Towards a bare-bone semantics for pure music

Reinhard Blutner [a]

[a] Retired member of ILLC, Universiteit van Amsterdam, Amsterdam, The Netherlands.

Abstract:
Music can have an extrinsic and/or an intrinsic meaning. The former is relevant in the case of program music, i.e. music that attempts to render an extra-musical narrative. The latter conforms to pure (absolute) music, i.e. music that can be understood without reference to extrinsic sources. Taking the intrinsic content of music as basic, we have to ask about its nature. Using a term of Immanuel Kant, I propose to identify it with aesthetic emotion. As tonal music is organized by series of chords relative to the context of a tonal scale, the question is how music forms can be mapped onto aesthetic emotions. In order to get a concise account of the affective response, this paper makes several simplifications. The most important simplification is to assume that affective responses can be represented by a two-dimensional space of emotions, where one dimension refers to surprise and the other dimension refers to pleasantness. Relating pleasantness with consonance and surprise with entropic uncertainty leads to an account which directly relates structural and probabilistic aspects of tonal music with its affective content. The present bare-bone semantics of pure music proposes an explicit modelling of the affective response based on an algebraic meaning conception.

Keywords: Music semantics, cognitive musicology, tonal attraction, aesthetic emotions, arousal–valence theory of emotions.

Inhalt

1. Introduction

The history of natural sciences is full of examples that demonstrate the role of *abstractions* and *idealizations* for the development of powerful explanatory theories. The simplest example is the free fall of bodies, where it is assumed that gravity is the only force acting upon the body (hence abstracting from air resistance) and that gravity is not changing during the fall. In this case, a very simple and systematic connection can be observed between the distance covered by an object and the elapsed time: the distance is proportional to the square of elapsed time. Interestingly, the Italian scientist Galileo Galilei (1564–1642) gave the first mathematical analysis of the law and he verified the law by observing bodies rolling down ramps (ignoring friction). Later, Isaac Newton (1643-1727) formulated the law of gravity, a force between two mass centers that is inversely proportional to the square of the distance between the two mass centers. Using this law, Newton was able to unify Kepler's laws of the movements of planets with the law of free fall. Newton also used the gravitation between moon and earth for calculating the frequency and amplitude of tidal waves. As pointed out by Stokhof and van Lambalgen (2011), a number of abstractions was necessary to calculate a definite result. For example, Newton assumed that the entire surface of the earth is covered by one ocean and that this ocean has no inertia of its own. Further, the effect of other celestial bodies, such as the planet Venus, was disregarded. Other effects that were ignored are earth rotation and Coriolis effects. Of course, neither Newton nor his followers never denied the reality of the phenomena that did not fit into the model.

In the development of cognitive science, analogies with techniques in theoretical physics are of particular relevance. In this connection, I refer to a remark by Paul Smolensky which was made during his contribution to the David Rumelhart Celebration held at Carnegie-Mellon University in 1999:

> Simple, mathematical ideas can be spectacularly successful even in cognitive science: As a recent refugee from physics, I early on found in Dave a crucial role model, much of his work demonstrating that the fundamental research method I had come to know and love in my physics training — mathematical-analysis-by audacious-oversimplification — could yield tremendous results even in a field as 'soft', and as remote from physics, as cognitive science. (Smolensky, 2000)

This paper investigates the relation between music and affective feelings and develops a bare-bone semantics for *pure* or *absolute* music. With these terms, I refer to tonal music that can be interpreted without reference to accompanying non-musical components. The main argument for being concentrated on pure music is that our understanding and evaluation of pure music will play a role in the understanding and evaluation of all kinds of 'impure' music as well, and so it cannot be ignored. I claim that the methodology of "analysis by audacious oversimplification" is appropriate for developing music semantics in case of pure music.

What we can learn from this historical excurse given at the beginning is threefold. First, one aim of science is the construction of *systematic* explanations. Scientific knowledge is primarily distinguished from other forms of knowledge, especially everyday knowledge, by being more systematic. In order to achieve the kind of systematicity that is illustrated by the three examples at the beginning it needs appropriate abstractions and idealizations. This leads me to the second aspect, the abstraction/idealization problem. The important question is what kinds of abstractions and idealizations are required in order to get a scientific approach that is sufficiently systematic. Hereby, the required level of systematicity regards either an existing research community or a potentially new research community that still has to be initiated. Third, *the proper conceptualization* of the domain under discussion is required. Which concepts are useful for describing the observed phenomena and for finding systematicity and universal laws? In case of tidal waves, what kinds of tidal forces are important for the explaining the dominant tidal constituents? In case of music, what types of musical forces should be involved? In connection with proper conceptualizations, the distinction between scientific concepts and folk concepts becomes obvious.

