In W. Auhagen, W. Ruf, U. Smilansky, & H. Weidenmüller (Eds.) (2005), *Music and science - The impact of music* (pp. 33-47). Halle, Germany: Deutsche Akademie der Naturforscher Leopoldina (*Nova Acta Leopoldina*, Bd. 92, Nr. 341).

# Perception of musical patterns: Ambiguity, emotion, culture Richard Parncutt

## DRAFT SUBMITTED 3.9.04.

Please do not cite since errors may have been corrected before publication.

#### Abstract

All music contains pitch-time patterns. Gestalt principles, which are learned from perceptual interaction with the environment, explain how the elements of auditory patterns are perceptually grouped and how auditory scenes are analyzed. This enables a perceptually grounded understanding of the rules of harmony and counterpoint. Patterns such as rhythms, chords, melodies, and entire tonal passages tend to be perceived relative to reference points in pitch-time space (downbeat, root, tonic tone or chord) and to have various identities and associations. All these references are, or can be, ambiguous or multiple. Musical emotions are generated either via (ambiguous or multiple) associations from outside the music or (ambiguous or multiple) structures and associated expectations within the music. Ambiguity does not itself generate emotion; instead, both ambiguity and emotion result from pattern recognition processes. Thus, the relationship between ambiguity and emotion is not causal. The art of musical composition and performance involves the manipulation of ambiguous expectations and emotions. Learning, context, ambiguity and multiplicity are central concepts in both psychoacoustics and cultural musicology, suggesting a considerable potential for interaction between musical sciences and humanities.

## 1. Introduction

A *pattern* is a spatial arrangement of elements that belong together, or a display of the essential elements of a more complex scene. In modern auditory psychology, patterns of sound are analyzed within the context of an *auditory scene* (BREGMAN, 1990). Visualized as a graph of frequency against time, the auditory scene is a representation of the contents of short-term (or working) auditory memory, i.e. sounds in memory that are available for simultaneous processing. The research field of *auditory scene analysis* is based on, and extends, the gestalt principles of perceptual psychology.

Most musical patterns are pitch-time patterns, or combinations of rhythm and melody. Essential to the perception of such patterns is the concept of *accent* (JONES, 1987) or the *salience* of an event (PARNCUTT, 1989, 1994a, 2003). The elements of a pattern are usually not perceived to be equally important, even if they are physically identical (e.g. in rhythm: POVEL & ESSENS, 1985; in melody: HURON & ROYAL, 1996). The more important (or salient) elements become points of reference (or accents), relative to which the rest of the pattern is perceived. The resultant hierarchical structures facilitate cognitive processing (DEUTSCH, 1999; KRUMHANSL, 1990).

An interesting and difficult question that tends to be avoided in music perception research is the relationship between pattern perception and musical emotion. In this paper, I will attempt to address this question by surveying typical patterns in western music and relevant theories of their perception and of musical emotion. I will consider the interdisciplinary nature of the question, the inherent ambiguity and multiplicity of musical patterns, and relevant theories of learning and culture.

## 2. Interdisciplinarity

The *(natural) sciences* shed light on the perception of musical patterns by, for example, considering how human sensory systems evolved to promote survival in typical environments. Such environments include not only a variety of sound sources, such as other humans and animals that may be dangerous as well as all kinds of everyday sounds, but also surfaces that obstruct, reflect or absorb the passage of sound between source and perceiver. In response

to these physical constraints, the auditory system has developed the ability to separate partials with different frequencies (TERHARDT, 1998), while retaining the ability to separate successive sound events and process complex temporal patterns (rhythms), and to store enough information about pitch-time patterns (primarily speech and environmental sounds) to allow them to be recognized. Physiologically, these processes are enabled by the basilar membrane in the cochlea and by the brain's neural networks (SANO & JENKINS, 1989).

The *humanities* address the cultural environments that shape sensory systems within the lifespan of each individual. A full understanding of the perception of musical patterns is not possible without reference to the specific patterns available to be perceived in a given historical or ethnological context. To understand the perception of pitch-time patterns in tonal western music, it is necessary to consider how the syntax of tonal western music evolved (EBERLEIN, 1994). For example, a theory that accounts for the finality of the authentic (dominant-tonic) cadence should equally be able to explain the finality of the double-leading-tone cadence of the 14th Century. A theory that purports to explain the avoidance of parallel fifths and octaves should be consistent with the considerable variations in the strictness with which parallel fifths and octaves were avoided in different periods of western music history.

Where the *experience* of music is of primary interest, it is necessary to address philosophical and phenomenological issues such as the reality of experience, the question of mind-body duality, and the separate existence of knowledge, information and culture (POPPER & ECCLES, 1977). For while the physical world and a cultural and historical context are necessary prerequisites for the existence of music, it is only in the world of experience or the "mind" that the "impact of music" – the topic of this volume – is felt. The experience of pattern recognition, like almost every other experience, is linked to emotional responses, whose understanding is essential for an understanding of music's essence.

#### 3. Gestalt principles

The most important function of the auditory (or visual) system is to identify environmental objects on the basis of the sound (or light) that they emit. To do that, it is necessary to recognize patterns of sound (or light). This involves grouping the parts of a pattern together and ignoring stimuli that do not belong to the pattern. It involves separating the foreground from the background, and understanding the foreground.

The well-known gestalt principles of similarity, proximity, good continuation, closure and so on operate similarly in the visual and auditory domains. According to the principle of *proximity*, the tones of a melodic phrase should be close to each other in pitch and time if the melodic phrase is to be perceived as a unit, and a chord is more likely to fuse (blend) if its tones begin synchronously (BREGMAN, 1990). According to the principle of *closure*, a harmonic complex tone can be perceived as such even if some of the harmonics are inaudible due to masking by other sounds (and even if one of those is the fundamental). According to the principle of *common fate*, the synchronous movement of an object's parts or contours suggests that they belong together; in the auditory case, a complex sound whose partials have coherently varying frequencies or amplitudes is likely to be perceived as a unit, even if the partial frequencies are not harmonic (MARIN & MCADAMS, 1991).

