Computational Models of Music Perception and Cognition II: Domain-Specific Music Processing

Hendrik Purwins[1], Maarten Grachten[1,2], Perfecto Herrera[1], Amaury Hazan[1], Ricard Marxer[1], and Xavier Serra[1]

[1]Music Technology Group
Universitat Pompeu Fabra, Barcelona

[2] Department of Computational Perception
Johannes Kepler Universität, Linz

Address of Correspoding Author:
Universitat Pompeu Fabra
Institut Universitari de Audiovisual
Music Technology Group
Ocata 1
08003 Barcelona, Spain
Tel: 0034-93 54 21 365
Email: hpurwins@iua.upf.es

PACS: 43.75.Cd Music perception and cognition
43.75.St Musical performance, training, and analysis
43.75.Xz Automatic music recognition, classification, and information retrieval
43.75.Zz Analysis, synthesis, and processing of musical sounds

Keywords: music cognition, auditory system, music perception, musical expectancy

Contents

Abstract

In Part I, we addressed the study of cognitive processes that underlie auditory perception of music, and their neural correlates. The aim of the present paper is to summarize empirical findings from music cognition research that are relevant to three prominent music theoretic domains: rhythm, melody, and tonality. Attention is paid to how cognitive processes like category formation, stimulus grouping, and expectation can account for the music theoretic key concepts in these domains, such as *beat*, *meter*, *voice*, *consonance*. We give an overview of computational models that have been proposed in the literature for a variety of music processing tasks related to rhythm, melody, and tonality. Although the present state-of-the-art in computational modeling of music cognition definitely provides valuable resources for testing specific hypotheses and theories, we observe the need for models that integrate the various aspects of music perception and cognition into a single framework. Such models should be able to account for aspects that until now have only rarely been addressed in computational models of music cognition, like the active nature of perception and the development of cognitive capacities from infancy to adulthood.

1 Introduction

Music as a socio-cultural phenomenon covers a very wide variety of human activities. Even with narrower interpretations, music is inhomogeneous, as forms of musical expression change constantly, both with time and with geographic location. Music theoretic notions are bound to be incomplete, or relevant to only a fraction of what could be validly called music. For example a commonly heard critique of musicology is that its central terminology has its origins in a very specific subset of music (namely 19th century Western European music), thereby neglecting large areas of music (for example popular music) that can not be adequately described with this terminology, or discarding it as degenerate [Middleton, 1990].

A related problem is the tension between music theorists and cognitive psychologists with respect to the way insights are gathered about musical notions and corresponding cognitive notions. One problem is for example that the notion of music perception that many music theorists seem to have, as being a largely conscious, voluntary process, is not at a par with the notion of perception as unconscious, and reflexive, which is current in cognitive psychology [Cross, 1998]. Furthermore, there are disparate views on methodological issues between music theory on the one hand and cognitive psychology on the other. Whereas music theorists have criticized studies of music by cognitive psychologists as focused on too low level, perceptual processes, using non-realistic stimuli [Honing, 2006], cognitive psychologists have discarded cognitive claims by music theorists as unwarranted by significant evidence, and too much lead by apriori conceptualization [Gjerdingen, 1999].

Even if such a tension exists, many studies in the last few decades have attempted to bridge the gap between traditional musicology and cognitive psychology. This has lead to the emergence of a number of branches like empirical, systematic, computational, and cognitive musicology. A common factor of these branches is that they move away from the classic notion of music as an art-form to be interpreted historically or aesthetically, towards music as a phenomenon that is to be primarily studied by measurement and modeling, as in any empirical science. This could lead to a proper grounding of music theoretic concepts in music cognition, and particularly music perception—as promoted by for example Cross [1998] and Gjerdingen [1999].

In Part I, we addressed the study of cognitive processes that underlie auditory perception of music, and their neural correlates. The aim of the present paper is to summarize empirical findings from music cognition research that are relevant to three prominent music theoretic domains: rhythm, melody, and tonality. Attention is paid to how cognitive processes like category formation, stimulus grouping, and expectation can account for the music theoretic key

concepts in these domains, such as *beat*, *meter*, *voice*, *consonance*. Issues that will be addressed are, for example, the relation between physical and perceived duration, the mental representations involved in the perception of melody, and the physiological basis for the perception of consonance.

Special attention is devoted to computational models of music perception [1]. Desain et al. [1998] point out that rather than being the definitive result of a theory of music cognition, computational models should be regarded as a starting point for comparing different hypotheses, or theories. Even if at present there are no models that integrate the various aspects of music perception and cognition into a single framework, models do exist for specific aspects of rhythm, melody, and tonality perception. In rhythm perception, a wide variety of computational models exist for *pulse finding* [Povel and Essens, 1985; Desain and Honing, 1999; Scheirer, 1998]. In melody, *expectation* has received much attention [Cuddy and Lunney, 1995; Schellenberg, 1997; Huron, 2006; Pearce and Wiggins, 2006; Temperley, 2007]. In tonality, the empirical findings on *probe tone ratings* described by Krumhansl and Kessler [1982] have lead to several computational models, like those of Huron and Parncutt [1993], and Leman [2000].

2 Rhythm

Rhythm is a musical concept with a particularly wide range of meanings, and as such it is essential to delimit the scope of what we will be talking about when discussing rhythm. An all-encompassing description of rhythm would be that it is about the perception of the temporal organization of sound. As far as we consider *musical* sound, rhythm in this sense is tightly connected with the concept of meter, which refers to the periodic structure in music. Several studies have even questioned the separation between the processes of meter and rhythm perception altogether [Hasty, 1997; Rothstein, 1981; Drake, 1998; Lerdahl and Jackendoff, 1983].

Although rhythm is ambiguous and could, as in the bottom-line definition given above, also include meter, a common use of the words rhythm and meter is expressed in a paraphrase of London [2006]: "[M]eter is how you count time, and rhythm is what you count–or what you play while you are counting". When we hear a sequence of events that can be located in time, periodicities in the timing of these events will cause in us the awareness of a *beat*, or *pulse*. By this, we mean a fixed and repeating time interval with a particular offset in time, that can be thought of as a temporal grid, forming

[1] See Part I for a discussion of cognitive validity of computational models in the context of music cognition.

a context in which the perceived events take place. The perception of events often makes some beats feel stronger, or more accented, than others, giving rise to a hierarchical grid of beats.[2] Demany et al. [1977]; Parncutt [1994] sustain the existence of a preferred tempo spectrum ranging from 60 to 120 beats per minute anchored approximately at 100 beats per minute. London [2002] provides a detailed review of cognitive constraints on beat perception.

As a context for the perception of musical events, meter serves to categorize events in terms of temporal position and accent, and inter-onset intervals (IOI) of events in terms of duration. A common interpretation of *rhythm* is the grouping of sequences of such categorized events, where groups often contain one accented event and one or more unaccented events [Cooper and Meyer, 1960]. This accounts for the influence of meter on the perception of rhythm. The perception of actual musical events on the other hand also affects the perception of meter. Firstly, in actual performances phenomenal cues like the loudness and duration of events are used to accent particular metrical positions, thereby suggesting a specific meter [Lerdahl and Jackendoff, 1983; Palmer, 1989]. In addition to this, the frequency distribution of musical events over metrical locations is mainly correlated with meter, rather than, for example, musical style or period [Palmer and Krumhansl, 1990]. This indicates that frequency distribution of events in a piece could be a perceptual clue to determining meter. Palmer and Krumhansl [1990] also show that listeners appear to have abstract knowledge of the accent structure of different kinds of meters, which become more fine-grained as a result of musical training. They suggest that the accent structures of different meters are learned by virtue of their ubiquity in Western tonal music.

