Parncutt, R. (2003). Accents and expression in piano performance. In K. W. Niemöller (Ed.), *Perspektiven und Methoden einer Systemischen Musikwissenschaft* (Festschrift Fricke) (pp. 163-185). Frankfurt/Main, Germany: Peter Lang.

# Accents and expression in piano performance

Richard Parncutt

## 1. Introduction

Computer simulations of expressive piano performance are becoming increasingly sophisticated. But they do not yet produce consistently satisfying results. In this regard, musical expression is lagging behind timbral synthesis; most instrumental timbres can now be convincingly synthesized by physical models (Smith, 1996). What is the problem?

Most recent computer-based attempts to create expressive musical performances (e.g., Clynes, 1985; Friberg, 1991; Honing, 1990; Mazzola & Zahorka, 1994; Sundberg, 1988; Todd, 1989) have focused on the role of musical *structure* (as defined for example by Clarke, 1988). It has been clear since (at the latest) Riemann (1884) that musicians' manipulations of performance parameters often clarify the musical structure for the listener. But two major problems arise in a computer model of music performance that is based *purely* on structure. First, there are many possible ways of representing musical structure in a computer model, and of expressing a given structure using timing, dynamics, amd so on; which, if any, of these models comes closer to the "right" representation? Second, musical structure is evidently not the *only* thing upon which interpretation depends.

Most obviously, interpretation depends on the *emotional* quality that a performer wishes, consciously or otherwise, to communicate. In an early statement of this idea, C. P. E. Bach (1753) moved away from the predominantly objectively determined *Affektenlehre* toward a more subjectively determined form of emotional communication (Kopiez, 1993). Recent research (Bresin & Friberg, 1998; Gabrielsson & Juslin, 1996; Juslin, 1997) has returned to the question of exactly what combinations of performance parameters are used to communicate which emotions.

Interpretation also appears to depend on the *dramatic* content of a piece of music, whereby "drama" may be understood to imply interactions between real people, or at least fictional characters. The psychological existence of such characters in absolute music was demonstrated by Watt and Ash (1998). Clarke (1995) attempted to explain dramatic aspects of performance expression in *semiotic* terms, citing Shaffer's (1992) proposal that a musical performance is a kind of narrative, in which *events* are determined by the musical structure, and the characters or personalities of *protagonists* are created by the performer through expressive interpretation. While this line of argument is convincing at a qualitative level, it does not immediately lend itself to application in a computer-based performance system. For this reason, it is hard to imagine how it could be systematically tested.

Yet another potentially important influence on musical interpretation may be the body of the musician - the brain, ears, lungs, lips, fingers and so on. Parncutt (1997) considered possible ways in which *physical models* could be introduced into piano performance. A possible starting point might be the model of temporal integration and integrated energy flux of Todd (1994) and Clarke (1995) – a kind of physical model of expression based on the physiology of the auditory periphery. Another possibility is music's apparent ability to imply movement (Shove & Repp, 1995) - as expressed, for example, by a conductor's baton. Todd's (1995) model of the kinematic implications of rhythm and phrasing may be regarded as a physical model of expression based on musical structure as conceived by the pianist. Similarly physical in their origins and implications are Truslit's implied motional forms (Repp, 1993), and Langner & Kopiez's (1995), Large & Kolen's (1994), and Large, Palmer, and Pollack's (1995) banks of coupled, resonating oscillators. But still a major problem remains – these ideas are hard to test systematically because of the difficulty of modeling them in a non-arbitrary way and then incorporating them into computer-based performance system.

The approach I will adopt in the present article is to return to the idea of *structure* and look at it again, from a new angle. It seems to me that a rather simple and straightforward aspect of the problem has been neglected: the relationship between *accents* and *expression*. As I will attempt to demonstrate, this aspect is already complex enough to account for a large part of the variations in timing and dynamics that make up musical expression - at least in piano performance.

## 2. Terminology for Accents in Speech and Music

At different times and in different places, music theorists have referred to many different kinds of accent, and coined a plethora of terms to describe them all. Consequently, modern music-theoretic terminology for accents can be inconsistent. Before proceeding, it is necessary to define the various accent types, and to provide supporting arguments for the terminology adopted here.

First, the term *accent* itself requires definition. An accent may be broadly defined as a relatively salient event - that is, an event that attracts the attention of a listener (Parncutt, 1987; Jones, 1987). In both speech and music, it is essential for clear understanding of the sound signal that the listener not only correctly decode the individual acoustic events (as given syllables or notes) but also get a feel for their importance relative to each other, to facilitate the inference of underlying structure and meaning. In other words, "accent" (as defined) is an integral part of acoustic communication. In this paper I will focus on accents in music, and in particular accents in piano music; but I will also compare the nature and function of musical accents with that of accents in speech.

My terminology for musical accents is set out in Table 1. It is based on that of Lerdahl and Jackendoff's (1983) *Generative Theory of Tonal Music* (henceforth *GTTM*). However, it departs significantly from their terminology in a number of respects.

| | IMMANENT ACCENTS | PERFORMED ACCENTS |
|---|---|---|
| time | - grouping<br>- metrical | - onset time (agogic)<br>- duration (articulatory)<br>- amplitude envelope |
| pitch | - melodic (contour)<br>- harmonic<br>- reductional | - intonation |
| loudness | - dynamic | - stress |
| timbre | - instrument/orchestration | - coloration |

Table 1: A Taxonomy of Accents in Music

*Immanent* accents are assumed to be apparent from the notated score; *performed* accents are effectively added to the score by a performer. Both have aspects associated with the four primary perceptual attributes of a sound: time (onset, duration), pitch, loudness, and timbre, as shown in the table. The distinction between immanent and performed relies for its existence of a notated score; the present approach is thus confined to notated music, or music that could be unambiguously notated.

I prefer the clearcut distinction between immanent and performed accents to *GTTM*'s somewhat more fuzzy distinction between *phenomenal, metrical* and *structural* accents. In *GTTM,* a phenomenal accent is defined as "any event at the musical surface that gives emphasis or stress to a moment in the musical flow"; "attack points ... sforzandi ... long notes ... leaps ... harmonic changes ..."; and as "a perceptual input to metrical accent" (p. 17). Yet it may be objected that *all* accents satisfy these criteria: all accents give emphasis, and all may conceivably play a role in the determination of perceived meter. *GTTM* further defines a *structural accent* as "an accent caused by the melodic/harmonic point of gravity in a phrase of section – especially the cadence, the goal of tonal motion" (p. 17). Later, it is noted that structural accents "articulate the boundaries of groups at the phrase and all larger grouping levels" (p. 30). Here, it may be objected that the kind of structure referred to in these definitions (the phrase or temporal group) is just one of several kinds of musical structure. A host of other accent types are directly related to other kinds of musical structure, and hence "structural". At the very least, the list includes metrical, melodic and harmonic accents; note that *GTTM* defines melodic and harmonic accents as phenomenal! One possibility would be to expand the definition of structural accents to include all these types. However it then becomes unclear which other accent types should be included as well. For example, the instrumentation of a piece of music has "structure". Should the accent that occurs when instrumentation changes be regarded as "structural"?

This raises the more general question: How "structural" are the various kinds of immanent accent? Music notation and theory suggest that pitch and time play a more important role than loudness and timbre in the building of musical structures. Thus, the most "structural" immanent accents are associated pitch and time, and pitch and time are the most important notational parameters. Two different explanations may be advanced, both of which refer to the human auditory environment. First, spectral frequencies and time intervals between sound events are normally not affected much by reflections off environmental objects in typical human environments. By contrast, spectral *amplitudes* (hence loudness and timbre) are often significantly affected by reflections. The ear seems to have adapted (evolved) to this situation by becoming more sensitive to frequency and time, and less sensitive to amplitude. Moreover, the ear is almost completely indifferent to phase relationships, because these are almost completely jumbled by reflections in typical environments and therefore carry almost no useful information about environmental sound sources (Parncutt, 1998; Terhardt, 1998). Second, the human auditory environment abounds with equally-spaced patterns of frequency (harmonic complex tones such as voiced speech sound) and time (isochronous signals such as the sound of footfalls and heartbeats) (Parncutt, 1994c), but not with equally-spaced patterns of sound amplitude (or intensity or SPL) or of spectral or temporal envelope (the physical correlates of timbre). As a result, pitch and time are the only two musical parameters that are typically perceived (or cognitively processed) relative to clearly defined reference frames (scales, metres). They are also the only musical parameters to exhibit *transpositional invariance*. To illustrate: The structural distance between the first and third beats of a 4/4 measure is, at the level of beats, the same as the distance between the second and fourth. Similarly, the structural distance between C4 and G4 is exactly the same as the distance between D4 and A4. But the distance between ***p*** and ***mp*** may not be directly compared with the distance between ***f*** and ***mf***. Similarly, the timbral distance between flute and oboe is not precisely comparable with the distance between violin and trumpet (assuming, that all instruments are playing a tone with the same pitch, loudness, and duration), although it is certainly possible to compare these distances in a more approximate way (e.g., Grey, 1977). Thus, while it is possible to construct musical forms based on loudness and timbre (McAdams & Saariaho, 1991), structures in pitch and time tend to be more complex and fine-grained.