The gestalt principles do not necessarily apply as clearly or strongly in hearing as they do in vision. For example, *good continuation* is a stronger cue in vision – it is easy to see two lines crossing as just that, but hard to hear two melodies cross unless other cues such as timbral or loudness differences between the melodies support this interpretation (DEUTSCH, 1999). This difference between vision and hearing may be explained in terms of experience with real physical environments. We often see lines cross that are otherwise identical (consider the letter x, or two similar sticks lying on top of each other the ground). But we seldom hear two harmonic complex tones with about the same timbre, loudness and pitch, such that one is gradually rising in pitch and the other falling; and if we do (for example, when listening to someone talking against a background of other people talking), we are

generally only interested in one of the two tones. Thus, part-crossing is avoided in music, but line-crossing is commonplace in the visual arts.

This example suggests that gestalt principles primarily reflect regularities in auditory stimuli produced by objects in typical human environments. The details of gestalt perception are acquired from the outside world, and may be acquired either within the course of a single lifetime (by learning) or over the course of several generations in which organisms that more successfully recognize specific patterns are more likely to survive and pass on their ability the next generation.

According to the gestalt principle of *closure*, a pattern can be perceived even if it is incomplete. This principle is important, and surprisingly analogous, for the perception both of pitch in a single sound and a rhythm's underlying pulse. The pattern in each case is physically equally spaced: the partials of a harmonic complex tone are equally spaced in frequency, and a musical pulse is equally spaced in time. In both cases the equal spacing is not always exact: for example the partials of piano tones are stretched relative to the harmonic series, and rhythmic beats deviate both randomly and systematically (or intentionally) from metronome ticks. Since (exactly or approximately) equally spaced patterns in frequency and time exist in abundance in the physical world, it is not immediately clear whether equally spaced patterns are recognized on the basis of their inherent regularity or their familiarity. While the inherent regularity of these stimuli is obvious, equally strong arguments exist to support the learning assumption: complex tones whose partials are equally spaced but are not clearly harmonics of an audible fundamental (e.g., a spectrum comprising the frequencies 45, 145, 245, 245 Hz) do not fuse well; and regular rhythms are only heard as rhythmic (that is, they are only associated with dance) within the range of human heartbeats and footsteps (i.e. with periods of about a second or less).

## 4. Ambiguity and multiplicity

If a given pattern has different interpretations, and an observer is generally only aware of one of these at a time, we may say that the pattern is *ambiguous*. If an observer is simultaneously aware of more than one interpretation, we may speak of *multiplicity* (PARNCUTT, 1989).

Both ambiguity and multiplicity are commonplace in music theory. Any piece of music may be analyzed or interpreted in several equally good ways (BENT & POPLE, 2001). Even within a given analysis, ambiguity and multiplicity can exist within and between different levels. An example of ambiguity within the level of the basic beat or tactus is *hemiola*: it is intuitively impossible to simultaneously maintain a cognitive representation of both 3/4 and 6/8 meter, but it is not hard for trained listeners to switch quickly from one to the other. When different interpretations exist simultaneously on different hierarchical levels, we may speak of multiplicity; for example, surface harmonies can contradict prolonged structural harmonies (TEMPERLEY, 2001).

Music theorists are sometimes reluctant to acknowledge ambiguity. Generations of North American music students have been drilled by their Schenkerian instructors that the tonic six-four chord, when it immediately precedes a dominant, is a double suspension, and therefore not a tonic but a dominant (FORTE & GILBERT, 1982). Another interpretation is that the tonic six-four has both tonic and dominant function simultaneously. This very ambiguity (or multiplicity) is one of the reasons why this chord is perceived as tense and as requiring of resolution in common-practice tonality.

Ambiguity may be divided into conscious and non-conscious ambiguity. MEYER (1956) was concerned with the conscious case: "A sound term can have different meanings at different times, but this does not prove that the term, or the hypothetical meaning which it first has, is ambiguous. (...) If we are certain in our minds as to the meaning of a sound term when it first appears, then it is not ambiguous at that time" (p. 51). This contrasts with TERHARDT's (1974) concept of the pitch ambiguity of a complex tone, according to which a listener hears the pitch to be different on different occasions – but on each occasion is aware of nothing other than "the" pitch at that time. Since the borderline between conscious and non-conscious in music perception is operationally so difficult is difficult to locate, and one can so easily become the other (depending on listener and context), I will ignore Meyer's point and consider both kinds of ambiguity.

Regarding non-conscious ambiguity, switching mentally between two perceptual interpretations of a musical event or passage can cause the identity of the music to change completely. For example, when SLOBODA (1983) asked pianists to sight-read a series of passages, two of which differed only in the placement of the barlines, none of the planists noticed the relationship between the two almost identical passages. The perceptual difficulty of switching from one pulse or metrical framework to another is reflected by the term rhythmic hysteresis (LARGE, 2000; PEPER, BEEK, & VAN WIERINGEN, 1995). In other cases of ambiguity, the identity of the music does not change, for example when we hear the same chord relative to a different root, even if the two root candidates are harmonically distant (e.g., tritone substitutions in jazz), or a passage relative to a different tonic. In these cases the distinction between ambiguity and multiplicity becomes unclear. An analytically perceived complex sonority appears to comprise several tones, so the sonority has multiple pitches; the same sonority heard holistically is ambiguous with respect to its pitch. In both cases, we are talking about the same pitches or tone sensations, whose saliences (and hence the degree of multiplicity or ambiguity) may can be predicted either in the time domain (CARIANI & DELGUTTE, 1996) or in the frequency domain (TERHARDT, STOLL, & SEEWANN; 1982; PARNCUTT, 1989).