A first step towards the understanding of rhythm perception is the study of perceived event durations. Differentiation of event durations happens in an early stage of auditory processing. Several factors are known to affect the perceived durations, leading to a difference between the physical duration and the perceived duration of an event. These include the phenomenon of auditory streaming, intensity/pitch difference between evenly spaced notes, and metric context. Likewise, perceived inter-onset intervals are influenced by event durations, and the different durations of successive events in their turn can affect the intensity and/or loudness perception of these [Terhardt, 1998].

Listening to two successive audio events gives rise to other cognitive processes than listening to each event separately. The judgment of successive events can be divided into two steps. The first step follows a modified version of Weber's law, in which the just-noticeable difference between two successive durations is proportional to their absolute length plus a constant of minimal

[2] Note that we use the term 'beat' both to refer to the grid, and to individual points in the grid.

discrimination [Allan, 1979]. The second step consists in comparing the two durations. If they are similar, the second duration is related to the first one. If the two durations are considered very different they will both be related to previously established perceptual categories of durations [Clarke, 1989].

The notion of a rhythmic group arises when a listener is exposed to a succession of two or more musical durations that he perceives as coherent. We can divide the process of rhythmic group perception into two interdependent steps. First the group boundaries are identified by judging the proximity and similarity of the stimuli [Meyer, 1956; Handel, 1989; Lerdahl and Jackendoff, 1983; Bregman, 1990]. Proximity in this context is the distance between the onsets, while the similarity refers to the distance between musical features like pitch, or timbre. The second step consists in recognizing and creating a representation for the internal structure of the group.

Desain and Honing [2003] present two experiments to analyze the formation of perceived discrete rhythmic categories and their boundaries. In the first experiment they study the phenomenon of rhythm categorization, which consists in the mapping from the performance space –the set of possible performed IOI durations– to a symbolic space of rhythm representation –the set of duration categories, or nominal durations. Such a mapping expresses how much the IOI's of a performed rhythmic pattern can deviate from the nominal IOI durations that define the pattern, before it is being perceived as a different rhythmic pattern. The deviation from the nominal IOI durations is called *expressive timing*. Listeners tend to simplify durational relations assigned to an expressively timed pattern of onsets. Here, high consistency in the listener's judgment is predominant if the durational ratios are simple. In the second experiment, Desain and Honing study how the metrical context affects the formation of rhythmic categories. As a context, an acoustic pattern is provided by tapping a triple, duple, and simply a bar structure. Then the stimulus is played. Given triple and duple meter contexts, the stimulus is identified as having the meter of the context stimulus. If only bars are sonified, the stimulus is as well identified as duple meter in the majority of the cases.

Dalla Bella and Peretz [2005] suggest that musical style differentiation (partly) relies on basic processes in auditory perception like rhythm perception. In this study, subjects are able to some extent to perform style differentiation and historical ordering of classical music excerpts, even when they lack familiarity with classical music.

Phillips-Silver and Trainor [2005] address the relation between body movement and rhythm perception. They show that duple/triple meter categorization in infants can be biased by the motor and proprioreceptive and vestibular (i.e. perception of body position and balance) systems.

2.1 Rhythm Models

Most of rhythm perception models address specific aspects of rhythm perception, and no comprehensive model exists. The choice of aspects to be modeled is often motivated by the available experiments, in order to allow a better understanding of the processes observed in those experiments. The processes modeled in rhythm perception are pulse finding, grouping, and internal representation. [3]

2.1.1 Pulse Finding

A relatively well studied phenomenon is that of pulse finding. Accordingly, a variety of pulse finding models have been proposed. The first type of model that appeared is based on symbolic data and applies *rule-based* algorithms. In such models a hypothesis of the pulse is proposed based on the pattern of the first few audio events. The hypothetical pulse will be shifted or scaled seeking to fit the upcoming temporal pattern [Steedman, 1977; Longuet-Higgins, 1987; Desain and Honing, 1999].

Povel and Essens [1985] introduce the concept of *inner clocks*. Inner clocks are constructed by temporal patterns and accents to provide forward chronological prediction, or internal expectancy. The period and phase of these expectancies can be adapted to correct for temporal variability. These models take quantized onset timings as input. Therefore they cannot cope with expressive pulse variation or beat irregularities that are present in real situations. In addition they do not account for the influence of audio features such as pitch, loudness, or timbre. Also the model cannot operate online. The idea of inner clocks is taken further in other works by Desain [1992]; Parncutt [1994]; Large and Kolen [1994].

In *oscillator models*, a signal processing point of view is adopted which take as input raw audio signals. An example of this is Scheirer's model [Scheirer, 1998]. In this model, the raw audio signal is separated in different sub-bands, the energy envelope of each output is then correlated with a bank of comb resonators covering the tempo space. These delay lines are phase locked to the input signals. The activation and state of the resonators allows the extraction of period and phase. This can be seen as a translation of the inner clocks model to continuous time space. As such, the problems of symbolism and noise robustness are avoided. Another branch of signal processing approaches is constituted by the models that make use of the autocorrelation, the related Discrete Fourier Transform, or its generalization, the Wavelet Transform [Smith, 1996], and similarity matrices [Foote and Uchihashi, 2001].

[3] For a thorough treatment of computational models of rhythm see Gouyon [2005].

Another framework that is being used for finding models are neural networks. An example of this is the *Graphical Model* by Lang and de Freitas [2004], which is based on Cemgil and Kappen [2003] and adapted to the continuous time domain and audio input. It also tries to address certain limitations implied by assumptions made in similar models such as the one by Goto [2001]. Lang and De Freitas approach pulse tracking by inference methods, exploiting the hypothesis of pulse periodicity and tempo continuity over time. They consider the pulse period and phase as underlying hidden states for each given frame of time. The observable variables are features extracted from the audio signal. The connection between the hidden state and the observable variable is done by considering that for a certain pulse period and phase we would have more musical events in timings that are coarser subdivisions of the beats.

The *primal sketch* pulse finding model by Todd [1994, 1999] is guided by neurophysiological experiments and related work in the visual domain, and is intended to be biologically plausible. It models the processing chain of the auditory cortex, and includes the gammatone basilar membrane model and the inner hair-cell (IHC) transduction model of Meddis and Hewitt [1991]. The summed output of the IHCs constitutes a spike train which is passed through a leaky integrator to model the time resolution of the auditory cortex. Subsequently, it is passed through a filter bank of Gaussian filters of different time scales in order to cover the space of perceived tempos. This is analogous to the notion of primal sketch in visual perception presented by Marr [1982]. The output of the filter bank is used to extract information about onset times and prominence. It gives rise to a *rhythmogram* which can be interpreted in the sense of the hierarchical grouping model by Lerdahl and Jackendoff [1983].

