts have been made to unify the many different definitions surrounding a musician. The aim of this study was two-fold: (i) to integrate the many definitions in music psychology associated with a musician into a multi-component model; and (ii) to implement the model in a study of music psychology literature in order to observe if a general consensus exists surrounding the definition of a musician.

Method

A review of literature covering a range of definitions for a musician was consulted. From this, the mechanisms governing each definition were then analyzed to give distinct and relatively independent components to the measurable concept of a musician. A larger scale investigation was then conducted over a focused area of published music psychology articles to determine how the 'musician' category was formed in studies comparing musicians with non-musicians (see Zhang *et al.*, submitted, for more details).

Results

The Three-Component Model of the Musician Definition (3CMMD)

A wide range of music psychology literature was investigated to collate the different ways in which a musician has been defined. Based on the literature, three broad aspects concerning the musician definition were identified: (1) musical expertise or skill; (2) self-identity or awareness; and (3) genetic predisposition or talent. The 'three-component model of the musician definition' (3CMMD) was then developed (Figure 1).

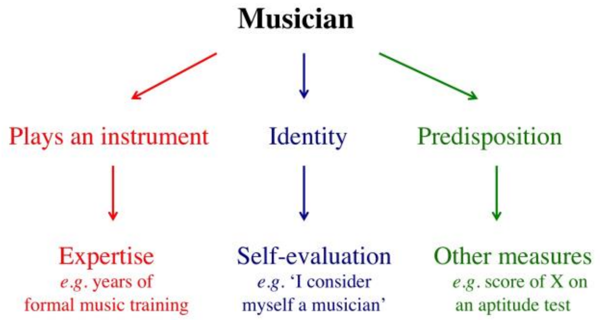


Figure 1. The three-component model of the musician definition (3CMMD).

Musical expertise component. Firstly, a musician identified through the conventional lens of expertise refers to someone who has received some degree of musical training, which enables him or her to be skilled on a musical instrument (Lamont, 2002; Sloboda, 2005). Although this is typically acknowledged and quantified through a 'years of formal music training' criterion (Rickard & Chin, 2017), other criteria such as music lessons, instrumental tuition, and musical experience have also been considered as alternate avenues where musical skill can be received (Kraus & Chandrasekaran, 2010). Evaluation of musical expertise is commonly through self-report, and although self-reports are potentially limited by misrepresentations (Chaffin & Imreh, 2001), the ease of administering a self-report question such as 'for how many years have you had formal music training' would require strong evidence against its utility to justify more complex ways of estimating expertise.

Musical identity component. Nevertheless, methods other than musical expertise have also been considered valid in estimating whether someone belongs in the category of musician. In particular, recent musicological insights have focused on the relationship between musician and identity

(Hallam, Cross, & Thaut, 2016; Hargreaves et al., 2002; Rickard & Chin, 2017). Membership of musician category determined through identity reflects the influence of social and cultural forces, which can shape and inform an individual's perception of themselves as a musician (Hallam et al., 2016). As a result, self-nomination into the musician category is often used as the criterion (Bigand & Poulin-Charronnat, 2006). While self-identity is a unique and intrinsic personal marker, with relatively greater face validity than the self-report measure of musical expertise, each individual can have different expectations about what qualifies them as having a musical identity due to personality and cultural predispositions, such as narcissism and self-effacement (Grijalva & Zhang, 2015; Kim & Chiu, 2011). Moreover, there are additional innate qualities which may not be directly accessible to the individual, such as their musical predisposition or 'aptitude'.

Musical aptitude component. A musician labeled through predisposition is an indication of the presumably natural possession of musical traits. This stems from the idea that one who may have the potential to be musical is not placed in an appropriate environment to realize that potential, or from the idea of prodigious musical talent that cannot be fully explained by environmental factors alone (for further discussion, see Gagné & McPherson, 2016). At the same time, those having had musical training may also exhibit musical inclinations. To account for this, recent developments in aptitude-based tests such as the Goldsmiths Musical Sophistication Index (Müllensiefen, Gingras, Musil, & Stewart, 2014) and MUSEBAQ (Chin, Coutinho, Scherer, & Rickard, 2018) have sought to include a combination of self-report and/or externally administered auditory tests in order to calculate the degrees of musical aptitude in addition to other components. Thus, a musician category can be formed based on the participant's performance on a continuous non-musical to musical scale.