The four main kinds of immanent accent on the left side of the table can be further subdivided. Temporal relationships in musical structures predominately involve IOIs (inter-onset times, rather than physical durations), and separate into *grouping* and *meter,* articulated by *grouping accents* and *metrical accents*. Grouping and metre are the two main hierarchical ways of organizing music in time (Cooper & Meyer, 1960; Deutsch, 1982; Drake, 1998; Handel, 1998). I follow Drake and Palmer (1993) in my use of the term "grouping accent"; it corresponds to GTTM's "structural accent". The terms *melodic* and *harmonic* accent correspond exactly to *GTTM*. A third kind of pitch-based immanent accent may be said to fall on the notes of *linear progressions* in Schenkerian analyses (non-consecutive tones that form a melodic line; see Forte & Gilbert, p. 237); I call these *reductional* (or linear) accents. Reductional accents may also be derived from *GTTM*'s prolongational reductions.

Performed accents may be classified in a similar way to immanent accents, as shown in Table 1. First, a performer may adjust the *duration* of an event by changing either its onset or its offset time. After Riemann (1884), performed accents produced by delaying or anticipating onset and offset times, or by changing inter-onset intervals (IOIs), are referred to as *agogic* accents (see also Drake & Palmer, 1993; Gabrielsson, 1974; Repp, 1990). Delaying the onset of a note relative to the prevailing metrical grid heightens expectation, increasing perceptual impact when the note finally arrives. This corresponds to the most usual use of the term agogic accent. Delaying the offset relative to the onset of the next note changes *articulation*, making the note more *legato*. Both these aspects of timing are important expressive strategies (Clarke, 1988; Palmer, 1989; Seashore, 1938; Sloboda, 1983).

A musical tone may also be made expressive by manipulating its pitch (changing its *intonation*), loudness (giving it *stress*), or timbre (changing its *coloration*). Each of these three parameters may be varied periodically during the course of the tone, producing *vibrato* and *tremolo*. On the piano, of course, vibrato is not possible; only key velocity may be varied, which simultaneously affects loudness and timbre, but not pitch.

The exact timing/duration, pitch, loudness and timbre of a notated musical event may be varied by a performer within the limits of associated *categories* of time, pitch, loudness, and timbre. Musical experience suggests that the exact boundaries of these categories are determined perceptually: no measuring instrument can predict exactly when a mistuned tone will start to sound out of tune, or under what circumstances it will be interpreted as a different tone (e.g., a semitone higher or lower), by a given set of listeners in a given musical context. Categories of pitch, time, loudness and timbre are thus (i) essentially psychoacoustic in nature and (ii) context-dependent.

The term *categorical perception* was originally invoked in speech research, and applied to the perception of phonemes (vowels, consonants) (Liberman, Harris, Hoffman, & Griffith, 1957). Categories of time, pitch, and loudness in speech may be regarded as ranges of values in which events with a given meaning are expected to occur; where these perceptual parameters deviate significantly from expected values, so too does the implicit message. The idea was later applied to the perception of musical intervals (Fricke, 1968; Burns & Ward, 1978). For example, an interval of 2.3 semitones between two tones in a tonal musical context will normally be heard to function as a major second (2 semitones) rather than a minor third (3 semitones). Immanent accents may thus define categories within which performance parameters can be manipulated.

## 3. Examples of Immanent Accents in Piano Music

### 3.1 Melodic accents

Melodic accents may be divided into *turns* and *skips* (Drake & Palmer, 1993; Thomassen, 1982; Huron & Royal, 1996). Turns are the peaks and valleys of melodic contours; the further a peak or valley lies from the mean pitch of the preceding tone(s), the more pronounced will be its melodic accent. Skips are disjunct intervals between consecutive tones: the wider the interval preceding a melodic tone, the stronger will be its accent. In unaccompanied melody, accents occurring at peaks tend to be more prominent than those at valleys – presumably because the most important phonemes in a speech utterance tend to be relatively high in pitch.

In the harmonic and polyphonic textures that are typical of piano music, melodic accents are generally stronger in the outer voices than in the inner voices, because the outer voices are generally more salient to the listener. The effect may be accounted for in the first instance by masking. Each voice partially masks other voices in its immediate vicinity, both higher and lower in pitch (Wegel & Lane, 1924). The inner voices are masked from both sides, whereas the outer voices are masked only from one side; the upper voice from below, and the lower voice from above. Things are actually considerably more complicated than this, because each voice contains several harmonic components that individually mask each other, as well as the components of other voices (Terhardt 1998). Moreover, pianists tend to exaggerate the difference in clarity between outer and inner voices by playing the outer voices more loudly – one of several ways of clarifying musical structure to the listener.

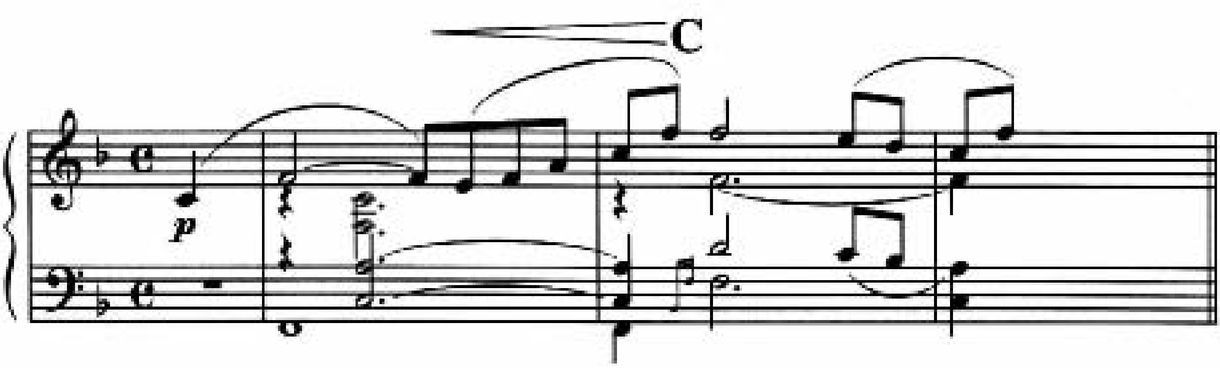
In polyphonic textures, melodic accents tend to be stronger at the peaks than at the valleys of upper-voice melodies, and at the valleys rather than the peaks of lower-voice melodies. The reason is probably that pitches at the extremes of the musical texture are more clearly audible than pitches within the texture – another effect that may be explained by masking. Thus, the further a melodic turn lies from the mean pitch of its context, the greater will be its perceived accent.

Ex. 1a. Chopin: Etude Op. 25 No. 7.

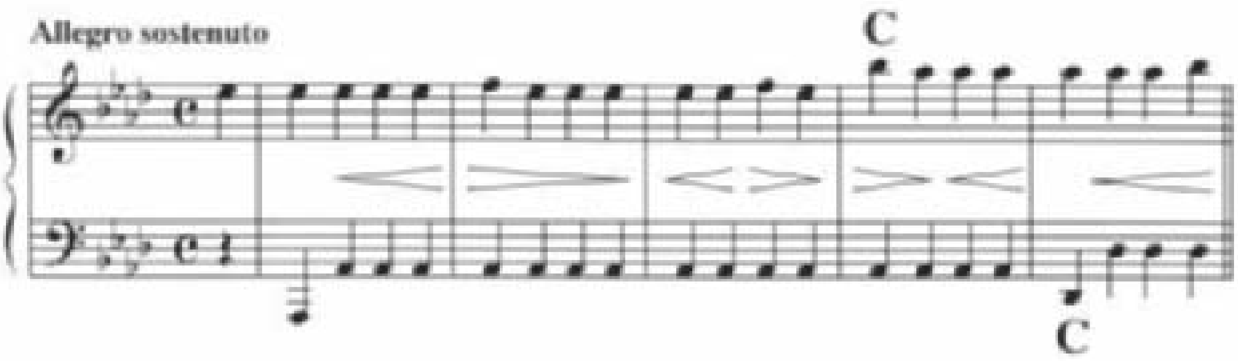

Ex. 1b. Mozart: Sonata K.304 for violin & piano, first movement.


Ex. 1c. Brahms: Sonata Op. 100 for violin & piano, first movement.


Ex. 1d. Schumann: *Träumerei*, Op. 15, No. 7.




Ex. 1e. Chopin: Etude Op. 25 No. 1 (reduction).




Ex. 1f. Bach: Prelude No. 1, Bk. 1, Well-*Tempered Clavier* (reduction).


Ex. 1: Melodic (Contour) Accents C.

Examples 1 demonstrate melodic accents with reference to extracts from the scores of some well-known piano pieces. The accents are marked C for contour; the letter M is reserved here for metrical accents (see Ex. 3).

Ex. 1a is an unaccompanied melody from a Chopin study. There are melodic accents at both peaks and valleys, but the peaks are more prominent than the valleys. The biggest accent is probably the second one marked, the note A3 at the top of the diminished-seventh leap – the largest interval between any pair of successive tones in the melody. Melodic accent also depends on the distance from the mean pitch of the local context; in this regard, the note B3 bears the greatest melodic accent. In the Mozart example (1b), the strongest melodic accent from the point of view of interval size is the one following the descending fifth interval (F# to B), whereas the strongest accent from the point of view of distance from mean pitch is the top G (G4).

In the highest voice of the Brahms example (1c), there is again an asymmetry between peaks and valleys. Melodic accents at contour peaks (B4, F#5) are more prominent than in the valley (E4) – this time, because the latter approaches the middle of the texture. The final accent (F#5) is the strongest of the three, because it is the highest pitch in the example, and is preceded by a relatively large leap (P4).