In robotics, Brooks [1991] uses the notion of *embodiment* to characterize agents that are endowed with an active perceptual process that senses e.g. gravity. Based on perceptual information from the environment, these agents react directly. Eck et al. [2000] build a pulse finding model as the interaction between perception, implemented as a network of oscillators, and a body model, simulated by a mass-spring system. Todd et al. [2002] emphasize the importance of the body for rhythm perception. Honing [2005] proposes to combine a rhythmic categorization model with a kinematic model in order to account for global tempo, note density, and rhythmic structure. [4]

2.1.2 Rhythmic Grouping and Categorization

Compared to pulse finding, rhythmic grouping and categorization have been experimentally explored to a lesser extent, and consequently less models exist for these phenomena. Lerdahl and Jackendoff [1983] suggest a hierarchical

[4] See Leman [2008] for a broader coverage of embodiment and music.

grouping model. Experiments [Deliège, 1987] support the predictions made by Lerdahl and Jackendoff [1983] and partially support the relative strength of the grouping rules. Longuet-Higgins and Lee [1982] build a model that creates expectations based on previous onsets. If the expectations are fulfilled, expectations are created on a higher structural level, yielding a hierarchy up to a time span of 5 s. In Parncutt [1994], a metrical hierarchy is generated by superimposing the three or four most salient and mutually consonant pulse trains.

Smith [2000] presents a method for rhythmic analysis from a signal-processing paradigm. By applying the Continuous Wavelet Transform (CWT) to rhythmic signals containing information about onsets of musical events, he obtains a multi-resolution time-frequency representation of the music. Although in the musical domain such time-frequency representations are typically used for pitch analysis of sound, at longer time scales they convey rhythmic information rather than pitch information. The amplitude and phase of different frequencies of the rhythmic signal reveals rhythmic phenomena such as meter change, agogic accent, ritardando and acceleration.

Desain and Honing [1991] present an algorithm for rhythmic categorization. They create a numerical model that only considers consecutive durations. The algorithm is designed in a way that the internal representation of the relations between two successive durations converge towards simple integer ratios.

3 Melody

Most music largely consists of successions of recognizable pitches. From a music theoretic point of view, such successions of pitches are usually structured into *voices*. A voice is a temporal succession of pitches that are perceived as belonging together, forming a larger whole. Typical voice-structures are *monophony*, where the music consists of a single voice, *polyphony*, consisting of several simultaneous voices, and *homophony*, where a single voice is accompanied by a succession of simultaneous pitches, forming *chords*, or *harmonies*. The most salient voice in a piece of music is usually called the *melody* of the piece. [5]

From a cognitive perspective, melody perception concerns primarily perceptual grouping. This grouping depends on relations of proximity, similarity, and continuity between perceived events (cf. Part I). As a consequence, what we are able to perceive as a melody is determined by the nature and limitations of perception and memory. Melody perception presumes various types of

[5] Although sometimes melody is used as a synonym for voice.

grouping. One type of grouping divides simultaneous or intertwined sequences of pitches into *streams*. The division into streams is such that streams are internally coherent in terms of pitch range and the rate of events. A second type of grouping concerns the temporal structure of pitch sequences. A subsequence of pitches may form a *melodic grouping* [Snyder, 2000], if preceding and succeeding pitches are remote in terms of pitch register, or time, or if the subsequence of pitches is repeated elsewhere. Such groupings in time may occur at several time scales. At a relatively short time scale the groupings correspond to instances of the music theoretic concepts of *motifs*, or *figures*. At a slightly longer time scale, they may correspond to *phrases*. According to Snyder [2000] the phrase, which is typically four or eight bars long, is the largest unit of melodic grouping we are capable of directly perceiving, due to the limits of short term memory. Rather than being fully arbitrary, (parts of) melodies are often instantiations of *melodic schemata*, frequently recurring patterns of pitch contours. The most common melodic schemata are *axial* forms, *arch* forms, and *gap-fill* forms. Axial forms fluctuate around a central pitch, the 'axis'; Arch forms move away from and back to a particular pitch; And gap-fill forms start with a large pitch interval (the 'gap') and continue with a series of smaller intervals in the other registral direction, to fill the gap [Snyder, 2000]. The pitch contours of these schemata are illustrated in Figure 1.

The existence of melodic schemata highlights the importance of *expectancy* in melody perception. Expectancy refers to the anticipation of an event based on the probability of occurrence [Chaplin, 1985]. In recent music cognition research, the role of expectancy has gained increased attention (see for example Huron [2006]), and along with it a probability theoretic perspective on musical memory, expectancy, and the characteristics of musical stimuli [Pearce and Wiggins, 2006; Temperley, 2007]. An empirical study on melodic expectancy is reported by Schellenberg et al. [2002], in which they evaluate variants of a model of melodic expectancy, the Implication-Realization model (see Section 3.1). A notable finding is that there is an effect of age on the model variant that best explains the subject's responses. This may provide insights in the evolution of the representation of melody in humans from childhood to maturity (see also subsection 3.1.1).

Dowling [1982] suggests that several *parallel representations* of melody occur in the brain at the same time. For example, as a collection of isolated pitches, as a collection of intervals with clear relationship between them, or as a contour. The salience of a particular representation may depend on characteristics of the melody (e.g. tonal or atonal), the age of the listener, and the amount of musical training of the listener. Experiments that measure the impact of brain maturity suggest that at early age, melody is represented only in a rough form, such as overall contour, and that cognitive maturation comes with increasingly subtle internal representations of melodic form [Dowling, 1982;

Schellenberg et al., 2002]. The effect of music education in childhood seems to be an acceleration of this cognitive development in the domain of music specific capabilities. Especially the awareness of tonal centers is enhanced by musical training [Dowling, 1982].

Among the innate abilities related to melody perception we could mention a pitch encoding mechanism based on distances between pitches [McDermott and Hauser, 2005], the distinction between different melodic contours, same-different distinctions in sequences of notes and sounds [Trehub, 2001], and the discrimination between different "emotional" contents that can be encoded not by the pitches themselves but by their temporal patterning and their rhythmic aspects [Papousek and Papousek, 1991].

3.1 Melody Models

The purpose of melody models is to describe the perception of melody by the listener. As with models of rhythm perception, most proposed models only reflect a specific aspect of melody perception. Depending on the focus of the work, they model for example expectancies about the continuation of melodies, formation of melodic categories, or the influence of exposure to music on the recognition of melodies.

3.1.1 Melodic Expectancy

The most well-known model for melodic expectancy is the Implication-Realization (I-R) model [Narmour, 1990, 1992]. It describes patterns of expectancies with respect to the continuation of melody. More specifically, it expresses both the continuation *implied* given a particular melodic interval, and the extent to which this (expected) continuation is actually *realized* by the following interval. According to the I-R model, the sources of the listener's expectancies about the continuation of a melody are two-fold: innate and learned. Innate expectancies arise from the architectural and functional constraints of the bottom-up auditory pathway, which act as the biological grounding of grouping principles of organization of the auditory scene. On the other hand, learned expectancies arise as a consequence of the listener's existing musical knowledge (most of it being of a statistical nature) due to long term exposure to music, which biases the estimations made on the possible continuations of a melody. The learned expectancies form a top-down influence that may interfere with or override innate expectancies.