Given the three broad and overlapping components in the literature that have been identified as accepted approaches towards the musician definition, the following study investigates how the 3CMMD aligns with the approaches used in experimental practice (see Zhang *et al.*, submitted, for more details).

Prevalence and nature of musicians in research studies

Following the development of the 3CMMD, 39 journal articles were identified as comparing musicians with non-musicians by using one or more of the components as part of their method criteria. Analysis of these studies demonstrated a significant dominance of the 'expertise' component in the 3CMMD for determining the category of musician. Furthermore, the use of comprehensive measures of musicality in the form of musical sophistication indexes or other psychometric scales occurred rarely. Overall, the results indicate a general consensus within the literature that a musician is someone who has at least 6 years of musical expertise (see Zhang *et al.*, submitted, for more details). The 6-year finding appears to be sufficiently consistent that we propose the implicit presence of a '6-year rule' used by music psychology researchers in determining the participant category of musician.

Conclusions

Music psychology literature has supported a number of acceptable musician definitions. However, until now, no attempt has been made to unify such definitions and to observe their prevalence within the music psychology research context. Here, these definitions have been summarized through the lens of the newly developed 3CMMD. Using this model, an analysis of music psychology research revealed a dominance of the 'expertise' component. Moreover, a general consensus in the literature was found to be a '6-year rule': a musician as defined by music psychologists is someone with a minimum of 6 years of musical expertise.

The present study suggests that while there is a considerable interest in the research community to investigate the advantages of being a musician, there is also significant flexibility in how a 'musician' category is formed. Therefore, the proposed '6-year rule' is intended to offer one approach to reconciling the discrepancies observed in the existing literature.

References

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*(1), 100-130.

Chaffin, R., & Imreh, G. (2001). A Comparison of Practice and Self-Report as Sources of Information About the Goals of Expert Practice. *Psychology of Music, 29*(1), 39-69.

Chin, T.-C., Coutinho, E., Scherer, K. R., & Rickard, N. S. (2018). MUSEBAQ: A Modular Tool for Music Research to Assess Musicianship, Musical Capacity, Music Preferences, and Motivations for Music Use. *Music Perception: An Interdisciplinary Journal, 35*(3), 376-399.

Gagné, F., & McPherson, G. E. (2016). Analyzing musical prodigiousness using Gagné's Integrative Model of Talent Development. In G. E. McPherson (Ed.), *Musical Prodigies: Interpretations from psychology, education, musicology and ethnomusicology* (pp. 3-114). Oxford: Oxford University Press.

Grijalva, E., & Zhang, L. (2015). Narcissism and Self-Insight: A Review and Meta-Analysis of Narcissists' Self-Enhancement Tendencies. *Personality and Social Psychology Bulletin, 42*(1), 3-24.

Hallam, S., Cross, I., & Thaut, M. (2016). *Oxford Handbook of Music Psychology* (2nd ed.). New York: Oxford University Press.

Hargreaves, D. J., Miell, D., & Macdonald, R. A. R. (2002). What are musical identities, and why are they important? In D. J. Hargreaves, D. Miell, & R. A. R. Macdonald (Eds.), *Musical Identities* (pp. 1-20). New York: Oxford University Press.

Kim, Y. H., & Chiu, C. Y. (2011). Emotional costs of inaccurate self-assessments: Both self-effacement and self-enhancement can lead to dejection. *Emotion, 11*(5), 1096-1104.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*, 599-605.

Lamont, A. (2002). Musical identities and the school environment. In D. J. Hargreaves, D. Miell, & R. A. R. Macdonald (Eds.), *Musical identities* (pp. 41-59). New York: Oxford University Press.

Lehmann, A., & Ericsson, K. (1997). Research on Expert Performance and Deliberate Practice: Implications for the Education of Amateur Musicians and Music Students. *Psychomusicology, 16*, 40-58.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE, 9*(2), e89642.