It is not immediately clear from the Schumann example (1d) whether the melodic accent belongs to the first or the second of the two highest tones (F5). At first glance, the first would appear to bear the accent, because it is preceded by a leap of a fourth, whereas the second is approached from the same pitch. If, however, one regards the first F5 as a non-essential tone that anticipates the second F5, then it is the second tone that bears the accent.

Ex. 1e is a reduction of the opening of a Chopin study; only the outer voices are shown. The strength of melodic accents in both the soprano and the bass is determined both by the sizes of preceding leaps and by the distance from the middle of the texture. Similar remarks apply to the melody and bass lines of the Bach reduction in Ex. 1f.

### 3.2 Harmonic accents

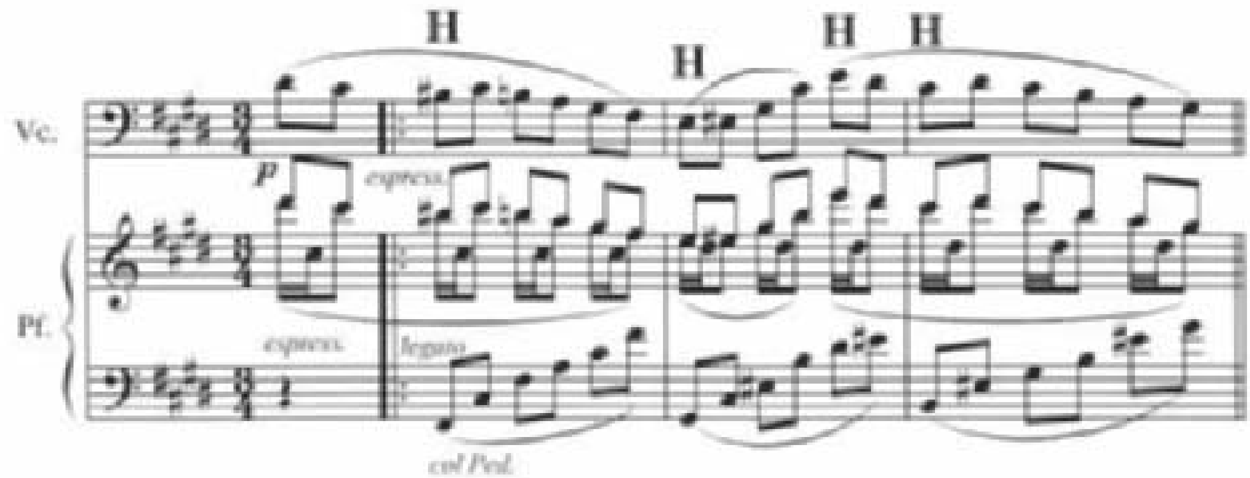
Harmonic accent corresponds closely to harmonic *tension*, as measured and modeled e.g. by Bigand, Parncutt, & Lerdahl (1996). Harmonic accents occur both at harmonic *changes* (the "horizontal" aspect, relative to conventional music notation) and at harmonic *dissonances* (the "vertical" aspect) (Smith & Cuddy, 1989). The salience of a "horizontal" harmonic change increases as the harmonic distance between a chord and its context increases (Dawe, Platt, & Racine, 1994) - an effect dubbed *harmonic charge* by Sundberg (1988). A "vertical" harmonic dissonance occurs when a single melodic tone clashes with its accompanying harmony – and effect called *melodic charge* by Sundberg. In the latter case, both roughness and harmonic distance can contribute to the overall perceived dissonance of a melodic tone (Terhardt, 1974). Harmonic dissonance depends on the spectral and temporal envelopes of a sound, which in turn depend both on the instrument playing and on the playing technique; it may therefore be either performed or immanent. By contrast, the harmonic distance between two chords is always immanent.

Ex. 2a. Mozart: *Fantasie* K. 397.

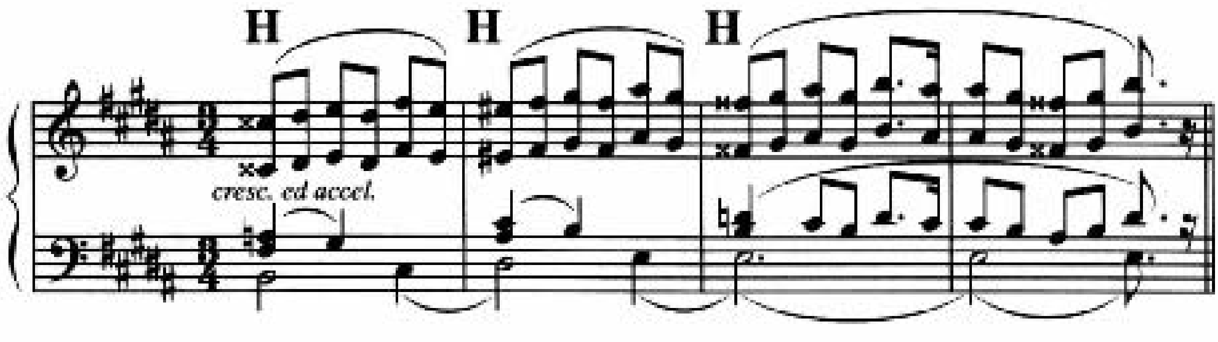

Ex. 2b. Brahms: Trio Op. 40 for horn, violin & piano, 3rd movt.


Ex. 2c. Bach: Prelude No. 12, Bk. 2, *Well-Tempered Clavier*.

Ex. 2d. Brahms: Sonata in E minor for 'cello and piano, 2nd movt.




Ex. 2e. Chopin: Etude Op. 25 No. 10.




Ex. 2f. Mozart: Concerto K. 466 for piano and orchestra, 1st movt.


Ex. 2: Harmonic Accents H.

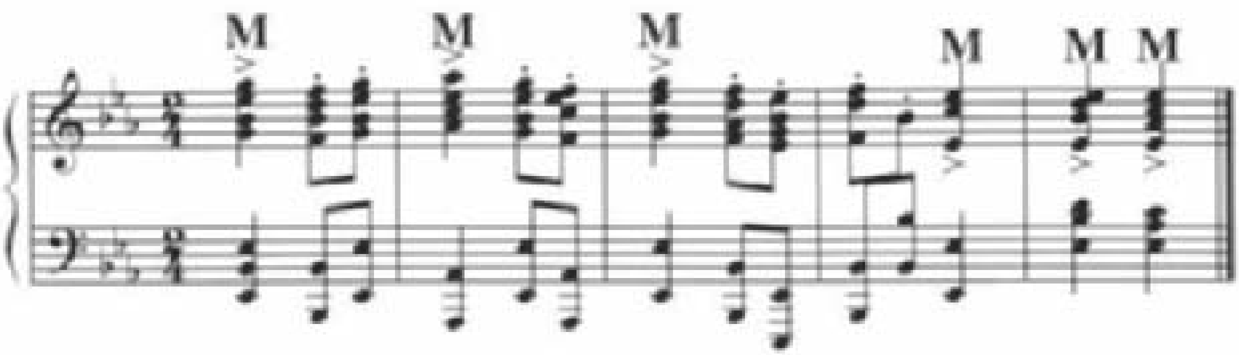
A straightforward example of accents occurring at changes of harmony is the Mozart (2a), where the emotive effect of quite simple harmonic changes is felt at the start of each measure. A more distant and dramatic change of harmony occurs in the second measure of the Brahms example (2b). The remaining excerpts in Ex. 2 focus on harmonic accents at points of relatively sharp dissonance. In the first full measure of the Bach Prelude (2c), the 6-4 chord is dissonant relative to the chords that precede and follow it. In Examples 2d (Brahms) and 2e (Chopin), dissonant chromatic tones in the melody are felt as harmonic accents. In Ex. 2f (Mozart) the dissonant tones are diatonic rather than chromatic.

### 3.3 Metrical accents

In the present definition, metrical accents are essentially confined to music; in speech, they only occur in metrically recited poetry. A possible definition is that metrical accents imply body movement or dance. If a rhythmic rendition of a poem evokes such a strong feeling of beat or pulse that listeners are inspired to dance, then the poem has, according to the present definition, become a piece of music.

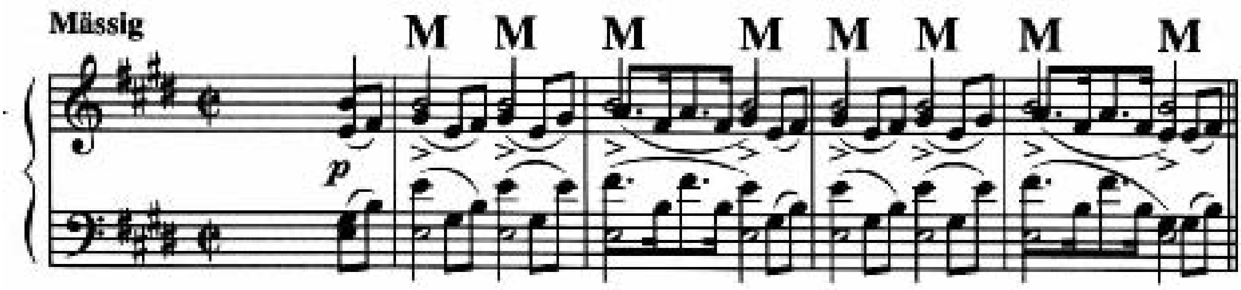
Metrical accents relate to the underlying beat, pulse or "stratum" as it is expressed at various hierarchical levels (Parncutt, 1994a; Yeston, 1976). In Western music, metrical accents are conceived of relative to standard notation as sub-beats, beats, sub-measures, measures, and hypermeasures. Listeners tend to focus on a single level of moderate tempo called the *tactus* or *referent level*, and perceive other levels relative to that level (Jones & Boltz, 1989). The tactus typically lies near 100 beats per minute (or IOI = 600 ms).