Some of the principles governing innate expectancy concern relations between pitch intervals and their registral directions (whether the intervals go up or down). Based on these principles, melodic archetypes, or *structures* can be

identified that either satisfy or violate the implication as predicted by the principles. Eight basic structures are shown in Figure 2 (left). The two principles that define the eight structures are firstly the principle of *registral direction* (PRD), and secondly the principle of *intervallic difference* (PID). PRD states that small pitch intervals imply an interval in the same registral direction, and large intervals imply a change in registral direction. PID states that a small interval implies a similarly-sized interval, and a large intervals implies a smaller interval. Table 1 summarizes the characteristics of each structure with respect to the principles. An additional principle concerns *closure*, the inhibition of expectancies by one or more factors like rhythmic events, meter, or harmony. Closure instills the termination of one structure, and the beginning of another. Lack of closure causes structures to share one or even two pitches, depending on the degree of closure. An example I-R analysis is shown below the score of a musical fragment in Figure 2 (right).

PRD, PID, and related principles have been shown to be statistically significant in various empirical studies [Cuddy and Lunney, 1995; Schellenberg, 1996]. Notably, a simpler version of the model [Schellenberg, 1997], including only two principles derived from the original principles, has been shown to have superior explanatory power [Schellenberg et al., 2002]. The first and most significant principle, *pitch proximity* (PPP), states that small pitch intervals are more expected than large pitch intervals. The second principle, *pitch reversal* (PPR), predicts expectancy of change of registral direction after a large pitch interval. Tests with adults, and children of different age, reveal that PPP holds from early age; it accurately predicts expectancies of children of about five years old and its accuracy increases with age. PPR on the other hand only develops relatively late (after age 11). This is explained by the fact that PPR is a second order principle (involving pitch intervals, not just the last pitch), which presumably takes more time to learn. The emergence of PPR expectancies may be due to increased perceptual differentiation, increased memory capacity, or due to exposure. However, it is difficult to determine which of these factors is actually responsible for the effect, since cognitive development is largely co-extensive with exposure to music that dominantly satisfies the PPR principle.

3.1.2 Internal Representation

Several works use a self-organizing architecture to model the emergence of mental concepts during learning and exposure to particular stimuli. [6] For instance, Page [1999] applies ART-2 [Carpenter and Grossberg, 1987] networks to the perception of musical sequences, and adopts an extension of ART-2 called SONNET [Nigrin, 1990]. Piat [2000] presents a model based on a

[6] Cf. Grossberg [1976]; Kohonen [1982]; Fisher [1987], summarized in Part I.

simple ART network that can categorize a melody containing distractors. In these studies, the trained networks provide a parsimonious account of empirical findings about perceived tones, chords, or key relationships, based on mere exposure to symbolic musical material. Marxer et al. [2007] use COBWEB/3 [McKusick and Langley, 1991] to model the gradual emergence of motivic categories during exposure to Bach's *Inventio no IV*. The partition is then evaluated and compared to the results of musicology studies [Lartillot and Toivianen, 2007].

4 Tonality

In Western music of the major-minor tonality period, roughly from Bach's birth (1685) to Wagner's *Tristan* (1857), the tonality of a piece of music is defined by the kinship of *tone centers* (local keys during a short time frame). A tone center can be characterized by a couple of features: 1) the collection of employed pitch classes defined by the *scale*, 2) the *key note* (the first note in the scale, often the most frequent note that may also be used as the beginning and final note), and 3) what *chords* are used how. The tone centers are interrelated via kinships that may be labeled as dominant, relative, and parallel. The tonic and dominant scale differ only by one tone (introduced by an additional accidental, a ♭ or a ♯), but start with key notes one fifth apart. Major and parallel minor scale start with the same key note but differ especially in having a major or minor third (counting from the key note). Major and relative minor scale, e.g. C-major and aeolic a-minor, use the same set of tones but start a small third apart.

Tonality has been extensively studied by musicologists and composers [Heinichen, 1728; Rameau, 1722; Riemann, 1877]. Of primary importance is the influence (especially in the Anglo-Saxon research community) of Schenker [1935], whose theory of harmony is ultimately absorbed into a thorough account of tonal structure. Since the mid-20th century theoretical approaches to tonality have developed in several directions [Krumhansl and Shepard, 1979; de la Motte, 1980; Werts, 1983; Noll, 1995].

There has been little work on tonality in psychoacoustics, as opposed to musicology, probably because of the many assumptions that have to be established. A special aspect is the relation between harmony and *consonance* (which is covered in-depth elsewhere, e.g. in Palisca and Moore [2008]). According to Helmholtz [1863], the degree of consonance of a pitch interval between two tones is based on the amount of coincidence between the partials of the two tones. Helmholtz argues for grounding consonance in physiology. Studies from Tramo et al. [2001] indicate that "there is a high correlation between tonal dissonance of musical intervals and the total number of auditory nerve fibers

that show beating patterns." These results may explain that "at least one aspect of consonance/dissonance is determined at the most peripheral level of the central auditory system"[Tramo et al., 2001]. However, the question remains whether the distinction of intervals into consonant and dissonant relations reflects a physiologically grounded constraint (as argued in Krumhansl [2000]) or if it is acquired through experience. A physiology-based approach to harmony is challenged by the treatment of the fourth, a simple 3:4 relation of frequencies, as a dissonance during important periods of Western music history. The consideration of the fourth as a dissonance is attributed to voice leading by Mazzola [1990].

Probe tone ratings [Krumhansl and Shepard, 1979] provide a quantitative description of key that creates the possibility of relating statistical or computational analysis of music (as a recording or a score) to cognitive psychology. Probe tone experiments consist of two stages: establishing a tonal context, and rating how well a new tone (the probe tone) fits with that context. The 12-dimensional vector containing the averaged answers for each probe tone is called the probe tone rating. There are two types of rating vectors, one for major and another one for minor mode. According to an observation reported in Krumhansl [1990] (p. 66–76), each component in the probe tone vector is related to the frequency and the overall duration of occurrence of the corresponding pitch class at metrically prominent positions of a piece written in a given key.

Purwins et al. [2000] developed a method (Constant-Q profiles) to accumulate pitch classes from an audio excerpt. As underlying DSP method serves the constant Q transform [Brown, 1991; Brown and Puckette, 1992] that is applied frame-wise and results in a frequency representation with equal resolution in the *logarithmic* frequency domain. To concentrate the so found information in a 12-dimensional vector (CQ-profile) all the values that belong to one pitch class are summed up across all octaves of the constant Q transform. The CQ-profiles can be added across all frames of a piece resulting in a vector that contains information about the tonal content of that piece since it is closely related to Krumhansl's probe tone ratings. We have to consider that the CQ-profile not only extracts the fundamental frequency of a tone but all its harmonics. For piano tones, the strongest frequency contribution falls (modulo octave) on the fundamental and its fifth in an approximate average ratio 3:1. Hence CQ-profiles should not be compared with the probe tone ratings, but with adapted ratings in which the harmonic spectrum of the analyzed tones is accounted for. Such a spectrally weighted rating is calculated by adding to the rating value for each tone one third of the rating value for the tone one fifth above (modulo octave). Figure 3 shows the highly consistent correlation of the average CQ-profiles of sampled piano cadences with the spectrally weighted probe tone ratings (cf. Figure 3, Purwins et al. [2000]).