Rickard, N. S., & Chin, T. (2017). Defining the Musical Identity of "Non-Musicians". In *Handbook of Musical Identities*. Oxford: Oxford University Press.

Sloboda, J. (2005). *Exploring the musical mind*. New York: Oxford University Press.

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology, 5*, 658.

Zhang, J. D., Susino, M., McPherson, G. E., & Schubert, E. (submitted). *The definition of a musician in music psychology: a literature review and the 6-year rule*. Manuscript submitted for publication.

Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

Different Processing Mechanism of Ambiguity in Music and Language: an Empirical Approach[1]

Yifan Ivan Zou

Department of Music, The University of Hong Kong, Hong Kong, PR China
ivan_yfzou@connect.hku.hk

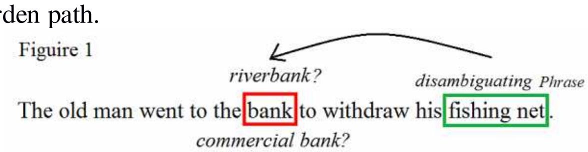
Abstract

Ambiguity is a ubiquitous phenomenon in both music and language. In the past several decades, scholars have acquired much understanding of it by examining its subtype called "garden-path ambiguity" in language. This type of ambiguity brings about several interesting questions in music: How can we define the garden-path ambiguity in music? Can we observe the garden-path effect in music? Does the human parsing mechanism (or the human parser) retain all the plausible interpretations in its initial processing (*parallel model*), or does it keep only the best-so-far interpretation at a time (*serial model*) when coping with such ambiguity? Though roughly touched upon by Meyer (1956), Bernstein (1976), Jackendoff (1991) and Temperley (2001), these questions have never been systematically theorized and empirically investigated. This paper first proposes a theoretical framework for what constitutes the garden-path ambiguity in music. Then it reports two experiments addressing the question about the parsing mechanism. In Experiment 1, a type of garden-path musical stimuli called "chimeric melodies" were presented auditorily to the participants. Results indicate that the processing of such garden-path stimuli does not cost significantly more time relative to the unambiguous controls. In Experiment 2, another similar type of chimeric melodies were presented but this time only visually in scores. In this experiment, the eye-fixation duration at the disambiguation region was significantly longer than other parts of the stimuli. Taken together, while the garden-path effect is verified in music, the processing model is dependent on the presentation modality. When presented auditorily (Experiment 1), the processing model of the garden-path ambiguity in music is best explained by a parallel model, whereas when presented visually (Experiment 2), it is best explained by a serial model. Further neuroscientific research such as EEG study is suggested for more modality-neutral comparison and more real-time information.

Introduction

Ambiguity is a ubiquitous phenomenon in both music and language, the two primary human communicative systems (Mithen 2005; Cross 2009). Indeed, ambiguity is a much advantageous feature that all efficient communication systems could possibly desire, because it fully exploits the already-informative context to free the communication from being unnecessarily crystal-clear yet redundant (Piantadosi et al, 2012). Therefore, it is crucial to understand how ambiguity is processed in both systems.

In the past several decades, scholars have acquired much understanding of the processing mechanism of ambiguity by examining its subtype called "garden-path ambiguity" in language. In the sentence (in Figure 1), for example, the highly expected interpretation of the word "bank" (financial bank) in the initial processing turns out to be erroneous as the sentence unfolds. The reanalysis is later made by reactivating (or re-ranking) an alternative meaning of the word "bank" (riverbank), which was not favored previously. Such an unexpected breakdown in the initial interpretation followed by a subsequent reanalysis is just like a wanderer trying to find another way out after realizing that he has been led down the garden path.



This type of ambiguity brings about several interesting questions: Can we find the garden-path ambiguity in music? Can we observe the garden-path effect in music? Does the human parsing mechanism (or the human parser) retain all the plausible interpretations in its initial processing (*parallel model*), or does it keep only the best-so-far interpretation at a time (*serial model*) when coping with such ambiguity? In linguistics, the eye-fixation duration at the disambiguating region (the noun phrase "fishing net" in the example above) is longer than the unambiguous control (Frazier & Rayner, 1982; Van Gompel et al 2001; Van Gompel et al 2005). Such longer eye-fixation time together with longer reaction time in other behavioral studies (where garden-path structure are involved) are taken as the crucial evidence of garden-path effect and serial model, since if all the plausible interpretations are activated and maintained throughout the processing, there would be no need to waste extra time in searching for an alternative one (Hopf et al 2003; Gibson & Pearlmutter, 2000 & etc.).