Ex. 3a. Brahms: *Rhapsodie* Op. 119 No. 4.

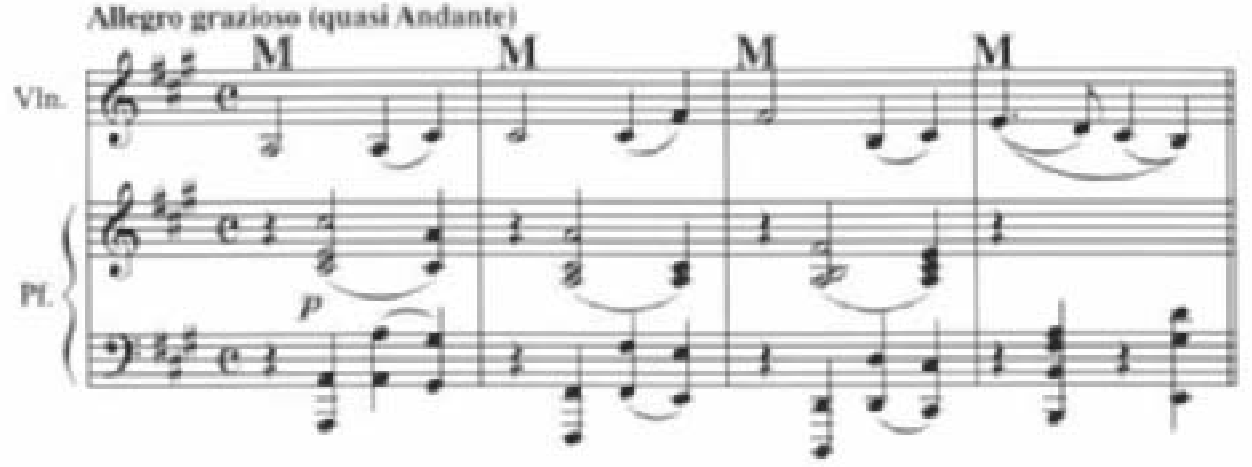



Ex. 3b. Brahms: *Ballade* Op. 118 No. 3.


Ex. 3c. Schubert: *Des Baches Wiegenlied* from *Die schöne Müllerin.*



Ex. 3d. Brahms: Sonata Op. 100 for violin & piano, last movt.




Ex. 3: Metrical Accents M.

Metrical accents are marked M in Examples 3. In the Eb Rhapsody (Ex. 3a), Brahms marks accents on the downbeats of the opening bars, evoking a march-like mood. The 5-measure phrase is divided into 2 + 3, where the final three measures are metrically ambiguous: as accented in the score, they may be divided into two halves of three beats each. Metrical ambiguity is more common in Western music than its notation would suggest (Vos, Collard, & Leeuwenberg, 1981). In ambiguous cases, listeners tend to choose one alternative soon after the piece begins, and stick to it in the face of evidence to the contrary (Lee, 1991; Longuet-Higgins & Lee, 1982).

Metrical accents with similarly powerful effect occur in the *Ballade* (Ex. 3b). In Schubert's *Des Baches Wiegenlied* (3c), the regularity of metrical accents is evocative of the rocking of a cradle. Metrical accents need not occur on every measure; hypermetrical accents beginning groups of 2 or 4 measures are commonplace in tonal music (for examples see Rothstein, 1989).

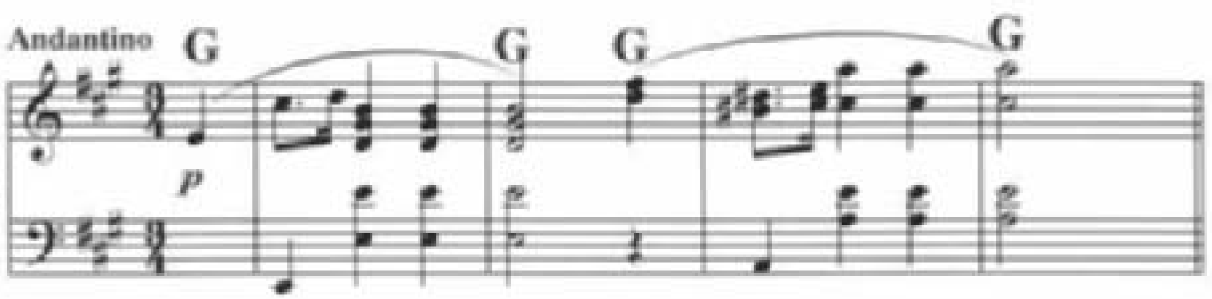
Metrical accents can be supported by changes of timbre, as Ex. 3d shows. Here, the start of each measure is characterized by the unaccompanied sound of the solo violin, contrasting with the blend of violin and piano that takes over for the remainder of the measure. Ex. 3d also serves as an illustration of the difference between notated and perceived meter (cf. Edlund, 1995); if the pianist plays too loudly, or the violinist too softly, a listener may perceive the downbeat to occur on the second 1/4-note of each measure.

### 3.4 Grouping accents

Both speech and music are characterized by hierarchical structures of phrases and sections of various sizes, where the start or end of each longer phrase or section generally coincides with the start or end of shorter phrases or sections (Povel & Essens, 1985). Group boundaries ae similar to points of stream segregation (Bregman, 1990) in that they typically occur when time and pitch intervals between successive notes are bigger than others in the immediate context (Clarke & Krumhansl, 1990; Friberg, Bresin, Frydén, & Sundberg, 1998). The strength of a grouping accent depends on the number of hierarchical levels that it demarcates (Gottschewski, 1993b; Parncutt, 1987; Rosenthal, 1992; Todd, 1985). According to this rule, the first and last grouping accents marked in both the Bach example (4a) and the Chopin example (4b) are more important than those that separate the two phrases in each case.

Ex. 4a. Bach: Chorale No. 98, *O Haupt voll Blut und Wunden.*


Ex. 4b. Chopin: Prelude Op. 28 No. 7.




Ex. 4: Grouping Accents G.

Grouping accents at the highest level are those that begin and end entire movements. At the next level down, a grouping accent may announce the beginning of a recapitulation section. A somewhat extravagant example of the latter is the C-major recapitulation of Rachmaninoff's 2nd Piano Concerto. A purposely subtle example is the first movement recapitulation of Mozart's Sonata in $B^{b}$ K. 333.

### 3.5 Reductional accents

Reductional (or *linear*) accents correspond to notes at a deeper reductional level than the musical surface. In Schenkerian analyses, reductional accents "hang together" perceptually by moving by step rather than leap (cf. Noorden, 1975). Table 2 shows some melodic reductions drawn from the above examples. The specific tones to which the reductions refer may be deduced by comparing the table with each example in turn.

| | |
|---|---|
| Ex. 1a | G# A B A G# |
| Ex.1c | C# B A A# |
| Ex.1f | E D C B |
| Ex.2a | F E D |
| Ex.2e | D# E F# G# A# B |
| Ex.3a | G Ab G F Eb |
| Ex.3c | G# A G# F# E |
| Ex.5 | C Bb Ab Bb C D |

Table 2: Reductional accents in Examples 1 to 4

Ex. 1e may be regarded as a further case of reductional accentuation. The melodic tones are separated from each other by harmonic flourishes (omitted from the reduction).

## 4. Expression in Speech and Music
### 4.1 Speech

To understand musical expression, it helps to take a careful look at expressive communication in speech. Speech may be regarded as the most important auditory stimulus for humans, and the auditory system is in many ways "tuned" to speech, to facilitate communication (e.g., Terhardt, 1998). We may therefore expect characteristic features of speech communication to affect the way other sounds are perceived - especially music, due to its expressive nature. In particular, the expression of *emotion* in speech has striking parallels with the expression of emotion in music (Gabrielsson & Juslin, 1996; Juslin, 1997; Zentner & Scherer, 1998).

Of the five varieties of *immanent* accent in music (grouping, metrical, melodic, harmonic, and reductional), only two have clear correlates in speech: grouping and reductional accents. An immanent accent in speech may be defined as an accent that is clear from written text, as read by a literate speaker of the appropriate language. *Grouping* accents correspond to the beginnings and endings of serial groups on various hierarchical levels: phrases, sentences, paragraphs (i.e., strings of sentences that concern a given topic), and even whole stories. *Reductional* accents apply to events that are important for syntactic and semantic reasons; for example, an unusual word appearing for the first time in a text will normally be semantically important. In the case of tone languages (e.g., Chinese), we may add *melodic* accents to this list; a Chinese speaker reading a Chinese text knows which syllables to articulate with which melodic shapes. In stress languages (including English and German), melodic accents are never immanent, but always performed: a text may be intoned with a range of different melodic shapes, depending on the reader's interpretation of it.