Probe tone ratings derived from listeners with strong musical background differ significantly from a group with very little education (Krumhansl [1990], p. 22–24), even if there is no general agreement in experimental studies on how to differentiate exactly between experienced and inexperienced subject's ratings of tonality [Vos, 2000]. Subjects with strong musical background give high ratings to the major or minor scale notes, especially the fifth. It is possible that subjects with strong musical background partly recognize intervals, apply their musical knowledge, and judge the probe based on their explicit knowledge of the prominence of the scale tone [Purwins et al., 2000]. On the other hand, the ratings of these notes given by naive listeners are inversely related to the distance of the note to the first note in the scale measured in half tone steps modulo octave. However, another study [Bigand, 2003] emphasizes how naive listeners assimilate chord structures and tonal knowledge by "passive exposure" to samples of Western music [Bigand, 2003]. This effect has been even observed in five year old children [Koelsch et al., 2003]. In spite of differences in tonality judgments between groups of listeners, the probe tone rating method seems to be a general concept for the description of tonal structures, applicable to non-Western music as well [Castellano et al., 1984].

4.1 Tonality Models

Leman [1995] implements the perception of inter-key relations as a chain of auditory images[7] and interaction with a self-organization or visualization schema. Following the hair cell auditory image, Leman uses the autocorrelation and tone context image, and the semantic image. A semantic image assumes characteristic groups of neurons that respond most strongly to particular keys (Leman [1995], p. 18-19). Cognitive models have also been built based on ART (Piat [2000], Section 3.1.2), and on Constant-Q profiles (cf. Section 4) in conjunction with Correspondence Analysis or alternatively Isomap [Purwins et al., 2004; Purwins, 2005]. However, all of these models lack biological plausibility [Purwins, 2005].

In Figure 4, Correspondence Analysis visualizes co-occurrence data. Pitch classes are described by a vector that contains the frequencies of how often pitch classes occur in one of the 24 major/minor keys (pieces), e.g. in the fugues of Bach's Well-Tempered Clavier, Part I (Figure 4). Correspondence Analysis aims at embedding the pitches in a lower-dimensional space such that the spatial relations in that space display the similarity of the features $\mathcal{P}$, expressed as frequencies of occurrence in the 24 keys $\mathcal{K}$. Correspondence Analysis

[7] An *auditory image* is a vector "where each number represents the probability of neural firing of a group of neurons during a short time interval." (Leman [1995], p. 38)

preserves the χ^2 distance (a generalization of the Euclidean distance) between data points. As a generalization of Principal Component Analysis, Correspondence Analysis decomposes the data matrix into two sets of factors. The vector of key frequencies can be rotated yielding a new co-ordinate system, given by a set of factors. Each factor is associated with a singular value, quantifying the amount of variance explained by that factor. In Figure 4, Correspondence Analysis can be considered as a rough metaphor of how high-dimensional data can be represented on a two-dimensional layer of the cortex, preserving a high amount of relevant information. We see how the circle of fifth, a key concept of Western music, has emerged [Purwins et al., 2004].

There are some studies on probe tone ratings that investigate the influence of sensory and short-term memory on key perception. Huron and Parncutt [1993] show that an echoic memory model, based on Terhardt [1979]'s pitch model, explains some of the data reported in Krumhansl and Kessler's study. Following this idea, Leman [2000] describes an auditory model including a short-term memory model that gives an explanation to Krumhansl and Kessler [1982]'s data. Also the neural network model by Bharucha [1999] takes into account short-term memory. According to Deutsch [1999], the extent to which probe tone ratings are driven by short-term memory, long-term memory, or yet other factors remains unresolved.

4.1.1 Learning Sequential Structure

Several error-driven recurrent neural network approaches exist to model the learning of melodies and more abstract musical tasks such as learning the structure of major and minor scales. Mozer [1994] presents a system called CONCERT that is able to solve a variety of musical tasks. Analogously to the work of Elman [1990], where neural networks discover structure in English language material, the system can produce internal representations of chords. Eck and Schmidhuber [2002] and Franklin and Locke [2005] follow a similar approach. Serrà [2007] has developed a tonal descriptor system that can recognize cover versions of an original musical piece.

5 Conclusion

In this review we have presented empirical findings and computational models in the field of music cognition, and in particular the domains of rhythm, melody, and tonality. A few concluding remarks can be made.

Firstly, various cognitive processes in different domains can be seen as forms of *self organization* and *dimensionality reduction*. Examples of the former are

most notably the perception of inter-key relations [Leman, 1995], and the emergence of motivic categories [Marxer et al., 2007]. The latter is exemplified in processes like the quantization of rhythmic durations to simple integer ratios [Desain and Honing, 1991], and the existence of melodic schemata [Snyder, 2000].

Secondly, *expectation* plays an important role in various aspects of music cognition for example in pulse finding, key finding, and in melody perception. An appealing computational framework for this is probabilistic (Bayesian) reasoning [Temperley, 2007; Pearce and Wiggins, 2006], but other methods have also been applied to model expectation [Longuet-Higgins and Lee, 1982; Povel and Essens, 1985; Narmour, 1990].

Furthermore, in melody perception and memory some studies have addressed the question of how internal representations of melody (and related concepts) evolve with age from infancy to adulthood [Dowling, 1982; Schellenberg et al., 2002; Salselas, 2007]. In general however, the evolution of cognitive music processing capacities in human development has rarely been studied. A related issue is the effect of musical training on music processing capacities. Although some studies detect an effect of musical training, for example in meter perception [Palmer and Krumhansl, 1990], and in the awareness of tonal centers [Dowling, 1982; Krumhansl, 1990], Vos [2000] reports that there is no general agreement on the effect of musical training on the perception of tonality induction.

Another general conclusion that can be drawn from the literature reviewed in this paper is that general computational frameworks for music processing hardly exist. The current state of the art is rather that many partial models exist for isolated tasks, that have been evaluated with isolated empirical data. On the one hand this is hardly surprising, given the complexity and broadness of music cognition as a research field, and given its recent emergence. On the other hand, this is not a very satisfying situation, as it does not foster the creation of encompassing theories. Such theories are desirable, since many phenomena in music cognition (and indeed in cognition as a whole) cannot be fully explained when observed in isolation, without the context in which they appear. Instead, we contend that a holistic approach will provide more insight into music cognition in all its aspects. Such an approach starts from the recognition of the inherent embodiment of cognition: rather than being a static feature, cognitive behavior is the result of a developmental process of continuous interaction of an agent with its environment. In this view, perception is not a passive act of acquiring information, but an active process that is an essential element of cognition. Theories that are based on such premises are likely to take advantage of the general advances in fields like computational neuroscience and cognitive psychology.

Adequate models of music cognition are not just of intrinsic value. They will also pave the way for understanding and solving applied problems such as: What are the significant similarities between musical excerpts? (music information retrieval) How do musical intelligence and competence develop? (music education) How can we predict the emotional and cognitive reaction of a listener? (emotion engineering) Furthermore, models of musical categories can lead to the design of adaptive interfaces for music making. Expectation can be exploited for intelligent machine composition and improvisation explicitly controlling points of surprise. Context-driven expectation may improve accuracy of low level descriptors compensating noise or uncertainty. Finally, the consideration of developmental issues can encourage the design of user-adaptive applications.