Remarkably, there are authors who fully doubt the investigation of music and emotion as a proper research goal. Famously, Eduard Hanslick doubted any real correspondence between a musical composition and the feelings it elicits. This clearly goes against the systematicity of emotive effects of tonal music:

> Definite feelings and emotions are unsuspectible of being embodied in music. … The physiological process by which the *sensations* of sound is converted into a *feeling*, a *state of mind*, is unexplained and will ever remain so. … Let us never appeal to a science for explanations which it cannot possibly give. (Hanslick, 1891)

Pinker (1997) is another author who plays down the intimate connection between music and emotion. He considered music as a harmless leisure activity and dismissed it as “mere auditory cheesecake". The legendary German philosopher Immanuel Kant defined *aesthetic emotions* ("ästhetische Gefühle") as an emotional response to music and other arts – produced when thought and emotion come together to create meaning. Unfortunately, this idea led Kant (1790) to rank music in the “lowest place” amongst the cultural values of arts. Christian Friedrich Michaelis (1770-1834), a German Private Docent in Leipzig and great music enthusiast interprets Kant's idea of "aesthetic emotions" as the key for understanding the effects of music. His ideas go clearly beyond Kant's own ideas and elucidate many issues in connection with the composition of tones, its aesthetic effects, and the role of mathematics in music (Mohr, 2007). For Michaelis (1795), the art of tones can be determined by its elicited effects and by his nature.

> Ich habe daher folgende Erklärungen der Tonkunst versucht. Sie ist die Kunst, durch mannichfaltige Verbindung der Töne das Gefühl zu rühren, die Fantasie zu beleben, und das Gemüth zu Ideen des Schönen und Erhabenen zu stimmen: oder die Kunst, durch verbundene Töne unmittelbar ästhetische Gefühle, und mittelbar ästhetische Ideen zu erregen. Sie besteht in modificjrter Darstellung der hörbaren Natur, dem Gesetz der vereinigten Mannichfaltigkeit gemäß in Form und Stoff bestimmt. (Michaelis, 1795, p. 39)

Hence, for Michaelis music can elicit aesthetic *feelings* in a direct way (and elicit aesthetic *ideas* indirectly).

Most contemporary authors see deep and profound connections between music and emotion and consider the issue as a proper research topic (for a recent view supporting this opinion, see Crickmore, 2017). This is true even when many authors do not share the present methodology of "audacious oversimplification" but rather find it important to sketch a "complete" picture of the relation between music and emotions. This is true for general models such as the BREVCEM approach (e.g., Patrik N. Juslin & Sloboda, 2013) or the component process model of emotion (Scherer & Coutinho, 2013). I will later use the BREVCEM approach in order to motivate proper abstractions and idealizations. Only recently, some authors have begun to develop models that rigorously abstract from certain aspects and idealize the real situation. I will consider the tension model of Lerdahl and Krumhansl (2007) and David Huron's ITPRA approach (Huron, 2006) in more detail. Further, I will discuss a recent approach to music semantics as proposed by Schlenker (2016).

Before a proper discussion of the connection between music and emotion can begin, some qualifications have to be made. A first qualification distinguishes the idea of *inducing feelings* from the idea of *expressing feelings* (or *representing feelings*). Ball (2010) traces this distinction to Arthur Schopenhauer, who in 1819 suggested to see the interplay between *inducing* and *expressing* emotions as a distinctive feature of music (Schopenhauer, 1819). A simple example may explain the difference (cf. Kania, 2016). Consider a painting showing a man crying, but it does it so in a rather "clinical style". Then the painting can be said to *represent* the person's sadness. However, the painting does not *induce* a sad feeling. In the field of music, there are many reports that demonstrate the one and the same piece of music can trigger different feelings at different occasions even when in all situations it is clear that the piece itself expresses sad music. Krumhansl (1997) calls the idea that music *elicits* emotions the 'emotivist' position. The other idea, that music *expresses* emotions that listeners recognize in the music is called the 'cognitivist' position.

Both positions are justified under certain conditions and we have to specify these conditions carefully before a proper and systematic analysis can begin. Music can *represent* emotion in three different ways: iconically, indexically, or symbolically (Peirce, 1867, 1991). This is why music is often called the "language of emotion," by reference to certain analogies between natural language semantics and music semantics (Schlenker, 2016). Second, music can *create* real, felt emotions. Here music plays *a causal role* in producing a specific mental and bodily process in the listener that we commonly call emotion.

Another important distinction is between *emotion* and *mood*. Emotion is reactive while mood is constituted under the control of consciousness. In the discussion of free will, the veto effect plays an important role. This effect can hinder evolving action plans from being realized (Libet, 1985). Similar veto effects can play a role in emotion induction and can hinder the transfer of certain emotions into long-living mood.