Performed accents in speech fall in much the same categories as in music. A speaker may attract attention to important syllables, words, or phrases by manipulating timing (onset time or duration), loudness, pitch, or timbre, while remaining within the expected category boundaries of these parameters. A word may simply be said more loudly (or perhaps, for special effect, more softly), higher in pitch (or perhaps lower), or with an unusual vocal timbre. In the case of timing, a speaker may delay the onset of the word, extend its physical duration, or change the duration of the following silence, either shortening it (making a smoother transition to the next word) or lengthening it (glottal stop or pause). Melodic forms (rising versus falling contours) are especially important for the communication of linguistic syntax and semantics (Dombrowski, 1995).

These changes do not normally happen independently of one another. To "raise one's voice" normally involves increasing both loudness and pitch. Reductional accents may be reinforced by increasing both pitch and loudness at important words ("Was that *Jenny*?" No, it was *James*."). In other words, in speech (as in music) performed accents are often coupled with each other, and this coupling tends to make the corresponding message easier to comprehend (Jones, 1987). Grouping accents are brought out by adjusting the length of the pause at each boundary to reflect its structural level; longer pauses are more likely to occur between paragraphs (or strings of thematically related sentences) than between sentences, and between sentences than phrases. Exceptions to these rules may generally be explained in terms of other accents or effects. For example, a speaker may run sentences or paragraphs into each other to achieve a sense of urgency, or to prevent others from interrupting. Grouping accents may also be clarified by shaping the pitch contour in characteristic ways.

The rhythm of speech is one of the main cues that allow listeners to anticipate future words (Martin, 1972). If the characteristic timing patterns of speech are changed, intelligibility can suffer. Intelligibility can also suffer if the performed accents supporting immanent accents are removed, or used in an unfamiliar way. Users of simpler text-to-sound software packages will readily testify that speech without correct or acceptible performed accentuation is difficult to follow. Only quite sophisticated rule systems can convincingly generate the pitch contour and timing profiles of natural speech; for an early example, see Carlsson and Granström (1974). These aspects of meaningful speech communication have surprisingly close analogs in musical communication. If, for example, piano music is played by converting the notated score directly to MIDI format and feeding the result directly to a MIDI keyboard, a listener will only receive a part of the message that the music might otherwise have transmitted. The performance may be made more musical by processing it within a rule system (such as Sundberg, 1988). Here, I call the part that is missing from the unprocessed performance the "performed accents".

To summarise, there is a clear relationship between performed and immanent accents in speech. Speakers do not stress syllables and words at random. Specific speech events are chosen for privileged treatment – those that carry most information, or have the most structural significance. The primary goal is to communicate linguistic structure and meaning.

The phenomenon may be encapsulated in a simple but powerful hypothesis: *performed accents reinforce immanent accents.* The main aim of the present paper is to apply this idea to the case of music performance.

## 4.2 Music

The idea that performed accents reinforce immanent accents is broadly consistent with the approach and findings of Clarke (1988). It implies that most observed and reproducible fluctuations in performance timing may be related back to the underlying immanent accents that a performer intends to communicate. Consistent with this idea, Clarke (1993) found that performances were easier to imitate when performed accents corresponded in musically typical ways to immanent accents. This may be because listeners cannot clearly cognitively separate different kinds of accents from each other (Tekman, 1997).

Drake and Palmer (1993) provided convincing experimental evidence that pianists bring out grouping, melodic, and metrical accents by systematically varying key velocity (stress), onset time (agogic accent), and articulation. The most obvious way to bring out an immanent accent in performance is by playing more loudly (stress). The idea that pianists should also take some extra time in the vicinity of important musical events (e.g., Crelle, 1823) is not quite to obvious. A possible explanation is simply that temporarily slowing the musical pulse, to allow more time around an event than listeners expect, will attract their attention to that event. Kurkela (1995) has studied five performances by famous pianists of a fragment of Chopin's Nocturne in E*b* Op. 9 No. 2, and in the light of the data developed a set of performance rules; each rule adds a small increment to the duration of a musically significant note that usually corresponds, in the present terminology, to a metrical or grouping accent. At the level of subdivisions of the beat, the observation that notated 2:1 ratios (as triple subdivisions of a beat or measure) are consistently performed in a ratio of less than 2:1 (Gabrielsson, Bengtsson, & Gabrielsson, 1983) has been accounted for by assuming that local tempo slows in the vicinity of metrical accents (Parncutt, 1994b).

The hypothesis that performed accents reinforce immanent accents allows for a wide variety of interpretations of a given piece of piano music. Musical scores typically include many different kinds of immanent accent, of various strengths. The performer is thus presented with a wide variety of accents, occurring both individually and in various combinations. It would appear to be a matter of artistry to decide which of these accents should be emphasized in performance, and which of the various possible means of expression should be used for this purpose. Stresses may be large or small, sudden or gradual – in the latter case, involving a gradual increase in loudness for notes leading up to the accent, and a gradual decrease afterwards. Agogic accent may involve a slowing of local or instantaneous tempo before the event and a speeding up after it. Articulatory accent may involve a temporary increase in *legato* for tones near a given event.

An immediate problem with the hypothesis that performed accents reinforce immanent accents is that it is difficult to falsify. First, different kinds of immanent accent tend to coincide. This makes it easier to understand, remember and reproduce the score (Jones, 1987); however, it also makes it difficult or impossible to determine which accents are in effect in each case. Second, just about any performance might be related to a pattern of accents that have been selected *post hoc* to account for the performance's expressive profile. The present approach to musical timing may therefore strike the reader as distinctly less systematic than that of many contemporary researchers in the area of music performance. I justify my "semi-scientific" approach as follows. Musical expression is a highly complex phenomenon that researchers in systematic musicology are only beginning to come to grips with. In spite of converging evidence from various research paradigms, including both direct measurement of individual performances and analysis-by-synthesis, it remains unclear, for example, exactly what kind of computer model might best produce musically satisfying expressive profiles for a reasonable range even of relatively simple piano styles. Perhaps there is no alternative but to accept an inevitable subjective element in the generation of musically satisfying computer-generated performances. And only when such performances can be convincingly generated can one be relatively confident of the validity of the underlying theory.

In spite of these caveats, it is possible to work productively with the above hypothesis. First, the hypothesis allows for musically "bad" performances to be generated in a consistent manner. It may be tested by comparing listeners' preferences for "good" and "bad" renditions of passages of music, as generated by a mathematical formulation of the hypothesis. Thompson, Sundberg, Friberg, and Frydén (1989) tested the rule-based system of Sundberg (1988) in this way. Second, the diversity of outcomes that can be generated by the hypothesis seems realistic in the light of the diversity of performances that may be observed for a single piece (see e.g. Repp, 1992).

### 4.3 Music examples

We return now to Examples 1 to 4, and consider how the timing and velocity of key-strokes in performance may be used to bring out the marked accents. In Ex. 1a, Chopin's dynamic markings indicate that he prefers an interpretation in which the first and third of the melodic accents (marked C) are stressed. In addition, the notes bearing the accents may be delayed and lengthened. In the Mozart example (1b), timing and loudness variations in the vicinity of the marked accents must, for stylistic reasons, be very subtle (they nevertheless tend in the same direction as expressive variations appropriate for Chopin). In Ex. 1c, Brahms indicates by his dynamic marks that the final contour accent should be brought out quite strongly. Performers of Schumann's *Träumerei* (1d) traditionally slow down considerably in the vicinity of the melodic accent at the top F; for experimental data see Repp (1992). Ex. 1e (Chopin) is a case in which appropriate timing results more or less automatically from the technical difficulty of the music: the leaps to the marked tones can be performed most reliably if the tones are slightly delayed relative to the underlying pulse. In the Bach Prelude (1f), performers may appropriately take a little extra time (or, occasionally, less time) and increase (or, sometimes, decrease) the overall loudness at the measures marked by melodic accents (measures 5, 7, 10); for experimental data see Cook (1987) and Shaffer (1995). These expressive variations may also be linked to timbre: expansion of the texture toward the soprano or toward the bass is inevitably accompanied by a change in overall sonority.

As a rule, all these examples include important accents that, for reasons of clarity, have not been marked, but which could be brought out in performance. Ex. 1a includes several harmonic accents, three of which coincide with the last three marked melodic accents. The first and third full measures of Ex. 1b begin with hypermetrical accents, while the beginning of the anacrusis to the third full measure may be regarded as a subsidiary grouping accent. Ex. 4b includes harmonic accents (dissonances) on the downbeats of full measures 1 and 3. And so on. Also, I have generally marked only those accents that seem to be strong enough to warrant mentioning. Thus, the melodic accents in the valleys of the melody in Ex. 1a (B#2, Fx3, E3, etc.) are not marked.

Different techniques are used by performers to bring out different kinds of immanent accent. These differences have so far been explored systematically for some, but not all, types of immanent accent (e.g., Sundberg, 1988). In most cases, agogic and dynamic accents are applied by slowing local tempo and increasing loudness in the vicinity of an immanent accent.

Regarding *metrical* accents, it may generally be assumed that a performer is striving to communicate the notated meter to the audience, although performers sometimes create pulses that contradict the notated meter (Gottschewski, 1993a). Performers can to some extent determine which pulse or metre will be perceived in a given pattern, and when perception will switch from one metrical interpretation to another, by manipulating of timing and dynamics (Sloboda, 1983).