Acknowledgments

Special thanks to Hugo Solis for contributing to the sections on melody in this article. We are grateful to Fabien Gouyon for useful comments. We would like to thank Graham Coleman for proof reading the manuscript. This work is funded by EU Open FET IST-FP6-013123 (EmCAP) and the Spanish TIC project ProSeMus (TIN2006-14932-C02-01). The first author also received support from a Juan de la Cierva scholarship from the Spanish Ministry of Education and Science. The second author is funded by the Austrian National Science Fund, FWF (project: P19349-N15).

References

Allan, L., 1979. The perception of time. Perception and Psychophysics 26, 340–354.

Bharucha, J. J., 1999. Neural nets, temporal composites and tonality. In: Deutsch, D. (Ed.), The psychology of music, 2nd Edition. Series in Cognition and Perception. Academic Press, pp. 413–440.

Bigand, E., 2003. More about the musical expertise of musically untrained listeners. Ann. NY Acad. Sci. 999, 304–312.

Bregman, A. S., 1990. Auditory Scene Analysis. MIT Press, Cambridge, MA.

Brooks, R. A., 1991. New approaches to robotics. Science 253, 1227–1232.

Brown, J., 1991. Calculation of a constant Q spectral transform. J. of the Acoustical Soc. of America 89 (1), 425–434.

Brown, J. C., Puckette, M. S., 1992. An efficient algorithm for the calculation of a constant Q transform. J. of the Acoustical Soc. of America 92 (5), 2698–2701.

Carpenter, G. A., Grossberg, S., 1987. Stable self-organization of pattern recognition codes for analog input patterns. Applied Optics 26, 4919–4930.
Castellano, M. A., Bharucha, J. J., Krumhansl, C. L., 1984. Tonal hierarchies in the music of North India. J. of Experimental Psychology 113 (3), 394–412.
Cemgil, A. T., Kappen, H. J., 2003. Monte Carlo methods for tempo tracking and rhythm quantization. J. of Artificial Intelligence Research 18 (1), 45–81.
Chaplin, J., 1985. Dictionary of Psychology, 2nd Edition. Laurel, New York.
Clarke, E., 1989. The perception of expressive timing in music. Psychological Research 51, 2–9.
Cooper, G., Meyer, L., 1960. The rhythmic structure of music. University of Chicago Press, Chicago.
Cross, I., 1998. Music analysis and music perception. Music Analysis 17 (1).
Cuddy, L. L., Lunney, C. A., 1995. Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. Perception & Psychophysics 57, 451–462.
Dalla Bella, S., Peretz, I., 2005. Differentiation of classical music requires little learning but rhythm. Cognition 96 (2), 65–78.
de la Motte, D., 1980. Harmonielehre, 3rd Edition. Bärenreiter, Basel.
Deliège, I., 1987. Le parallélisme, support d'une analyse auditive de la musique: Vers un modèle des parcours cognitifs de l'information musicale. Analyse musical 6, 73–79.
Demany, L., McKenzie, B., Vurpillot, E., 1977. Rhythm perception in early infancy. Nature 266, 718–719.
Desain, P., 1992. A (de)composable theory of rhythm perception. Music Perception 9 (4), 439–454.
Desain, P., Honing, H., 1991. The quantization of musical time: A connectionist approach. In: Todd, P. M., Loy, D. G. (Eds.), Music and Connectionism. MIT Press, Cambridge, pp. 150–160.
Desain, P., Honing, H., 1999. Computational models of beat induction: The rule-based approach. J. of New Research Music 28, 29–42.
Desain, P., Honing, H., 2003. The formation of rhythmic categories and metric priming. Perception 32, 341–365.
Desain, P., Honing, H., Vanthienen, H., Windsor, W., 1998. Computational modeling of music cognition: Problem or solution? Music Perception 16 (1), 151–166.
Deutsch, D., 1999. The processing of pitch combinations. In: Deutsch, D. (Ed.), The psychology of music, 2nd Edition. Series in Cognition and Perception. Academic Press, pp. 349–411.
Dowling, W. J., 1982. The Psychology of Music. Academic Press, Ch. Melodic Information Processing and Its Development, pp. 413–429.
Drake, C., 1998. Psychological processes involved in the temporal organization of complex auditory sequences: universal and acquired processes. Music Percept. 16, 11–26.
Eck, D., Gasser, M., Port, R., 2000. Rhythm Perception and Production. Swets and Zeitlinger, Ch. Dynamics and embodiment in beat induction, pp. 157–

170.
Eck, D., Schmidhuber, J., 2002. Finding temporal structure in music: blues improvisation with LSTM recurrent networks. Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on, 747–756.
Elman, J. L., 1990. Finding structure in time. Cognitive Science 14 (2), 179–211.
Fisher, D. H., 1987. Knowledge acquisition via incremental conceptual clustering. Mach. Learn. 2 (2), 139–172.
Foote, J., Uchihashi, S., 2001. The beat spectrum: A new approach to rhythm analysis. In: Proceedings of the International Conference on Multimedia and Expo (ICME). pp. 881–884.
Franklin, J. A., Locke, K. K., 2005. Recurrent neural networks for musical pitch memory and classification. International J. on Artificial Intelligence Tools 14 (1-2), 329–342.
Gjerdingen, R. O., 1999. An experimental music theory? In: Cook, N., Everest, M. (Eds.), Rethinking Music. Vol. 2. Oxford University Press, Oxford.
Goto, M., 2001. An audio-based real-time beat tracking system for music with or without drum-sounds. J. of New Music Research 30 (2), 159–171.
Gouyon, F., 2005. A computational approach to rhythm description. Ph.D. thesis, Universitat Pompeu Fabra.
Grossberg, S., 1976. Adaptive pattern classification and universal recoding: I. parallel development and coding of neural feature detectors. Biological Cybernetics 23, 121–134.
Handel, S., 1989. Listening: An Introduction to the Perception of Auditory Events. MIT Press.
Hasty, C., 1997. Meter as Rhythm. Oxford University Press.
Heinichen, J. D., 1728. Der General-Baß in der Composition. Dresden, reprint 1969, Georg Olms Verlag, Hildesheim.
Helmholtz, H. v., 1863. Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. Vieweg, Braunschweig.
Honing, H., 2005. Is there a perception-based alternative to kinematic models of tempo rubato? Music Perception 23 (1), 79–85.
Honing, H., 2006. On the growing role of observation, formalization and experimental method in musicology. Empirical Musicology Review 1 (1).
Huron, D., 2006. Sweet Anticipation. The MIT Press, Cambridge, MA.
Huron, D., Parncutt, R., 1993. An improved model of tonality perception incorporating pitch salience and echoic memory. Psychomusicology 12 (2), 154–171.
Koelsch, S., Grossmann, T., Gunter, T., Hahne, A., Schroger, E., Friederici, A., 2003. Children processing music: electric brain responses reveal musical competence and gender differences. J. Cogn. Neurosci. 15, 683.
Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybernetics 43, 59–69.
Krumhansl, C., 1990. Cognitive Foundations of Musical Pitch. Oxford University Press, Oxford.