How about music? The garden-path pattern and the parallel vs. serial controversy were first introduced into music by Meyer (1956) and Bernstein (1976), and later they were theoretically explored by Jackendoff (1991) and Temperley (2001). However, the garden-path pattern in music has never been systematically theorized, nor have the garden-path effect and the parsing mechanism been empirically studied.

Theorizing garden-path ambiguity in music

Although quite a few musical examples of garden-path ambiguity can be found in Meyer (1956), Bernstein (1976), Jackendoff (1991), Temperley (2001) and Moreno (2003),

[1] The author has proposed a modified title for this conference proceeding paper after the submission. The proposed modified title is *The Comparison between the Processing Mechanism of Garden-Path Ambiguity in Music and Language: an Empirical Approach.* The reason for proposing this modified title is because the accumulative data collected during the preparation for the conference proceeding paper made the author more careful when drawing the conclusion that the processing mechanism of ambiguity in music and language are indeed different. Statistically, more subjects' data tend to make the "different claim" weaker. Therefore, the conclusion is still unclear. Before more data coming in to complete the whole picture, the author thought it is better to change the current title to the modified one.

these musical examples have never been systematically theorized. Based on what criteria, does a musical stimulus constitute the garden-path ambiguity? To theorize and define it, this paper proposes to identify the following five stages as criteria for the musical garden-path ambiguity. It should be emphasized here that these criteria are to be understood, not on the physical level, but instead, in the context of "ambiguity is a state of mind in the listener" as suggested by Meyer (1956).

1.Initial Interpretation [***MISINTERPRETATION***]: A highly expected interpretation emerges to be the only valid interpretation dominant at the beginning. However, this turns out to be a misinterpretation as the sentence unfolds.

2.Point of Error Detection [***SURPRISE***]: There must be an exact moment which clearly "tells" the music listener or the score reader that their initial interpretation is erroneous. It is called a "critical moment" or "region of disambiguation" in the linguistics, because it is the moment when the disambiguation begins and the alternative plausible interpretations begin to take shape.

3.Subsequent Interpretation [***REANALYSIS***]: After *SURPRISE*, there must be go-back procedure of self-correction. This is an involuntary/automatic brain mechanism to recover from the failure of the initial interpretation by looking for alternative interpretations.

4.Successful Reanalysis [***DISAMBIGUATION***]: There must be a word in language or a musical event, which clearly signals the moment when a successful reanalysis is achieved. At this stage, the ambiguation is completed.

5.Reflection or Appreciation [***RE-EVALUATION***]: In the end, there must be a re-evaluation of the whole process in hindsight. Multiple meanings can only be appreciated or perceived retrospectively.

These five stages (abbreviated as "***MSRDR***") are proposed as the criteria for the garden-path ambiguity in music. They are discussed in a quasi-temporal order which can supposedly capture the state of mind in music listeners or score readers.

Methods

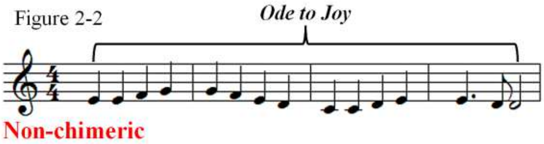
Experiment 1

Stimuli A set of "chimeric melodies" were constructed according to the above-mentioned MSRDR criteria in Experiment 1. In a chimeric melody (Figure 2-1), the tail of the first melody overlaps with the head of the second. The rationale here is that, in order to perceive the second melody *Yankee Doodle*, one has to reanalyze bar 3, which is supposed to be processed initially as part of the first melody *Ode to Joy*. Stimuli with such structure provide an ideal case for the study of garden-path ambiguity in music.