Ambiguity and multiplicity can emerge from competition between gestalt principles (BREGMAN, 1990). Consider the role of the principle of *proximity* in determining the perception of implied polyphony: a single melodic line that implies two or more lines. If temporal proximity more strongly determines grouping than pitch proximity, a single line will be perceived – one tone after the other. If the effect of pitch proximity is stronger than the effect of temporal proximity, the higher tones may separate perceptually from the lower tones, creating two separate lines, each with its own shape and rhythm – implied polyphony.

The principle of *closure* leads to ambiguities when a stimulus can be matched to more than one possible pattern. In the simplest case, an equally-spaced spectrum of frequencies can be matched to equally-spaced pitch templates in different octave registers; similarly, a rhythm that is mainly equally spaced may be matched to a pulse either at the same speed or at half or double the speed (PARNCUTT, 1994b). The psychological reality of both kinds of ambiguities may easily be demonstrated in psychoacoustical experiments in which listeners match a pure tone to a complex tone and regularly commit "octave errors" (TERHARDT, STOLL, SCHERMBACH, & PARNCUTT, 1986) or tap out the underlying beat of a rhythm on different metrical levels (EDLUND, 1995; PARNCUTT, 1994a; POVEL & ESSENS, 1985). Another example of ambiguity is the perception of musical phrasing, which is primarily determined by the principles of similarity and proximity (DELIÈGE & MÉLEN, 1997). It is generally possible to segment a piece of music in different ways into phrases, and to join phrases in different ways into different hierarchical levels, because various cues compete to determine phrase endings and beginnings: the time gap between phrases is usually longer than between successive notes within a phrase (a strong cue to segmentation; LERDAHL & JACKENDOFF, 1983), and phrases tend to rise at the start and fall at the end (a weak cue; HURON, 1996).

These examples of perceptual ambiguity suggest that musical patterns are *inherently* ambiguous, and raise the question of the extent to which ambiguity is deliberate or preferred. Music tends to be preferred if its *complexity* is moderate (not too simple, not too complex: BERLYNE, 1974). Ambiguity, seen as an aspect of complexity, may be one of the most interesting aspects, and possibly a definitive characteristic, of literature, poetry and music. Musical ambiguity is like a game that composers and performers play with listeners' expectations. By contrast, the main function of everyday speech is to communicate explicit information, making ambiguity generally undesirable.

In both language and music, context disambiguates. That is, an element that is ambiguous when presented in isolation may lose some or all of that ambiguity when it appears in context. "The fact that as we listen to music we not only interpret present stimuli on the basis of past events but also view past events and expect future ones on the basis of present stimuli means that a process at first felt to be ambiguous may later be seen as less so. Similarly processes at first considered unambiguous may later be seen as involving or leading toward ambiguity. In other words, ambiguity depends upon the structural architectonic viewpoint taken toward the stimulus series in question" (MEYER, 1956, p.52).

## 5. Specific musical patterns

Musical patterns include rhythms, melodies, chords, chord progressions, and counterpoint (combinations of simultaneous melodies). The reference points relative to which musical patterns are perceived include downbeats, roots and tonics. The recognition of a specific pattern within the auditory scene, such as a known melody, involves not only the segmentation of the musical "surface" into chords, melodies, and so on and the identification of salient elements of the pattern, but also interaction between that surface and memory (e.g. the memory for musical melodies). How is this memory organized?

In the case of melody, DOWLING (1978) demonstrated the importance of contour for musical memory. A melody is recognized if its pitches rise and fall in a given sequence, even if the pitch and time intervals between successive tones are changed. But they should not change too much: it helps if the distinction between steps and leaps, and – in the case of rhythm – between relatively long and short tones (FRAISSE, 1963) is maintained. This suggests that a kind of information reduction (or higher-level categorical perception) is involved in memory for melodies. That would be consistent with the surprisingly large number of melodies that most people can recognize, and the short time interval within which recognition takes place.

Modern musical cultures typically include a very large number different melodies, each of which has its own identity (title, text, composers, performers, symbolic meanings and so on). These many melodies are constructed from a limited number of possible contours and rhythms. Melodies may be perceived to be similar according to a variety of structural criteria (DELIÈGE, 2001) as well as extramusical associations, the context in which the music is heard, and the listener's social identity (BAUMANN & HALLORAN, 2004; CLARKE & DIBBEN, 1997). Since similar melodies can be confused with each other (e.g., HÉBERT & PERETZ, 1997), a melody may have an ambiguous or multiple identity. This principle is exploited in everyday composition and improvisation, as an original statement of a melody is varied. A variation that is too distant from the original will, at least at first, sound like a new melody. The ambiguity between "new " and "derived" may be regarded as part of the composition. Music analysis often involves discovering and exploring relationships of this kind.

Empirically, it is easy to test whether someone recognizes a melody – just play it to them and ask them to name it or sing it back. It is not so easy to test the recognition of a chord, since only musicians with the appropriate training can label a chord in terms of its root and quality. Cognitive psychologists addressing questions of musical structure (e.g., KRUMHANSL, 1990) have paid surprisingly little attention to the phenomenon of chord roots, by comparison to the perception of melody and of tonality. This is surprising, given that a chord's quality depends on the intervals between the root and the other tones: to recognize a chord according to its quality (e.g. minor triad, dominant seventh), it is necessary to identify a reference pitch and a pattern of intervals relative to that pitch. Moreover, standard harmonic theory, in both classical traditions and in popular music and jazz, conceives of harmonic progressions in terms of roots.