Krumhansl, C., 2000. Rhythm and pitch in music cognition. Psychol. Bull. 126 (1), 159–179.

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.

Krumhansl, C. L., Shepard, R. N., 1979. Quantification of the hierarchy of tonal function with a diatonic context. J. of Experimental Psychology: Human Perception and Performance.

Lang, D., de Freitas, N., 2004. Beat tracking the graphical model way. In: Advances in Neural Information Processing Systems (NIPS). Vol. 18.

Large, E. W., Kolen, J. F., 1994. Resonance and musical meter. Connection Science 6 (2, 3), 177–208.

Lartillot, O., Toivianen, P., 2007. Motivic matching strategies for automated pattern extraction. Musicae Scientiae 4A, 281–314.

Leman, M., 1995. Music and Schema Theory. Vol. 31 of Springer Series in Information Sciences. Springer, Berlin, New York, Tokyo.

Leman, M., 2000. An auditory model of the role of short term memory in probe-tone ratings. Music Perception, Special Issue in Tonality Induction 17 (4), 481–509.

Leman, M., 2008. Embodied Music Cognition and Mediation Technology. MIT Press, Cambridge, MA.

Lerdahl, F., Jackendoff, R. A., 1983. A Generative Theory of Tonal Music. MIT Press, Cambridge, MA.

London, J., 2002. Cognitive constraints on metric systems: Some observations and hypotheses. Music Perception 19 (4), 529–550.

London, J., 2006. How to talk about musical meter. Colloquium talks, Carleton College, MN.

Longuet-Higgins, H. C., 1987. Mental Processes. MIT Press.

Longuet-Higgins, H. C., Lee, C., 1982. The perception of musical rhythms. Perception 11, 115–128.

Marr, D., 1982. Vision. Freeman, New York.

Marxer, R., Holonowicz, P., Purwins, H., Hazan, A., 2007. Dynamical hierarchical self-organization of harmonic, motivic, and pitch categories. In: NIPS Music, Brain and Cognition Workshop. Vancouver, Canada.

Mazzola, G., 1990. Geometrie der Töne. Birkhäuser Verlag, Basel.

McDermott, J., Hauser, M., 2005. The origins of music: Innateness, uniqueness, and evolution. Music Perception 23 (1), 29–59.

McKusick, K. B., Langley, P., 1991. Constraints on tree structure in concept formation. In: IJCAI. pp. 810–816.

Meddis, R., Hewitt, M. J., June 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. J. of the Acoustical Soc. of America 89 (6), 2866–2882.

Meyer, L., 1956. Emotion and Meaning in Music. University of Chicago Press, Chicago, IL.

Middleton, R., 1990. Studying Popular Music. Open University Press,

Philadelphia.
Mozer, M. C., 1994. Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multi-scale processing. Connection Science 6 (2 & 3), 247–280.
Narmour, E., 1990. The analysis and cognition of basic melodic structures: the Implication-Realization model. University of Chicago Press.
Narmour, E., 1992. The analysis and cognition of melodic complexity: the Implication-Realization model. University of Chicago Press.
Nigrin, A. L., 1990. Sonnet: A self-organizing neural network that classifies multiple patterns simultaneously. In: IEEE International Joint Conference on Neural Networks (IJCNN). Vol. II. IEEE, San Diego, pp. 313–318.
Noll, T., 1995. Fractal depth structure of tonal harmony. In: Proc. of the Int. Computer Music Conf. ICMA, Banff.
Page, M. P. A., 1999. Modelling the perception of musical sequences with self-organizing neural networks. In: Musical networks. MIT Press, Cambridge, MA, USA, pp. 175–198.
Palisca, C. V., Moore, B. C. J., 2008. Consonance. In: Macy, L. (Ed.), Grove Music Online. Accessed 31 Jan. http://www.grovemusic.com.
Palmer, C., 1989. Mapping musical thought to musical performance. J. of Experimental Psychology - Human Perception and Performance 15 (12), 331–346.
Palmer, C., Krumhansl, C. L., 1990. Mental representations for musical meter. J. of Experimental Psychology - Human Perception and Performance 16 (4), 728–741.
Papousek, M., Papousek, H., 1991. The meaning of melodies in motherese in tone and stress languages. Infant Behaviour and Development 14, 415–440.
Parncutt, R., 1994. A perceptual model of pulse salience and metrical accent in musical rhythms. Music Perception 11, 409–464.
Pearce, M. T., Wiggins, G. A., 2006. Expectation in melody: The influence of context and learning. Music Perception 23 (5), 377–405.
Phillips-Silver, J., Trainor, L. J., 2005. Feeling the Beat: Movement Influences Infant Rhythm Perception. Science 308 (5727), 1430–.
Piat, F., 2000. Artist a connectionist model of musical acculturation. In: International Conference on Music Perception and Cognition. Keele University, UK.
Povel, D.-J., Essens, P., 1985. Perception of temporal patterns. Music Perception 2, 411–440.
Purwins, H., 2005. Profiles of pitch classes - circularity of relative pitch and key: Experiments, models, computational music analysis, and perspectives. Ph.D. thesis, Berlin University of Technology.
Purwins, H., Blankertz, B., Obermayer, K., 2000. A new method for tracking modulations in tonal music in audio data format. In: Amari, S.-I., Giles, C. L., Gori, M., Piuri, V. (Eds.), Int. Joint Conf. on Neural Networks (IJCNN-00). Vol. 6. IEEE Computer Society, pp. 270–275.
Purwins, H., Graepel, T., Blankertz, B., Obermayer, K., 2004. Correspon-

dence analysis for visualizing interplay of pitch class, key, and composer. In: Mazzola, G., Noll, T., Luis-Puebla, E. (Eds.), Perspectives in Mathematical and Computational Music Theory. Osnabrück Series on Music and Computation. Electronic Publishing Osnabrück, pp. 432–454.

Rameau, J. P., 1722. Traité de l'harmonie réduite à ses principes naturels. Ballard, Paris.

Riemann, H., 1877. Musikalische Syntaxis. Breitkopf und Härtel, Leipzig.

Rothstein, W., 1981. Rhythm and the theory of structural levels. Ph.D. thesis, Yale University.

Salselas, I., 2007. The development of melodic representations at early age: Facts towards a computational model. Masters Thesis, Universitat Pompeu Fabra, Barcelona.

Scheirer, E. D., 1998. Tempo and beat analysis of acoustic musical signals. J. of the Acoustical Soc. of America 103 (1), 588–601.

Schellenberg, E., Adachi, M., Purdy, K., McKinnon, M., 2002. Expectancy in melody: Tests of children and adults. J. of Experimental Psychology: General 131, 511–537.

Schellenberg, E. G., 1996. Expectancy in melody: Tests of the implication-realization model. Cognition 58, 75–125.

Schellenberg, E. G., 1997. Simplifying the Implication-Realization model of melodic expectancy. Music Perception 14 (3), 295–318.

Schenker, H., 1935. Der freie Satz. Vol. 3 of Neue musikalische Theorien und Phantasien. Universal, Wien.