Figure 2-1

In addition to the chimeric melodies, non-chimeric melodies

Figure 2-2



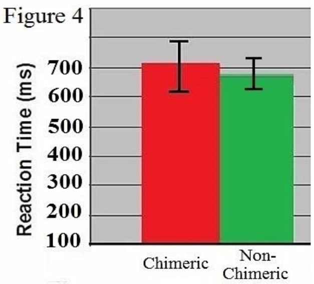
(Figure 2-2) were also presented as control, forming couplets. There were in total 24 couplets, and the starting melodies were all well-known ones (see Appendix), whereas the ending melodies for the chimeric were newly composed in a tonal manner and were taken as the "target."

Participants & Procedures. 16 music-major university students participated in the Experiment 1. In each trial (Figure 3), the target melody was first presented and the participants were instructed to memorize it. It was then followed by a longer melody (probe), which could be either chimeric (which includes the target) or non-chimeric (which is the original familiar tune). The task was to decide whether the target melody was included in the probe melody or not by speeded button pressing.

Figure 3

Results & Discussion. The overall percentage of accuracy is 89.1%(±4.4) and 94.3%(±1.7) for the chimeric and non-chimeric respectively. As for the reaction time, when only that from the accurate responses counted, it is 707ms±(89ms) and 682ms±(61ms) for the chimeric & non-chimeric melodies respectively (see Figure 4). There is no main effect of the stimulus type as indicated by the one-way ANOVA [$F(1,30)=1.63, p>.05$]. In summary, there is no significant difference in the accuracy or the reaction time between the two types of stimuli.

Figure 4



The result indicates that the reanalysis of the garden-path ambiguity in the chimeric melodies takes similar time to achieve the same level of success in processing. This is in stark contrast to the significantly more effortful processing observed in the linguistic garden-path ambiguity. Moreover, it implies that musical garden-path ambiguity is processed in a parallel model, since if only one interpretation is kept until the error is detected, the human parser will need extra time in activating an alternative one. With this respect, we might view the processing of ambiguity in musical garden path as more holistic and post-hoc than its linguistic counterpart, which means it evades the risk of misinterpretation by keeping all the valid interpretations run in the background. Such processing model is desirable when the to-be-process stimuli are indeed incorrect or misleading, because it is much securer when there is a "safety net". However, when it comes to the normal or canonical stimuli, the law of diminishing return begins to kick in: the incremental production of the sentence processing drops when it is overcautious about the possible processing failure.

Experiment 2

Stimuli The stimuli in Experiment 2 used the same 24 familiar tunes as adopted in Experiment 1 (see Appendix). Each tune (for instance, *Minuet in G major* by Bach) was then turned into a couplet, one for unambiguous control (see Figure 5-1), and one for garden-path ambiguity where the first phrase of the familiar tune was spliced into its own sequence (see Figure 5-2). There were 24 such couplets, making it 48 musical stimuli in total. All the 48 musical stimuli were randomized so that the participants could not predict what came next. All the stimuli in Experiment 2 were presented visually in musical scores. Take Figure 5-2 for instance, the rationale here is that readers were supposed to be misled into the deceptive illusion that the melody was going to unfold as the tune *Minuet in G major* by Bach upon hearing the opening phrase of it (the first 2 bars). However it turned out that their initial interpretations were wrong, and they had to reanalyze the first two bars as the motif for the subsequent sequence instead.

Figure 5-1 

Figure 5-2 

Participants & Procedures. Another cohort of 12 music-major university students participated in Experiment 2. These participants had all passed a pre-test for score reading ability. The stimuli were visually presented on a Lenovo Thinkpad 13 screen. The eye-fixation movements were recorded by a Eyelink 1000 eye tracker. The participant's eye was kept 42 cm from the screen with a visual angle of 1 °. In each trial, notes in the melodies were visually presented in a beat-wise manner and these chunks of notes were shown in a gating paradigm with an interstimulus iterval (ISI) of 750ms (Figure 6). To ensure the attentive reading and the understanding of the task, there were questions periodically after some trials, asking the participants if a continuation was sensible or if a musical phrase (which was only slightly different than the stimuli prior to it) had appeared in the previous stimuli. For instance, whether the continuation as shown in Figure 7-1 is sensible or not for the melody in Figure 5-2, or whether the musical phrase as shown in Figure 7-2 had appeared in the melody prior to it.