Generations of music theorists tried but failed to develop a satisfactory general theory of the root of a chord that could account for the root of the minor triad – especially during the latter part of the 19th century. This research program not only failed to discover the specific functional nature of the relationship between the harmonic series and chord roots – it also did not take proper account of the ambiguity of chord roots, and to develop a plausible, general account of that ambiguity. If, as TERHARDT (1974) and PARNCUTT (1988) assumed, the perception of a chord root involves much the same process as the perception of the pitch of a single complex tone, we may expect the root of a chord to be ambiguous. For example, the root of a Dminor chord functioning as a supertonic triad in the key of C major can be regarded either as D (more likely if the chord is in root position) or as F (more likely if the chord is in first inversion). This ambiguity is reflected in RIEMANN's (1893) functional approach, in which the supertonic harmony is called *Subdominantparallel*.

Of course, tonality is ambiguous, too. Music theorists often disagree about the key of a musical passage: Schenkerians hear entire pieces relative to a single tonic, while other analysts speak of sequences of modulations to near and far tonalities in the same pieces. Listeners in psychoacoustical experiments disagree about which tone represents the tonic, or sounds most final, after a given passage, and both theorists and psychological models disagree about the exact point in time at which modulations occur (AUHAGEN, 1994; HURON & PARNCUTT, 1993; VOS & LEMAN, 2001). Since ambiguity is typical of pattern recognition, it is appropriate to regard the perception of tonality, like the perception of chord roots, as a form of pattern recognition. If so, what is the nature of the pattern?

>>>insert figure 1: major and minor chords and keys<<<

Musical keys are usually instantiated by tonal cadences such as progressions from subdominant to dominant to tonic. In order to explore how cadences determine tonality, KRUMHANSL and KESSLER (1982) asked listeners how well individual tones in the chromatic scale follow such progressions. The resultant tone profiles (Figure 1) quantified the music-theoretical intuition that the tonic tone is more stable than the dominant, which in turn is more stable than the mediant, which is more stable than the other diatonic tones in a key, which are more stable than the non-diatonic tones.

The question then arises as to the origin of the profiles. Since they evidently existed psychologically in the 17th century, by which most commentators (e.g., DAHLHAUS, 1967, EBERLEIN, 1994; RANDEL, 1971) agree that majorminor tonality had completely "emerged", a causal explanation must lie in an earlier period of music history; and since they are essentially patterns of pitch, their origin must involve (other) patterns of pitch that were familiar to listeners of the 17th and earlier centuries.

A good candidate for such a pattern is the pattern of pitch evoked or suggested by a major or minor triad. These sonorities gradually became more prevalent in polyphonic music over a period spanning several hundred years, from the 13th to the 16th centuries (RANDEL, 1971). During that time, they also increasingly functioned as reference points (departure and resting points) in the gradually "emerging" major-minor system.

According to PARNCUTT (1988), the perceptual salience of a particular pitch class within a chord is a measure of the probability that that pitch class will function as chord's root. That probability also depends on the temporal context in which the chord appears. My tone profiles (or pitch-salience profiles) of major and minor triads are compared with Krumhansl's tone profiles of major and minor keys in Figure 1. The correlation is high, as KRUMHANSL and KESSLER (1982) themselves noted when comparing tone profiles following an isolated (tonic) triad with profiles following cadential progressions. Thus, pitch salience within the tonic triad is a good predictor of stability within the corresponding major or minor key. Put another way, Krumhansl's key profiles are none other than pitch-salience profiles of tonic triads, and can be entirely reduced to the chord-root phenomenon.

In summary, neither the root of a chord nor the major-minor tonality of a passage can easily be defined, determined experimentally, or predicted using a model. Recent mainstream research in cognitive psychology (e.g., VOS & LEMAN, 2001) has devoted much more attention to the more difficult problem of tonality. In so doing, it has largely ignored the role of chord roots in determining tonality. The approach just described establishes a connection between the two that is both qualitative and quantitative. That is, it is not only informative about the nature and function of roots and tonics – it is also enables quantitative predictions to be made on the basis of relatively few assumptions or adjustable parameters (the principle of parsimony). It also

explains why Krumhansl's tone profiles are so informative, although they (intentionally) say nothing about voice-leading aspects of tonality (BUTLER, 1989). The comparison with the pitch-salience profile of the tonic triad makes it clear that (and why) the key profiles encapsulate only the vertical or harmonic aspect of major-minor tonality, and entirely ignore the melodic or voice-leading aspect. In this sense, they may be regarded as cognitive representations of DAHLHAUS's (1967) *harmonic tonality*.

These considerations offer a partial answer to the question of what exactly is "recognized" or "identified" when we hear music relative to a given tonality or tonic. In explaining the inherent ambiguity of tonality, they indirectly contribute to an understanding of musical emotion in tonal music.

## 6. Emotion

In his seminal monograph on musical emotion, MEYER (1956) distinguished between two schools of thought. "Referentialists" claim that musical emotions are generated externally, through *associations* with extra-musical experiences (e.g., vocalizations in infant-adult vocal play, in which melodic contours have specific meanings; PAPOUSEK, PAPOUSEK, & SYMMES, 1991). "Absolutists" claim that musical emotions are generated internally, through the *structure* of the music itself – or through supposedly "purely musical" processes, such as the fulfillment or denial of expectations within musical pitch-time structures. Clearly, both are right. Musical emotions can both be influenced from outside and be generated within the music. The interesting point for the present purpose is that both aspects of musical emotion involve pattern recognition.