Serrà, J., 2007. A qualitative assessment of measures for the evaluation of a cover song identification system. In: International Conference on Music Information Retrieval (ISMIR). Vienna, Austria.

Smith, L., 1996. Modelling rhythm perception by continous time-frequency analysis. In: Proceedings of the International Computer Music Conference. pp. 392–395.

Smith, L., 2000. A multiresolution time-frequency analysis and interpretation of musical rhythm. Ph.D. thesis, University of Western Australia.

Snyder, B., 2000. Music and Memory: An Introduction. The MIT Press, Cambridge, MA.

Steedman, M. J., 1977. The perception of musical rhythm and metre. Perception 6, 555–569.

Temperley, D., 2007. Music and Probability. MIT Press.

Terhardt, E., 1979. Calculating virtual pitch. Hearing Research 1, 155–182.

Terhardt, E., 1998. Akustische Kommunikation. Springer.

Todd, N., 1999. Implications of a sensory-motor theory for the representation and segregation of speech. J. of the Acoustical Soc. of America 105 (2), 1307.

Todd, N. P. M., 1994. The auditory "primal sketch": A multiscale model of rhythmic grouping. J. of New Music Research 23, 25–70.

Todd, P., Lee, C., Boyle, D., 2002. A sensory-motor theory of beat induction. Psychological Research 66 (1), 26–39.

Tramo, M. J., Cariani, P. A., Delgutte, B., Braida, L. D., 2001. Neurobiological foundations for the theory of harmony in Western tonal music. Annals of the New York Academy of Sciences 930, 92–116.

Trehub, S. E., 2001. Musical predisposition in infany. Annals of the New York Academy of Sciences 910, 1–16.

Vos, P. G., 2000. Tonality induction: theoretical problems and dilemmas. Music Perception, Special Issue in Tonality Induction 17 (4), 403–416.

Werts, D., 1983. A theory of scale references. Ph.D. thesis, Princeton.

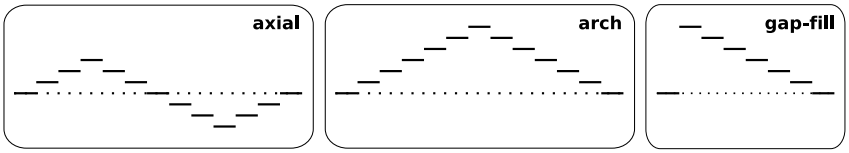


Fig. 1. Pitch contour diagrams of three melodic schemata: axial, arch, and gap-fill (from Snyder [2000]).

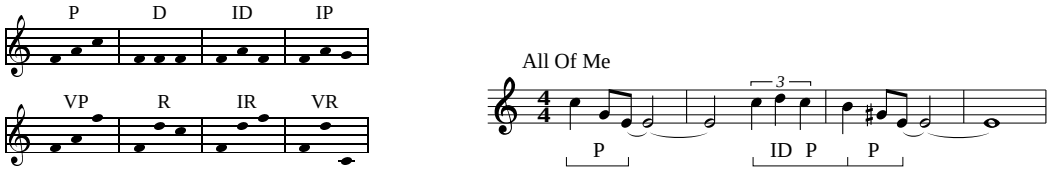
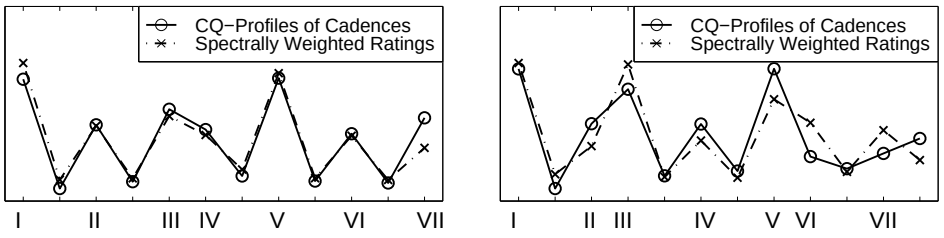


Fig. 2. Eight of the basic structures of the Implication-Realization (I/R) model (*left*). First measures of *All of Me* (The Beatles), annotated with I/R structures (*right*).

Structure	Interval sizes	Same direction?	PID satisfied?	PRD satisfied?
P	S S	yes	yes	yes
D	0 0	yes	yes	yes
ID	S S (equal)	no	yes	no
IP	S S	no	yes	no
VP	S L	yes	no	yes
R	L S	no	yes	yes
IR	L S	yes	yes	no
VR	L L	no	no	yes

Table 1
Characterization of eight basic I/R structures. PID and PRD denote the principles of intervallic difference and registral direction, respectively. In the second column,'S' denotes small, 'L' large, and '0' a prime interval. See Section 3.1 (page 13).



(a) Major ratings (b) Minor ratings

Fig. 3. High consistency between spectrally weighted probe tone ratings from a psychological experiment and calculated accumulated pitch class intensities from sampled piano cadences in audio (from Purwins et al. [2000]). The horizontal axis corresponds to the 12 pitch classes (the twelve keys on a piano keyboard within one octave). The scale degrees in major and minor are indicated in roman numerals. On the piano, the scale degrees would correspond to the white keys in C-major and a-minor. The vertical axis indicates the accumulated energy per pitch class (cq-profiles) and the prominence of the pitch classes weighted by the characteristic strength of the partials of a typical piano timbre (cf. Section 4.1).

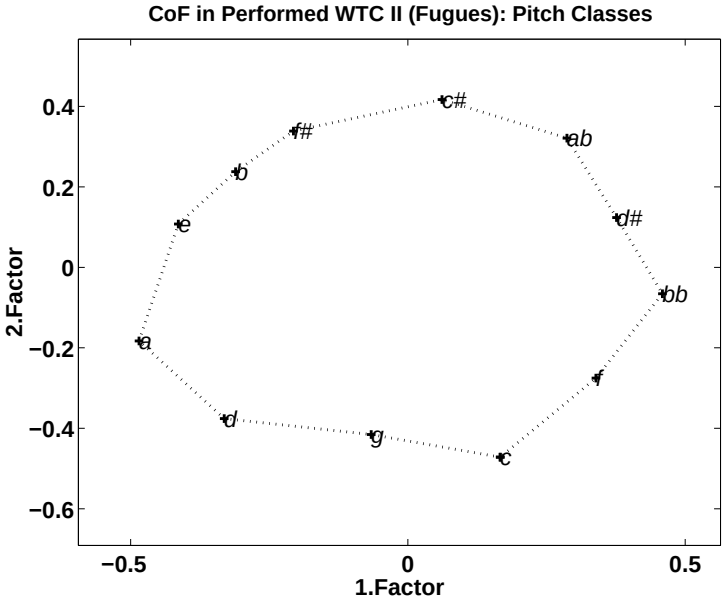


Fig. 4. The circle of fifths (lined out) appears in a recording of Bach's Well-Tempered Clavier (WTC). The analyzed data are the overall intensities of pitch classes in the fugues of Bach's WTC II in the recording of Glenn Gould. Correspondence Analysis as a metaphor of high level cognition is applied to project the 24-dimensional pitch class vectors onto a two-dimensional plane, spanned by the two most prominent factors (cf. Section 4.1). These two factors of performed WTC II capture 88.54 % of the variance of the data [Purwins et al., 2004].