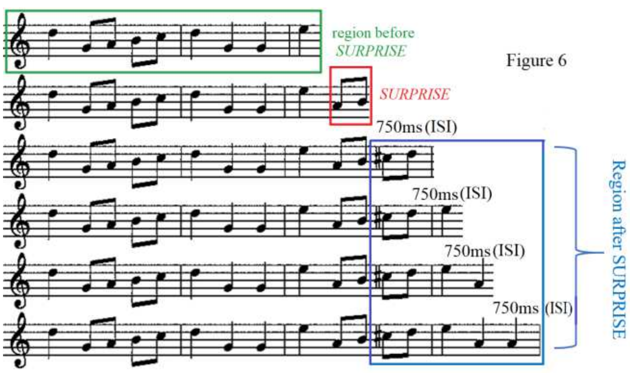


Figure 7-1

Figure 7-2

Results & Discussion. This experiment has a 2 × 2 factorial design, with one independent variable being *Region* and the other being *Ambiguity*. *Region* has to do with which portion of the musical stimuli is examined, and it has two levels (*before SURPRISE* vs. *at SURPRISE*). *Ambiguity* has to do with whether the musical stimuli is ambiguous or not, and it has two levels (*Unambiguous* vs. *Garden-path Ambiguous*). As shown in Table1, for garden-path ambiguous stimuli, the overall mean eye-fixation time per beat in the region *before SURPRISE* is 42.1 (±4.3), whereas for the region *at SURPRISE*, it is 58.7 (±6.9). A follow-up ANOVA test verifies that there is a simple main effect of *Region* on the overall mean eye-fixation time per beat in the garden-path ambiguous sentence, $F_1(1,22)=7.92$, $p<.05$. This means there is a significant garden-path effect (slow-down reading rate per beat) at the *SURPRISE* region. The result here conforms to the garden-path effect found in language as measured by the reading rate per letter (Frazier & Rayner, 1982). An ANOVA test was performed to test whether there is a significant interaction effect of *Region* × *Ambiguity*. The result indicates that there is a significant one, $F(1,22)=13.11$, $p<.05$. It is also interesting to note that the mean eye-fixation duration for the region before *SURPRISE* are qualitatively the same for the two different stimuli, F (1,22)=0.67, p>.05. It implies that the processing models for garden-path ambiguity resembles that in the unambiguous control, which is supposedly a serial type.

Table 1 Mean eye-fixation time per beat (msec) in different regions for the two different types of stimuli

Region / Ambiguity	Before *SURPRISE*	At *SURPRISE*
Unambiguous Control	41.5(5.7)	39.4(4.8)
Garden-path Ambiguous	42.1(4.3)	58.7(6.9)

Conclusion

Taken together, while the garden-path effect is verified in music, the processing model is dependent on the presentation modality. When presented auditorily (Experiment 1), the processing model of the garden-path ambiguity in music is best explained by a parallel model, whereas when presented visually (Experiment 2), it is best explained by a serial model. One potential reason why the garden-path effect did not emerge to the surface in Experiment 1 might be precisely because of this different presentation modality. Since auditory signal is ephemeral, unlike in the music score reading where eyes can go back and forth to rehearse what have just seen, later reanalysis when losing track of what have just heard will be more costly in terms of the cognitive resource. Therefore, compared to the serial (or incremental) model, the parallel (or holistic) model is more desirable where listeners do not embark on the processing immediately after receiving the signals, but instead, they wait patiently until the whole thing unfolds before digesting it in one go. In this way, the potential cost incurred by the reanalysis can be minimized. In linguistics, the garden-path ambiguity is almost always investigated in visual modality. But as we know, music is much more an auditory activity than visual. Therefore, it is more meaningful for us to probe into the processing mechanism of the musical garden-path ambiguity when it is presented auditorily. In this respect, further neuroscientific research can be conducted to solve this modality-related mystery. EEG study is highly desirable since it can monitor the brain activity when the participants either read or listen to the musical stimuli (whereas eye-tracking can only shed light on reading but not listening). Moreover, since it records what occurs in the brain in a real-time manner, therefore, it frees the monitoring procedure from being clumsy and unnecessarily task-oriented. Since many tasks in the behavioural experiments are added merely because behavioural data such as accuracy and reaction time can only emerge to the surface when certain behavioural tasks are conducted. However, the trade-off of such tasks is that they are sometimes not directly testing what the research questions is concerned about.