Timing patterns near *grouping* accents tend to be more consistent and robust than timing patterns near other kinds of accent (Drake & Palmer, 1993). Riemann (1884) observed that a phrase often begins with a simultaneous *crescendo* and *accelerando*; this lasts until the melodic climax and is followed by a simultaneous *decrescendo* and *decelerando*. This combination of "tempo curve" and "dynamic curve" may be regarded as a kind of performance accent that is extended in time over several tones. Its exact shape has been explored by Repp (1992) and Todd (1995). The specific case of the final retard (the end of a long group, leading to a highly salient grouping accent – the end of a section or an entire piece) was addressed by Friberg and Sundberg (1999). Grouping accents are not necessarily entirely immanent; performers may manipulate timing and dynamics to imply a grouping structure different from that notated or implied in the score (Lester, 1995). This kind of interpretational freedom is characteristic of the Western romantic performance tradition.

## 5. Applications

### 5.1 Computer-Assisted Performance

Learning to play an instrument invariably involves thousands of hours of practice (Sloboda & Howe, 1991; Ericsson, Krampe, & Tesch-Römer, 1993). Much of this time is devoted to the development of instrumental technique by seemingly endless repetition of scales, arpeggios, exercises, and studies. Pianists are motivated to devote considerable effort to technical practice in the expectation that they will be rewarded for their dedication by acquiring the ability to perform great works from the piano repertoire.

In our time, there is no compelling reason why performers should not take advantage of computer technology to reduce the amount of time they spend on technical work, thereby allowing more time for matters of interpretation. This prospect appears particularly attractive for musicians with physical disabilities that prevent them from reaching high levels of technical expertise.

It is not immediately clear how such a computer system may be designed. It would clearly be impractical for users to adjust the timing and intensity of each individual note in a score; that would amount to writing the MIDI file directly. Instead, the system would need to allow adjustment of the timing and dynamics of several notes at once - even of an extended passage. And it would have to do this in characteristic ways that mimic music performance practice. Thus, *some* kind of underlying theory is indispensible.

Just as indispensible would be active, creative, artistically informed input from the user. After all, as mentioned earlier, no currently available computer performance software, regardless of its sophistication, can (yet) consistently produce satisfying renditions of musical works in a reasonable range of styles, using only the scores as input. For the moment, there seems little chance that computers will replace human musical interpreters.

A relatively theory-free music interpretation system has recently been developed by Dalgarno (1997), with encouraging results. How would such a system operate if the present theory of immanent and performed accents were adopted as a theoretical basis? The user might first be presented with a structural analysis of the piece in question, featuring the various types of immanent accent along with preliminary estimates of their relative strengths or saliences. Since such an analysis is quite subjective, the user could be asked to adjust this analysis to correspond to his or her concept of the piece. The user would then be invited to choose which accents to emphasize, and by how much.

The system would allow modulations of timing and dynamics associated with chosen immanent accents to be freely manipulated. In the case of dynamics, stresses may be large or small, sudden or gradual. A "gradual stress" would involve a gradual increase in loudness for notes leading to the accent, and a gradual decrease afterwards. In the case of timing, agogic accents may involve a slowing of local tempo before the event and a speeding up after it; and articulatory accents may involve a temporary increase in *legato* for tones near a given event.

The creation of user-friendly interfaces for such adjustments, including local and global curves for modulation of timing and dynamics in the vicinity of specific accents, would be a relatively straightforward programming exercise. After each change to the interpretation, the result would be synthesized, listened to, and appraised by the user, who would then make further changes.

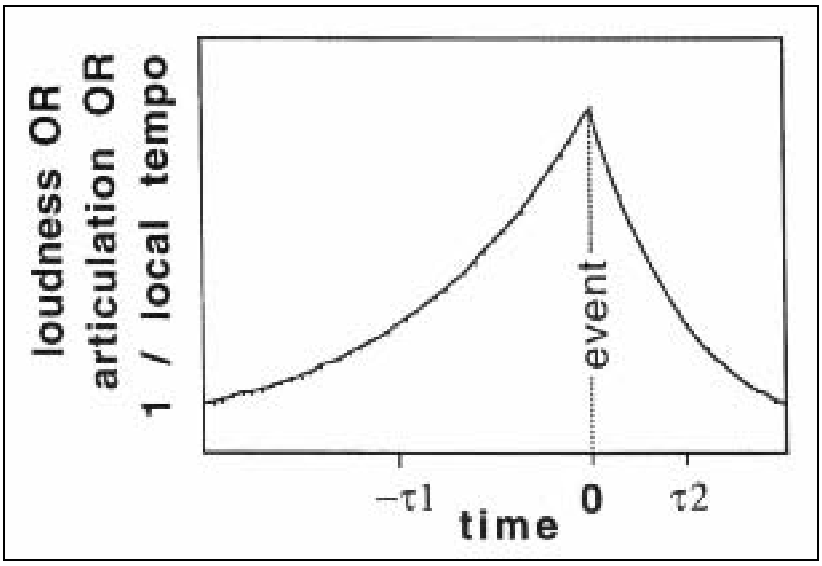


Figure 1. A Schematic Timing Curve For Individual Accents.

Ex. 5. Chopin: Posthumous Study No. 1.

The following stepwise-iterative procedure could enable the creation of convincing expressive interpretations:

- The musical score is displayed on the computer screen. Various kinds of immanent accent are notated automatically on the score, in a similar way that accents were marked in Ex. 5. Note that more than one immanent accent may occur at one time. For example, the first note of Ex. 5 bears both metrical accent (as the start of both 2- and 4-measure hypermeters) and grouping accent (as the start of a large section). The second melodic tone in the second measure marks a local contour peak and receives melodic accent (C); it is also dissonant relative to the underlying harmony, and bears harmonic accent.
- The user sets parameters for each performed accent. The parameters will depend on the model of expression that is being used (which in turn depend on musical style). A simple model of timing is shown in Figure 1. The strength of the accent corresponds to the height of the peak. The width of the curve $(\hat{o}^1 + \hat{o}^2)$ corresponds to the duration of delaying and lengthening. A parabolic or exponential shape is consistent with recent experimental data and models on timing

(Repp, 1992; Friberg and Sundberg, 1999; Todd, 1995). A differently shaped curve would be required for expressive variation of loudness (key velocity) and articulation in the vicinity of an immanent accent; moreover, the shape of the curve may depend on the kind of emotion being expressed (Clynes, 1977). Note that more than one note may be included under a single curve; in the case of a final ritard, for example, the left portion of the curve may be elongated over several measures, and the peak corresponds to the end of the piece.

- In a user-friendly program, the timing curve associated with each accent is seen in "pop-up" form simply by clicking on the note or chord in question. The size and shape of the curve is altered by "nudging" the curve at its "handles" using the mouse; no numerical values need be entered.
- Where similar timing curves are applied to a number of accents (e.g., the downbeat of every bar), a group of notes can be selected, and their average timing curve adjusted in a single operation.
- The tempo curves associated with all the accents marked in the score are added to obtain a global timing curve for the piece. The duration of the timing curve is normalized to maintain a suitable total duration, or average tempo.
- The result is synthesized, listened to, and appraised by the user, who then makes changes to the original parameters, and returns to the start of the cycle. After many such iterations, the user aims to arrive at an interpretation of the piece that is not only musically acceptable (assuming that the user has sufficient musical skills, talent, and patience) but also unique and personal.

### 5.2 Piano teaching

Musical interpretation is traditionally taught by a combination of imitation, intuition, and trial and error. Teachers often describe the details of interpretation to their students using high-level imagery ("angry", "flowing") that give little direct, detailed information about expression to their students.

The present taxonomy of accents, coupled with the principle that performed accents reinforce immanent accents, provides music teachers with a technical vocabulary for the teaching of musical interpretation, just as the theory of sentics (Clynes, 1977) made explicit the relationship between emotional qualities and corresponding musical gestures. Mawer (1999) had string students perform voice-leading analyses of their scores as a basis for exploring both expressive/interpretive and technical possibilities in performance. My approach here is similar, except that I try to cover a broader range of structural features: the various immanent accent types addressed above.

Suppose that a teacher asks a student to make more of a given immanent accent. A student conversant with the approach I have outlined could proceed to identify other similar immanent accents in the score, and then decide by what appropriate kind of expressive shaping they may be brought out in performance. The student might even mark an entire score with immanent accents of various kinds, as shown in Ex. 5, and then discuss with the teacher what kind of expressive accentuation might be suitable for each.

Too many accents are marked in Ex. 5 for all of them to be brought out in a single performance. A pianist might select a particular subset of the accents to bring out, and deliberately skim over the rest. In one performance, for example, the focus may be on the contour accents; in another, on harmonic accents. In this way, a pianist may systematically explore interpretations of a given passage that are clearly distinguishable, but nonetheless internally coherent.

### 5.3 Performance as analysis

Recent research in expressive music performance has repeatedly shown that expressive musical performances communicate information about musical structure to audiences. For example, Behne and Wetekam (1993) showed pianists can communicate intended rhythmic aspects of their performances. In this sense, performers may be regarded as "ecologically sound" music analysts.

Clarke (1995) demonstrated how analysis of MIDI data can enable interesting conclusions to be drawn about the structure of a piece of music, as conceived by a performer. Here, this may be done by applying the hypothesis that performed accents reinforce immanent accents, but in reverse. In a first stage, fluctuations in timing and loudness typical of those occurring in the vicinity of accents may be recognized. These may be used to map out the various accents that a pianist has brought out in a performance, and to estimate the relative importance of those accents in the performer's conception of the piece. In the process, it may be necessary to identify the specific strategies that a given performer uses to communicate structure – what Shaffer (1995) has termed the *character* of the performance. In a second stage, a cognitive representation of the music may be sought that is consistent with that accent structure. Such a structure may best be found by top-down matching, mimicking cognitive processes occurring during music listening. The end result would be a performance-based analysis.