Emotional associations between musical pitch-time patterns and non-musical patterns are commonplace. Consider the experiential (perceptual and emotional) world of an infant. Infants are very good at extracting and storing information from their environment. They are also very interested in the emotional state of the adults that surround them, because their ability to communicate emotionally – to detect and manipulate the emotional state of others – is essential for their survival. When babies are active, they need a *happy* adult to entertain them; when they are sleepy or sleeping, they need a

*tender* adult to look after them. Adults who are *sad* or *afraid* are less useful to infants, and those who are *angry* can be downright dangerous. Infants recognize these states by listening to adult speech (PAPOUSEK et al., 1991). Happy speech is relatively fast and loud, clearly articulated, and covers a relatively wide pitch range. Tender speech is slower and quieter; the individual tones have softer attacks, and are more legato. Thus, long before infants start to understand the *lexical* meaning of speech (that is, the meaning that is maintained when speech is written down), they are good at deciphering its *gestural* meaning. Beyond these general states or moods, speech can include gestures that communicate specific meanings. For example, a sudden rise in pitch can be used to elicit attention or encourage the taking of a turn. Such gestures also have strong emotional connotations.

Emotional cues in speech (e.g. the loudness, abruptness, accentuation and distortion of angry speech) correspond closely to emotional cues in music, suggesting that the speech cues determine the musical cues. In both speech and music, listeners are quite good at decoding such cues, although the message can sometimes be ambiguous (JUSLIN & PERSSON, 2002). The existence of complex relationships between emotions and patterns of pitch, loudness and time (rhythm) suggests that an infant's experience of the voice of adults, which presumably begins well before birth, may be an important precursor of music and may even represent music's origin (PARNCUTT, 1993). Since the internal sounds of the mother's body are structurally similar to melody (voice) and rhythm (footsteps, heartbeat), the emotions generated by music may be closely related to maternal love as experienced by the fetus and infant.

The emotionality of musical pitch-time patterns may also be explained by the structure of those patterns. MEYER (1956) assumed that "emotion or affect is aroused when a tendency to respond is arrested or inhibited" (p. 14). At any point in a piece of music, listeners have expectations that the music will continue in a specific fashion; if no such expectations exist, the music is meaningless or incomprehensible. For example, if the music follows a pattern that is typical of the end of a phrase, and the listener is familiar with music

with this kind of syntax, that listener will expect the phrase to end soon. According to Meyer, the emotional flavor of a passage is influenced by whether such expectations are realized or not: the realization of implications is associated with feelings of fulfillment, satisfaction or relief, whereas nonrealization is associated with surprise, disappointment or frustration. "Ambiguity is important because it gives rise to particularly strong tensions and powerful expectations. For the human mind, ever searching for the certainty and control which comes with the ability to envisage and predict, avoids and abhors such doubtful and confused states and expects subsequent clarification" (MEYER, 1956, p.51).

This theory sounds fine in principle, but for many it is too abstract to connect with real emotional experiences. The findings of SLOBODA (1991) made Meyer's theory more concrete. His experimental participants were keen listeners to "classical" music. They showed him the exact points in musical scores that corresponded to their strongest emotional experiences. Certain kinds of patterns occurred repeatedly in the data: harmonic progressions descending by a cycle of fifths to the tonic, passages containing repeated melodic appogiaturas, melodic and harmonic sequences, enharmonic changes, harmonic or melodic acceleration to a cadence, the delay of final cadences, new or unprepared harmonies, sudden dynamic or textural changes, repeated syncopations, and prominent events that arrive earlier than expected. These results are consistent with Meyer's theory, because in all of these cases, listeners had strong expectations that were either fulfilled or denied.

How can expectations explain the emotional quality of one of the structures to emerge from Sloboda's study, the harmonic progression descending by fifths to the tonic? In Western music of the 18th and 19th centuries, progressions in which roots fell by a fifth or third (e.g., C major to F major or A minor) were much more common than progressions in which roots rose by a fifth or third (e.g., C major to G major or E minor) (cf. data of EBERLEIN, 1994). Modern listeners who are familiar with this repertoire should, therefore, be entirely unsurprised by such progressions. According to this logic, such progressions should be uninteresting and hence unemotional.

In searching for a resolution to this paradox, we might begin by asking why the asymmetry between rising and falling fifths and thirds between chord roots exists at all. AGMON (1994) proposed that a progression of two chords is preferred or "makes a stronger effect" (p. 245) if the root of the second chord is not included among the tones of the first chord. Agmon's *root newness* principle is reminiscent of Meyer's *implication-realization* idea: a chord progression may "make a stronger effect" if the root of the second chord is implied by the first chord, but not played. The implication arises through familiarity: since progressions through a falling fifth or third are more common than progressions through a rising fifth or third, listeners tend to expect the falling progressions.

Agmon's root newness and Meyer's implication-realization may be combined into a single concept based on the *perceptual saliences* of the involved pitches and the resultant *pitch commonality* of the progressions (PARNCUTT, 1989). According to Terhardt's theory of pitch perception, each individual chord in a progression implies pitches that are not physically present. Both major and minor triads imply pitches a fourth and a sixth above the root: CEG implies F and A, and CEbG implies F and Ab (see Figure 1). These implied pitches correspond to missing fundamentals – fundamentals of partials that are physically present in the chord, but belong to different tones. If, for example, the fundamental (1st harmonic) of a harmonic complex tone is an F (in any octave), then the 3rd harmonic is a C and the 5th is a G; therefore, C and G together (and hence CEG or CEbG) can imply F. If the 1st harmonic is an A, then the 3rd harmonic is an E and the 7th is a G; therefore, E and G together (and hence CEG) can imply A. And so on. The reason for the asymmetry in the distribution of chord progressions (falling fifths and thirds between roots being more common than rising) may be that the pitch salience profile of major and minor triads is itself asymmetrical. In the falling-fifth progression from C major to F major, the implied tones in the first chord (F and A) are realized in the second chord; but in the rising-fifth progression

from C to G, the implied pitches in the first chord (F and A) are not realized in the second, and the new tones in the second chord (B and D) are more weakly implied in the first (D is implied more weakly than F or A, and B is not implied at all, according to the pitch model). A similar line of argument can explain why falling-third progressions are more prevalent than rising.