Acknowledgements. I am deeply grateful to Dr. Youn Kim, my PHD supervisor, who is always supportive and inspiring for my academic work, and I am most indebted to her for music theories and analyses. I would also like to thank professor William Shi Yuan Wang, with whom I have been working closely for many years since I completed my master study, and I am most indebted to him for linguistic knowledge and empirical methodology.

References

Bernstein, L. (1976). *The unanswered question: Six talks at Harvard* (Charles Eliot Norton Lectures; 1973). Cambridge, Mass.: Harvard University Press.

Cross, I. (2005). Music and meaning, ambiguity and evolution. *Musical communication*, 27-43.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178-210.

Gibson, E., & Pearlmutter, N. J. (2000). Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research*, *29*(2), 231-240.

Hopf, J. M., Bader, M., Meng, M., & Bayer, J. (2003). Is human sentence parsing serial or parallel?: Evidence from event-related brain potentials. *Cognitive Brain Research*, *15*(2), 165-177.

Jackendoff, R. (1991). Musical parsing and musical affect. *Music Perception: An Interdisciplinary Journal*, *9*(2), 199-229.

Lewis, R. L. (2000). Falsifying serial and parallel parsing models: Empirical conundrums and an overlooked paradigm. *Journal of Psycholinguistic Research*, *29*(2), 241-248.

Meng, M., & Bader, M. (2000). Ungrammaticality detection and garden path strength: Evidence for serial parsing. *Language and Cognitive processes*, *15*(6), 615-666.

Meyer, L.B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Mithen, S. (2005). *The singing neanderthals: The origins of music, language, mind, and body*. London: Weidenfeld Nicolson.

Moreno, J. (2003). Subjectivity, Interpretation, and Irony in Gottfried Weber's Analysis of Mozart's "Dissonance" Quartet. *Music Theory Spectrum*, 25(1), 99-120.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, *122*(3), 280-291.

Pickering, M. J., & Van Gompel, R. P. (2006). Syntactic parsing. In *Handbook of Psycholinguistics (Second Edition)*(pp. 455-503).

Temperley, D. (2001). *The cognition of basic musical structures*. MIT press.

Van Gompel, R. P., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in sentence processing: Evidence against current constraint-based and two-stage models. *Journal of Memory and Language*, *45*(2), 225-258.

Van Gompel, R. P., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *52*(2), 284-307.

Appendix

The 24 familiar melodies from 3 genres (nursery rhyme, folk song & classical music) of music used in experiment1&2

Nursery Rhymes	**Folk Songs**	**Classical Music**
London bridge is falling down	*Oh Susanna*	Beethoven *Ode to Joy*
Yankee doodle	*Happy Birthday*	Bizet *Toreador Song*
Oh my darling clementine	*Auld Lang Syne*	Mozart *Eine kleine Nachtmusik*
If you are happy	*Jingle Bell*	Wagner *Wedding March*
Old macdonald had a farm	*Moon River*	Bach *Minuet in G*
Twinkle twinkle little star	*Silent Night*	Strauss *The Blue Danube*
Mary had a little lamb	*Old Folks at Home*	Tchaikovsky *Swan Lake*
Frère Jacques	*Edelweiss*	Brahms *Wiegenlied: Guten Abend, gute Nacht*

Das Land Steiermark

sempre:

Society for Education, Music and Psychology Research

Université Concordia University

Karl-Franzens-Universität Graz
University of Graz

Universidad Nacional de La Plata

UNSW Sydney

ICMPC
International Conference on Music Perception and Cognition

European Society for the Cognitive Sciences Of Music

Karl-Franzens-Universität Graz
Geisteswissenschaftliche Fakultät

SMPC
Society for Music Perception and Cognition

AMPS