An "expressive deconvolution" of this kind should be possible. It simply mimics what human listeners do when they understand a performer's structural interpretation. But the above procedure is impractical due to the multiplicity of different interpretations at each stage. A more realistic alternative might be to have expert listeners analyse the sound of the music directly. Musicians would listen to recorded performances by themselves and others, and annotate immanent and performed accents on the score, using a coding system. They might be asked to listen to several of performances of the same piece (in different random orders). For piano music, annotations might focus on expressive changes in timing (agogic accents: delay and anticipation of note onsets relative to when listeners expected them; noticeable accelerations and decelerations; shaping of phrases) and dynamics (notes played noticeably louder or softer than other notes in their context; crescendo, diminuendo, dynamic shaping of phrases). The perceptual salience of each performed accent could also be estimated. Later, in collaboration with music analysts, salient events would be categorised as specific immanent or performed accents. The performer's structural concept of the work would emerge more or less clearly from the results.

"Subjective" structured listening would have the following three advantages over "objective" analysis of MIDI data:

1. Physical regularity is not the same as perceptual regularity. A music performance that sounds perfectly regular, mechanical, or metronomic generally deviates systematically from physical regularity (Drake, 1993; Gérard, Drake, & Botte, 1993; Penel & Drake, 1998; Seashore, 1938) in the direction of a typical expressive performance (Repp, 1998). A direct analysis of MIDI data is insensitive to this effect. A possible solution would be to adjust MIDI timing and velocity data by subtracting a mean or deadpan performance (Repp, 1997); but because tempo fluctuates on different hierarchical levels simultaneously, the tempo of the baseline would need to be adjusted quasi-continuously to that of the expressive performance, using arbitrarily determined time constants. Such a complex procedure is unlikely to be reliable. A structured listening experiment would automatically take the effect into account.
2. Inaudible variations of SPL and timing cannot affect data obtained from a structural listening experiment. But to remove inaudible effects from MIDI data would require the application of sophisticated procedures, whose effects on the data would be uncertain. Further perceptual parameters that could be taken into account during such an adjustment are the perceptual difference limens for SPL and timing, which – to make matters even more complicated – are context dependent; and masking, which would render some notes (especially inner parts) inaudible.
3. Structured listening enables the relative perceptual salience of significant events in a piece, such as specific expressive gestures, to be estimated directly. To estimate perceptual salience of events from MIDI data in a systematic fashion would require the development of a perceptual theory for this purpose, and the results would be more abritrary.

## 6. Conclusion

I have presented a theory of musical expression in which fluctuations in agogics and dynamics are related primarily to *local* (rather than global) features in the music. These local features are identified as varieties of *accent*. This use of the word accent is largely a matter of convenience, and is supported by similar usages in other areas of music theory; those who prefer to restrict the use of the word accent to *stress* or *dynamic accent* might instead, for the purpose of this theory, refer to "relatively salient events" or even "musical attention-getters". The emphasis on local features makes the theory similar to that of Sundberg (1988); the present theory may be regarded as an attempt to place that theory on a new, unified theoretical basis by relating most of its detail to a general music-theoretic concept of accent.

Of course the relationship between expressive performance and accent is only part of the story. Several aspects of expressive performance are difficult to fit into the "accent" mould. For example, a sustained crescendo may lead to a grouping accent on a relatively high structural level (the start of a new section). But the acceleration that often accompanies such a crescendo (e.g., Wetekam, 1995) cannot easily be accounted for by a theory of accent, since – as demonstrated in several of the above examples - tempo typically slows rather than accelerates before important events.

I have formalized the relationship between timing/dynamics and musical structure by breaking it down into two conceptual stages. In the first stage, the musical structure determines the positions, types, and relative strengths of immanent accents. Second, the positions, types, and relative strengths of immanent accents influence fluctuations in timing and dynamics in performance. Expressive fluctuations are assumed to be centered on a given point in time (the corresponding performed accent) but to span a time period which may include several events at the musical surface. Moreover, several nearby immanent accents may contribute to a shift in local tempo or dynamic at a particular point in time, by addition (or, possibly, non-linear combination) of the timing/dynamic curves associated with each individual accent. In this way, the model may account for expressive effects that occur on both local and global levels.

Performers need not, of course, be aware of specific relationships between immanent and performed accents in order to produce convincing musical interpretations. It may be that performers instead learn to link specific timing and dynamic profiles to particular kinds of musical passage by a rather global and intuitive kind of pattern matching. If that is the case, the question arises as to the *origin* of the link between immanent and performed accents. Perhaps the structural-expressive relationships posited in this paper can only develop in a gradual way, as generations of musicians interact to build up a performance tradition. If that is the case, then the present model can only account *indirectly* for aspects of musical expression. Performance traditions may also include relatively arbitrary timing/dynamic gestures or clichés that cannot be embodied in any kind of reductionist model. This observation, coupled with the idea that performers bring aspects of their own extra-musical life experience to their performances, ensures that the present model can only account for a part of the complex web of relationships that link musical structure to musical expression.

**Acknowledgments.** I am grateful to John Sloboda and Eric Clarke for helpful comments on an earlier draft.

## References

BACH, C. P. E. (1753). *Versuch über die wahre Art, das Clavier zu spielen*. Berlin. Reprint (1957) Leipzig: C.F. Kahnt.

BEHNE, K.-E., & WETEKAM, B. (1993). Musikpsychologische Interpretationsforschung: Individualität und Intention. *Musikpsychologie, 10,* 24-37.

BIGAND, E., PARNCUTT, R., & LERDAHL, F. (1996). Perception of musical tension in short chord sequences: Influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics, 58*, 125-141.

BREGMAN, A.S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

BRESIN, R. & FRIBERG, A. (1998). Emotional expression in music performance: synthesis and decoding. *TMH-QPSR* (Speech Music and Hearing Quarterly Progress and Status Report, Stockholm), *3-4*, 85-94.

BURNS, E.M., & WARD, W.D. (1978). Categorical perception – Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63,* 456–468.

CARLSSON, R., & GRANSTRÖM, B. (1974). A phonetically oriented programming language for rule description of speech. In G. Fant (Ed.), *Proceedings of the speech communication seminar*, Vol. 2: *Speech production and synthesis by rules* (pp. 245–253). Stockholm: Almqvist and Wicksell.

CLARKE, E.F. (1988). Generative principles in music performance. In Sloboda, J. (Ed.): *Generative processes in music. The psychology of performance, improvisation and composition.* Oxford: Science Publications.

CLARKE, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception, 10*, 317-341.

CLARKE, E.F. (1995). Expression in performance: Generativity, perception, and semiosis. In J. Rink, (Ed.), *The Practice of Performance. Studies in Musical Interpretation* (pp. 21-54). Cambridge: Cambridge University Press.

CLARKE, E. F., & KRUMHANSL, C. L. (1990). Perceiving musical time. *Music Perception, 7*, 213-252.

CLYNES, M. (1977). *Sentics, the touch of emotion.* New York: Doubleday Anchor.

CLYNES, M. (1985). Music beyond the score. *Communication & Cognition, 19,* 169–194.

COOK, N. (1987). Structure and performance timing in Bach's C Major Prelude (WTC I): An empirical study. *Music Analysis*, *6,* 257–272.

COOPER, G., & MEYER, L. B. (1960). *The rhythmic structure of music.* Chicago: University of Chicago Press.

CRELLE, A.L. (1823). *Ausdruck und Vortrag für Fortepiano-Spieler*. Berlin: Maurer'sche Buchhandlung.

DALGARNO, G. (1997). Creating an expressive performance without being able to play a musical instrument. *British Journal of Music Education, 14* (2), 116-171.

DAWE, L. A., PLATT, J. R., & RACINE, R. J. (1994). Inference of metrical structure from perception of iterative pulses within time spans defined by chord changes. *Music Perception, 12*, 57-76.

DEUTSCH, D. (1982). Grouping mechanisms in music. In D. Deutsch (Ed.), *The Psychology of Music.* New York: Academic.

DOMBROWSKI (1995). *Musikpsychologie*.

DRAKE, C. (1998). Psychological processes involved in the temporal organization of complex auditory sequences: universal and acquired processes. *Music Perception, 16,* 11-26.

DRAKE, C., & PALMER, C. (1993). Accent structures in music performance. *Music Perception*, *10,* 343–378.

EDLUND, B. (1995). Making meter evident – On the playing of an ambiguous Bach melody. Musikpsychologie, 12, 28-41.

ERICSSON, K. A., KRAMPE, R. T. & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100,* 363-406.

FORTE, A., & GILBERT, S.E. (1982). *An introduction to Schenkerian analysis.* New York: Norton.

FRIBERG, A. (1991). Generative rules for music performance. *Computer Music Journal*, *15* (2), 56–71.

FRIBERG, A., BRESIN, R., FRYDÉN, L., & SUNDBERG, J. (1998). Musical punctuation on the microlevel: Automatic identification and performance of small melodic units. *Journal of New Music Research, 27*, 271-292.

FRIBERG, A. & SUNDBERG, J. (1999). Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners. *Journal of the Acoustical Society of America, 105*, 1469-1484.