In both cases, the second chord in the more common progression may be considered as a realization of an implication created by the first. The perceptibility of this implication can easily be demonstrated: In experiments in which musical chords are followed by probe tones and listeners asked how well the tone goes with the chord, listeners clearly distinguish between tones that go well (such as F or A following C major) and tones that do not go well (such as F# or Ab following C major) (PARNCUTT, 1993). Thus, the emotional effect of harmonic progressions descending by a cycle of fifths to the tonic may be linked to the implication-realization effect just described. The repeated fulfillment of expectations may lead to a strong feeling of relief or even abundance. The expectation is generated not only by the implied and realized pitches but also by the predictability of the pattern.

This kind of academic theorizing sounds impressive and can keep psychologists and philosophers happily occupied for hours, But it is hardly sufficient to explain the *intensity* of musical experiences as repeatedly and systematically observed, for example, by GABRIELSSON & LINDSTRÖM (2003). My theory of the prenatal origins of music (PARNCUTT, 1997) is an attempt to solve this problem. It assumes that the source of musical emotion is mainly external to music. Musical emotions are held to be cultural elaborations of emotions perceived and encoded in non-linguistic memory by the fetus, as it associates maternal emotions with typical patterns created the internal sounds of its mother's body – her voice, breathing, heartbeat, digestion, footsteps and other movements.

Is there a causal connection between ambiguity and emotion? If so, it should be valid outside of music – specifically, in the visual arts. But art theory does not offer a clear answer to this question. Most psychology texts include a discussion of famous ambiguous figures, such as the one that can be seen either as two faces or as a vase, depending on which part of the figure is seen in the foreground (figure-ground perception), or the one that can be seen either as a young or an old woman (BORING, 1930). But there is no suggestion that this ambiguity may *itself* have an emotional connotation. The famous ambiguous paintings by Escher (e.g. the continuously rising staircases) are not regarded as particularly emotional – unless a fascination with ambiguity is regarded as emotion.

This suggests that the association of ambiguity and emotion in music observed by Meyer is not a causal relationship. The ambiguity does not cause the emotion; instead, something else must cause both of them. That something is evidently the process of pattern recognition itself. Every pattern that is recognized has associations, which include emotions – otherwise the pattern would be meaningless and there would be no motivation to recognize it. And pattern recognition generally results in ambiguity: if the pattern is regarded as a template that is matched to a stimulus, there are generally different template positions that produce a reasonable match. Seen in this light, the ambiguity-emotion connection is an artifact.

An approach of this kind can explain the emotional quality of the simplest musical passages, such as a single tone or a string of equally space beats. The equally-spaced pattern that is recognized in each case is not emotionally neutral, if the harmonic pitch pattern is linked to the spectrum of a voiced speech sound, and the rhythmic pulse is linked to human heartbeats and footsteps. So even at this simple level, where (as we have seen) pattern recognition already produces ambiguity, we would expect pattern recognition to be associated with emotion.

## 7. Auditory learning

The theoretical ideas presented in this paper rely strongly on the assumption that the auditory system can quickly learn a large number of auditory patterns and associate them with specific meanings (GIBSON, 1953). This is consistent with a range of empirical observations. BREGMAN's (1990) theory of auditory scene analysis assumes that the auditory system learns complex pitch-timeloudness patterns and uses this acquired knowledge to "parse the auditory" scene", that is, to determine which parts of the scene belong together and hence to which environmental sound sources. DOWLING's (1978) approach to the question of melody recognition focuses on how melodies are encoded in memory and is consistent with the ability of people to learn a very large number melodies, regardless of their musical training. PALMER and KRUMHANSL (1987) assumed that musical meters such as 2/4 and 6/8 have characteristic profiles of strong and weak beats on different hierarchical levels, and that these are learned by exposure to music in these meters. Central to BURNS' (1999) account of musical intonation is the idea that intonation is preferred if performers and listeners are familiar with it; thus, instrumentalists such as violinists who can easily vary their intonation over a wide range may nevertheless stay close to the familiar tuning of the piano. EBERLEIN's (1994) analysis of the evolution of western tonal syntax relies heavily on the notion that the people (musicians, listeners) of a given historical period are familiar with the sound of the music of that period including the details of its syntax, and that developments in syntax are always constrained or driven by this familiarity.

Physiologically, all these kinds of auditory learning can easily be explained by the concept of neural networks (GJERDINGEN, 1990). The same concept can explain the ambiguity of pattern recognition and how it is possible for the response to a given pattern to change when a listener is exposed to new patterns.

The pitch model of TERHARDT et al. (1982)<sup>1</sup> is consistent with the assumption of perceptual learning: the pattern of spectral pitches within a single complex tone (i.e. the harmonic series) is assumed to be learned from exposure to harmonic complex tones in early life (although this information could also be genetically transmitted and linked to periodicity). The concept of rhythm that I presented in PARNCUTT (1994a) relies on a similar idea: the perception of pulse in musical rhythms is assumed to be based on the perception of extramusical pulses in heartbeats and footsteps.

At this level, the mystery of musical meaning may be considered to be solved. At another level, it remains a mystery. Music psychologists are still a long way away from predicting emotional responses to specific music experienced in specific situations. Neural networks cannot be considered in isolation from the human environment; it is essential to consider the cultural context. But that is constantly changing and, at least for the moment, scientifically intractable. To

<sup>&</sup>lt;sup>1</sup> This paper relies on a spectral approach to pitch perception. Since periodicity in the time domain corresponds to harmonicity in the frequency domain, and pitch perception primarily involves harmonic complex tones, spectral and temporal approaches to pitch often differ little in their predictions and implications. Periodicity information is present throughout the auditory system, since neural firing in different parts of the auditory pathways is to a large extent phase-locked (cf. PATTERSON, this volume). However, the brain also includes tonotopic representations of pitch (PANTEV, HOKE, LUETKENHOENER, & LEHNERTZ, 1989; SCHREINER & LANGNER, 1997). The perception of slightly inharmonic sounds such as piano tones can be explained either in the frequency or the time domain - in both cases it involves permissible departures from harmonicity or periodicity, which may be tuned to margins of error present in physical sounds. Terhardt's attempts to algorithmically predict pitch shifts did not produce clear enough evidence to disprove time-domain theories; further experimental work is necessary to clarify this question. An advantage of a time-domain approach is its proximity to the real physical signal, both in the air and in the brain. By contrast, frequency is a not a direct property of sound, but a parameter that is derived from frequency analysis (TERHARDT 1998); in this sense, a theory of pitch based on frequency is neither direct nor parsimonious. An advantage of the frequency-based approach is that it is relatively easily to implement in a computer algorithm and to apply in music-theoretical contexts. Terhardt's model estimated the saliences of all pitches evoked by a sound, which makes it both easy to test and to criticize. Temporal models have been slow to emulate this feature, which is particularly relevant to the question of music-structural ambiguity.

address a problem of such enormous dimensions, it will be necessary to bring humanities and sciences closer together.

#### 8. Culture

"Culture" is often used as a collective term for the physical and intellectual products of human society. It includes commonly held ideas, their intersubjective generation and development, and their sharing and communication. It also includes all forms of art, including music.

From a cognitive-psychological viewpoint, culture is limited by cognitive constraints such as memory limitations (D'ANDRADE, 2001). The syntax of music, like any other aspect of culture, is subject to cognitive constraints and interactions with other aspects of culture. Perception of music within a given culture is inextricably linked to the historical development of the syntax and semantics of the development of musical language in that culture. For example, the syntax of western music in the middle ages and renaissance depended strongly on the compositional conventions of the time. These depended in part on the social functions of the music in question, and partly on universal perceptual principles (EBERLEIN, 1994; HURON, 2001). Compositional conventions influenced the music-stylistic norms of their time, that is, the statistical properties of music (how often specific patterns happened, and how often they were followed by other specific patterns). These in turn influenced how people perceived music (based on information stored in neural networks), such as their pitch-time expectations at a particular point in a piece of music, how people reacted emotionally, and what meaning they ascribed to the sounds that they heard. There may therefore be no direct connection between universal principles of perception (such as the gestalt principles, or the perception of harmonic complex tones in speech) and the principles of music perception.

These scientifically oriented ideas are consistent with parallel developments in the humanities, where the meaning of a musical text (or of any other text) is considered to be multiply determined; it is not restricted to its original, intended meaning but develops as the culture within it is embedded develops (KRAMER, 1993). Scholars in the humanities stress that musical actors (performers, composers, improvisers, listeners, consumers) are constantly learning and developing, and that everything they learn depends on the cultural context of which they themselves are part. At the same time, scientists stress that the auditory system is also constantly "learning" – within the same cultural context. The concepts of learning, context, ambiguity and multiplicity link together the apparently incompatible disciplines of *psychoacoustics* and *cultural studies*. This combination appears to have considerable potential for productive exploitation in future research.

#### References

Agmon, E. (1994). Principles of chord progression. In E. Deliège (Ed.), *Proceedings of the Third International Conference on Music Perception and Cognition* (pp. 245-246). Liège, B: Centre de Recherches et de Formation Musicales de Wallonie.

 Auhagen, W. (1994). Experimentelle Untersuchungen zur auditiven
 Tonalitätsbestimmung in Melodien. Kölner Beiträge zur Musikforschung (Band 180). Kassel: Bosse.

- Baumann, S., & Halloran, J. (2004). An ecological approach to multimodal subjective music similarity perception. In Parncutt, R., Kessler, A., & Zimmer, F. (Eds.), *Conference on Interdisciplinary Musicology: Abstracts* (pp. 44-45) and *Proceedings* (http://gewi.uni-graz.at/~cim04/).
- Bent, I. D., & Pople, A. (2001). Analysis. In S. Sadie (Ed.), *The new Grove dictionary of music and musicians* (2nd ed.), 1, 526-589.
- Berlyne, D. E. (1974)., *Studies in the new experimental aesthetics*. Washington: Hemisphere.
- Boring, E. G. (1930). A new ambiguous figure. *American Journal of Psychology, 42*, 444-445.
- Bregman, A.S. (1990). *Acoustical scene analysis*. Cambridge, Massachusetts: MIT Press.
- Burns, E. M. (1999). Intervals, scales and tuning. In D. Deutsch (Ed.), *Psychology of Music* (2<sup>nd</sup> ed., pp. 215-264). San Diego: Academic Press.
- Butler, D. (1989). Describing the perception of tonality in music: A critique of the tonal hierarchy theory and a proposal for a theory of intervallic rivalry. *Music Perception*, 6 (3), 219-242.
- Cariani P. A., & Delgutte B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology 76*, 1698-1716.
- Clarke, E., & Dibben, N. (1997). An ecological approach to similarity and categorisation in music. In M. Ramscar et al. (Eds.), *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation* (pp. 37-41).
   Department of Artificial Intelligence, Univ of Edinburgh.
- Dahlhaus, C. (1967). Untersuchungen über die Entstehung der harmonischen Tonalität. Kassel: Bärenreiter.