FRICKE, J.P. (1968). *Intonation und musikalisches Hören.* Habilitationsschrift Köln.

GABRIELSSON, A. (1974). Performance of rhythm patterns. *Scandinavian* Journal of Psychology, 15, 63-72.

GABRIELSSON, A., BENGTSSON, I., & GABRIELSSON, B. (1983). Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology, 24,* 193-213.

GABRIELSSON, A., & JUSLIN, P. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music, 24*, 68-91.
GÉRARD, C., DRAKE, C., & BOTTE, M.-C. (1993). Rhythm perception: Interactions between time and intensity. *Contemporary Music Review*, *9*, 179-189.
GOTTSCHEWSKI, H. (1993a). *Theorie und Analyse der musikalischen Zeitgestaltung.* Dissertation, Universität Freiburg.
GOTTSCHEWSKI, H. (1993b). Tempoarchitektur – Ansätze zu einer speziellen Tempotheorie. *Musiktheorie, 8* (2), 99–118.
GREY, J.M. (1977). Multidimensional scaling of musical timbres. *Journal of the Acoustical Society of America*, *61,* 1270.
HANDEL, S. (1998). The interplay between metric and figural rhythmic organization. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1546-1561.
HONING, H. (1990). POCO: An environment for analysing, modifying, and generating expression in music. In *Proceedings of the 1990 International Computer Music Conference*. San Francisco: Computer Music Association.
HURON, D., & ROYAL, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Perception, 13,* 489-516.
JONES, M. R. (1987). Dynamic pattern structure in music: Recent theory and research. *Perception and Psychophysics, 41*, 621-634.
JONES, M. R., & BOLTZ, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*, 459-491.
JUSLIN, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception, 14*, 383-418.
KOPIEZ, R. (1993). Interpretationsforschung mit Hilfe des Computerflügels. Eine Studie zur Wahrnehmung von Interpretationsmerkmalen. *Musikpsychologie, 10*, 7-23.
KURKELA, K. (1995). Tempo deviations and musical signification: A case study. In E. Tarasti (Ed.), *Musical signification. Essays in the semiotic theory and analysis of music* (pp. 301-317). New York: Mouton de Gruyer.
LANGNER, J. & KOPIEZ, R. (1995). Entwurf einer neuen Methode der Performanceanalyse. *Musikpsychologie*, 12, 9–27.
LARGE, E. W. & KOLEN, J. F. (1994). Resonance and perception of musical meter. *Connectionist Science, 6*, 177–208.
LARGE, E. W., PALMER, C., & POLLACK, J. B. (1995). Reduced memory representations for music. *Cognitive Science, 19*, 53-96.
LEE, C. S. (1991). The perception of metrical structure: Experimental evidence and a model. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure* (pp. 59-127). London: Academic.
LERDAHL, F., & JACKENDOFF, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.
LESTER, J. (1995). Performance and analysis: Interaction and interpretation. In J. Rink (Ed.), *The practice of performance*. Cambridge University Press.
LIBERMAN, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.S. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358–368.
LONGUET-HIGGINS, H. C., & LEE, C. S. (1982). The perception of musical rhythms. *Perception, 11*, 115-128.
MARTIN, J.G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, *79,* 487–509.
MAWER, D. (1999). Bridging the divide: embedding voice-leading analysis in string pedagogy and performance. *British Journal of Music Education, 16*, 179-195.
MAZZOLA, G., & ZAHORKA, O. (1994). Tempo curves revisited. *Computer Music Journal*, *18* (1), 40–52.
McADAMS, S, & SAARIAHO, K. (1991). Qualités et fonctions du timbre musical. In J. B. Barrière (Ed.), *Le timbre, métaphore pour la composition*, pp. 164–181. Paris: IRCAM & Christian Bourgois.

NOORDEN, L. van (1975). *Temporal coherence in the perception of tone sequences.* Doctoral dissertation, Institute for Perception Research, Eindhoven, The Netherlands.

PALMER, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331-346.

PARNCUTT, R. (1987). The perception of pulse in musical rhythm. In A. Gabrielsson (Ed.), *Action and Perception in Rhythm and Music* (pp. 127-138). Stockholm: Royal Swedish Academy of Music.

PARNCUTT, R. (1994a). Pulse salience and metrical accent in musical rhythms. *Music Perception*, *11,* 409–464.

PARNCUTT, R. (1994b). Categorical perception of short rhythmic events. In A. Friberg et al. (Eds.), *Proceedings of the 1993 Stockholm Music Acoustics Conference* (pp. 47–52). Stockholm: Royal Swedish Academy of Music.

PARNCUTT, R. (1994c). Template-matching models of musical pitch and rhythm perception. *Journal of New Music Research*, *23*, 145-168.

PARNCUTT, R. (1997). Modeling piano performance: Physics and cognition of a virtual pianist. T. Rikakis (Ed.), *Proceedings of theInternational Computer Music Conference*, Thessaloniki, Greece (pp. 15-18). San Francisco: International Computer Music Association.

PARNCUTT, R. (1998). Listening to music in the real world? A critical discussion of Marc Leman's (1995) *Music and Schema Theory: Cognitive Foundations of Systematic Musicology. Journal of New Music Research, 27,* 380-408.

PENEL, A., & DRAKE, C. (1998). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research, 61,* 12-32.

POVEL, D.J. and Essens, P. (1985). Perception of temporal patterns*. Music Perception, 2*, 411-440.

REPP, B.H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America, 88*, 622-641.

REPP, B.H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". *Journal of the Acoustical Society of America*, *92*, 2546–2568.

REPP, B.H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik*. *Psychology of Music*, *21*, 48–73.

REPP, B. H. (1997). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception, 14,* 419-444.

REPP. B. H. (1998). The detectability of local deviations from a typical expressive timing pattern. *Music Perception, 15,* 265-289.

RIEMANN, H. (1884). *Musikalische Dynamik und Agogik.* Hamburg: D. Rahter.

ROSENTHAL, D. (1992). Emulation of human rhythm perception. *Computer Music Journal, 16* (1), 64-76.

ROTHSTEIN, W. (1989). *Phrase rhythm in tonal music.* New York: Schirmer.

SEASHORE, C.E. (1938). *Psychology of music.* New York: McGraw-Hill. (Dover, 1967).

SHAFFER, L. H. (1992). How to interpret music. In M.R. Jones & S. Holleran, *Cognitive bases of musical communication* (pp. 263-278). Washington, DC: American Psychological Association.

SHAFFER, L. H. (1995). Musical performance as interpretation. *Psychology of Music*, *23*, 17–38.

SHOVE, P., & REPP, B. H. (1995). Musical motion and performance. In J. Rink (Ed.), *The Practice of Performance. Studies in Musical Interpretation*. Cambridge: Cambridge University Press.

SLOBODA, J. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35A,* 377-396.

SLOBODA, J.A., & Howe, M.J.A. (1991). Biographical precursors of musical excellence: An interview study. *Psychology of Music*, *19,* 3–21.

SMITH, J. O. III (1996). Physical modeling synthesis update. Computer Music Journal, 20 (2), 44–56.

SMITH, K. C., & CUDDY, L. L. (1989). Effects of metric and harmonic rhythm on the detection of pitch alternations in melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 457-471.

SUNDBERG, J. (1988). Computer synthesis of music performance. In J. A. Sloboda (Ed.), *Generative processes in music.* Oxford: Clarendon.

TEKMAN, H. G. (1997). Interactions of perceived intensity, duration, and pitch in pure tone sequences. *Music Perception, 14*, 281-294.

TERHARDT, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, *55*, 1061–1069.

TERHARDT, E. (1998). *Akustische Kommunikation: Grundlagen mit Hörbeispielen*. Berlin: Springer.

THOMASSEN, M. T. (1982). Melodic accent: experiments and a tentative model. *Journal of the Acoustical Society of America*, *71,* 1596-1605.

THOMPSON, W.F., SUNDBERG, J., FRIBERG, A., & FRYDÉN, L. (1989). The use of rules for expression in the performance of melodies. *Psychology of Music, 17,* 63-82.

TODD, N. P. McA. (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33-58.

TODD, N. P. McA. (1989). A computational model of rubato. *Contemporary Music Review*, *3*, 69-88.

TODD, N. (1994). The auditory 'primal sketch' – a multiscale model of rhythmic grouping. *Journal of New Music Research, 23,* 25–70.

TODD, N. P. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America*, *97*, 1940–1949,

VOS, P. G., COLLARD, R. F., & LEEUWENBERG, E. L. (1981). What melody tells about meter in music. *Zeitschrift für Psychologie, 189*, 25-33.

WATT, R. J., & ASH, R. L. (1998). A psychological investigation of meaning in music. *Musicae Scientiae,* 1, 33-53.

WEGEL, R. L., & LANE, C. E. (1924). The auditory masking of one sound by another and its probable relation to the dynamics of the inner ear. *Physical Review*, 23, 266–285.

WETEKAM, B. (1995). Was passiert beim Crescendo? Ein Beitrag zur musikpsychologischen Interpretationsforschung. *Musikpsychologie, 12,* 42-55.

YESTON, M. (1976). *The stratification of musical rhythm.* New Haven, CT: Yale University Press.

ZENTNER, M., & SCHERER, K. R. (1998). Emotionaler Ausdruck in Musik und Sprache. *Musikpsychologie*, *13,* 8-25.