- D'Andrade, R. (2001). A cognitivist's view of the units debate in cultural anthropology. *Cross-Cultural Research: The Journal of Comparative Social Science*, *35*, 242-257.
- Deliège, I. (2001). Similarity perception  $\leftrightarrow$  Categorization  $\leftrightarrow$  Cue abstraction [Special issue]. *Music Perception, 18* (3).
- Deliège, I., & Mélen, M. (1997). Cue abstraction in the representation of musical form. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition* of music (pp. 387-412). Hove, East Sussex: Psychology Press.
- Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 299-348). New York: Academic.
- Dowling, D. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341-354.
- Eberlein, R. (1994). *Die Entstehung der tonalen Klangsyntax*. Frankfurt/Main: Lang.
- Edlund, B. (1995). Making meter evident On the playing of an ambiguous Bach melody. *Musikpsychologie*, *12*, 28-41.
- Forte, A., & Gilbert, S.E. (1982). *An introduction to Schenkerian analysis.* New York: Norton.
- Fraisse, P. (1963). *The psychology of time.* New York: Harper & Row.
- Gabrielsson, A., & Lindström, S. (2003). Strong experiences related to music. A descriptive system. *Musicae Scientiae*, *7*, 157-217.
- Gibson, E. J. (1953). Improvement in perceptual judgments as a function of controlled practice or training. *Psychological Bulletin,* 50,z 401-431.

Gjerdingen, R. O. (1990). Categorization of musical patterns by selforganizing neuronlike networks. *Music Perception, 7*, 339-370.

- Hébert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory & Cognition*, 25, 518-533.
- Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology, 10,* 3-23.
- Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception, 19*, 1-64.

- Huron, D., & Parncutt, R. (1993). An improved model of tonality perception incorporating pitch salience and echoic memory. *Psychomusicology, 12,* 154-171.
- 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.
- Juslin, P. N., & Persson, R. S. (2002). Emotional communication. In R. Parncutt & G. E. McPherson (Eds.), *Science and psychology of music performance* (pp. 219-236). New York: Oxford University Press.
- Kramer, L. (1993). *Music as cultural practice, 1800-1900*. Berkeley: University of California Press.
- 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 representation of musical keys. *Psychological Review*, *89*, 334-368.
- Large, E.W. (2000). On synchronizing movements to music. *Human Movement Science, 19,* 527-566.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.
- Marin, C. M. H., & McAdams, S. (1991). Segregation of concurrent sounds. II: Effects of spectral envelope tracing, frequency modulation coherence, and frequency modulation width. *Journal of the Acoustical Society of America*, *89*, 341-351.
- Meyer, L.B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
- Palmer, C., & Krumhansl, C. L. (1987). Independent temporal and pitch structures in determination of musical phrases. *Journal of Experimental Psychology: Human Perception & Performance*, 13,116-126.
- Pantev, C., Hoke, M., Luetkenhoener, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science, 246,* 486-488.

- Papousek, M., Papousek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, 14, 415-440.
- Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Perception, 6*, 65–94.
- Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Heidelberg: Springer.
- Parncutt, R. (1993). Pitch properties of chords of octave-spaced tones. *Contemporary Music Review*, *9*, 35-50.
- Parncutt, R. (1994a). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11,* 409-464.
- Parncutt, R. (1994b). Template-matching models of musical pitch and rhythm perception. *Journal of New Music Research, 23*, 145-168.
- Parncutt, R. (1997). Pränatale Erfahrung und die Ursprünge der Musik. In L. Janus & S. Haibach (Eds.), *Seelisches Erleben vor und während der Geburt* (pp. 225-240). Neu-Isenburg: LinguaMed.
- Parncutt, R. (2003). Accents and expression in piano performance. In K. W.
   Niemöller (Ed.), *Perspektiven und Methoden einer Systemischen Musikwissenschaft* (pp. 163-185). Frankfurt/Main, Germany: Peter Lang.
- Peper, C. E., Beek, P. J., & van Wieringen, P. C. W. (1995). Multifrequency coordination in bimanual tapping: Asymmetrical coupling and signs of supercriticality. *Journal of Experimental Psychology: Human Perception* and Performance, 21, 1117-1138
- Popper, K. R., & Eccles, J. C. (1977). The self and its brain. Berlin: Springer
- Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*, 411-440.
- Randel, D. M. (1971). Emerging triadic tonality in the fifteenth century. *Musical Quarterly, 57,* 73–86.
- Riemann, H. (1893). Vereinfachte Harmonielehre. London: Augener.
- Sano, H., & Jenkins, B. K. (1989). A neural network model for pitch perception. *Computer Music Journal*, *13* (3), 41-48.
- Schreiner, C. E., & Langner, G. (1997). Laminar fine structure of frequency organization in auditory midbrain. *Nature, 388,* 383-386.

- Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35A*, 377–396.
- Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music, 19,* 110-120.
- Temperley, D. (2001). *The cognition of basic musical structures.* Cambridge, Massachusetts: MIT Press.
- 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.
- Terhardt, E., Stoll, G., & Seewann, M. (1982). Pitch of complex tonal signals according to virtual pitch theory: Tests, examples and predictions. *Journal of the Acoustical Society of America*, *71*, 671-678.
- Terhardt, E., Stoll, G., Schermbach, R., & Parncutt, R. (1986).
  Tonhöhenmehrdeutigkeit, Tonverwandschaft und Identifikation von Sukzessivintervallen. *Acustica*, *61*, 57–66.
- Vos, P. G., & Leman, M. (Eds.) (2001). Tonality induction [Special issue]. *Music Perception, 17* (4).



**Figure 1.** Comparison of the major and minor key profiles of Krumhansl & Kessler (1982) with calculated chroma salience within the corresponding tonic triad according to Parncutt (1988). Open triangles: key profiles, i.e. experimental goodness-of-fit ratings on a 7point scale, averaged over three cadential progressions (IV-V-I, II-V-I, VI-V-I) and a single tonic triad (I). Filled squares: pitch-class weight according to Parncutt (1988) with root-support weights P1/P8 = 10, P5 = 5, M3 = 3, m7 = 2, M2/M9 = 1, m3 = 0, divided by 3 for ease of comparison.