The organization of this paper is as follows. In Section 2, I argue why I consider the *emotivist position* as the proper way to analyse the connection between music and emotions. Further, taking this position, I discuss the BREVCEM approach and find out what components should be involved in our bare-bone semantics of music. I also discuss two particular models that are based on the pioneering work of Meyer (1956), Lerdahl's and Krumhansl's (2007) tension model and Huron's (2006) ITPRA approach. Section 3 is devoted to discussing the structural grid of tonal music. In this section, the proper conceptualization of the domains of "tone semantics" is specified, considering tones in tonal contexts and defining different functional roles, such as tonic centers, musical forces, stability, musical movements etc. (Leman, 2012). The proposed tone semantics represents tones in an algebraic way satisfying certain restrictions of symmetry. The section further shows how the proposed semantic model accounts for available data concerning tonal attraction, graded consonance/dissonance, and musical similarity. Section 4 starts with a particular picture of emotions, the arousal-valence model. I explain the advantages of this model and contrast it with alternative ideas. The rest of this section develops an idealized model of how to map the tonal characteristics of music to its emotional content. One part of this mapping is based on the proposed tone semantics. Other parts are crucially based on tonal pragmatics, which involves the inclusion of tonal contexts and melodic expectancy. The model works for pure or absolute music. It addresses certain well-known puzzles such as the distinction between major and minor modes and the role of an *ostinato*. Section 5 summarizes the main theses of the paper and draws some preliminary conclusion including a discussion of how the present approach can be improved.

2. Earlier approaches of modelling the emotional response of music

In the introductory section, I stressed that audacious oversimplification and idealizations are a key methodology of theoretical sciences, especially in cases where we find multiple factors that influence a phenomenon. In the following section, I will list several ways how music could cause emotions and I will argue which factors are essential for a bare-bone semantics of pure music.

Musical meaning can be based on music as icon or music as index. Music as icon means that musical elements act as signs in virtue of their likeness to a designated object. Music as index means that musical elements act as signs in virtue of causal connection with its designatum. The symbolic use of music (signs in virtue of a convention) can largely be excluded (ignoring military and hunting signals). This contrasts with language, which is primary symbolic. Interestingly, natural language can have an indexical base as well. This case includes the deictic use of language and the phenomenon of connotation.[1]

What about music and communication? Does the composer or the performer of music communicate musical meaning or musical content? I think it can clearly be stated that music is not a system of communication in the sense of the American philosopher Paul Grice (Grice, 1975, 1989). Bierwisch (1978) expresses this idea as follows: "Language says, music shows what it means" (p. 60, my translation). A similar idea is expressed by Juslin and Sloboda (2013), who stress that the concept of communication requires that there must be an intention to express a particular emotion (and the listener has to recognize this intention). However, it is extremely difficult to find out the real intention of a composer as the example of Shostakovich illustrates (Gojowy, 1983). In addition, even when we recognize the intention, then it is not essential for understanding the music of the composer. The music itself shows what it means.[2]

[1] Already the German philosopher Michaelis (1795) sees tonal melodies as signs. However, for him they are not arbitrary signs that get their meaning by convention (as in language). Instead, he sees them as signs that directly and reactively express and elicit certain feelings ("Empfindungen"). This idea goes back up to Plato's doctrine of affections. The idea relates to indexical signs.

[2] However, there are uses of music where communicative intentions play a role. A standard example is Haydn's "Joke" quartet (Op. 33, no. 2). Without grasping Haydn's intention to tell a musical joke, we are faced with a passage that can only be regarded as a kind of musical mistake (London, 2008). However, this is not an example of pure music we are primarily interested in.

I think it is not wise to identify musical meaning with the emotional response of music. Rather, the emotional response is only *one* reflex of musical meaning. From the point of view of musical analysis, is it is important to have a whole battery of different measures that are appropriate to grasp the intrinsic content of tonal music. Tonal attraction, tonal forces, stability, tension and relaxation, consonance and dissonance are only a few labels that come to mind. An important task is the possibility to derive these functions from more basic and possibly more abstract functions. The situation is similar to the situation in linguistic semantics (e.g., Katz, 1972; Katz & Fodor, 1963): there are many complementary concepts such as synonymy, contrast, entailment, semantic equivalence, polysemy, etc. that can be derived from a basically theoretical (non-observable), abstract meaning entity.

The content of the present section is to give examples and to motivate the peculiarities of the underlying meaning function of music and its relation to emotions. Hence, I will prepare the grounds for finding the appropriate abstract entity by discussing possible simplifications and idealizations. Section 3 will make an explicit proposal of how to model this abstract entity in the simplest case assuming numerous simplifications.

2.1 Inducing or expressing feelings?

In the introductory section, we have distinguished the idea of inducing feelings from the idea of expressing feelings. The first idea conforms to the emotivist position, the second to the cognitivist position. Both ideas have their advantages and disadvantages. I will start with the cognitivist position. This position has the clear advantage that it abstracts from the variety of factors that may influence the induction of emotions such as the individual disposition at the moment of induction, the influence of other subjects that observe the scene etc. Another aspect of this position is the close analogy between natural language and music. Is it not plausible that the three ways in which language can symbolize external reality, namely iconically, indexically, and symbolically, also apply to music?

Examples for indexical functions include music with a programmatic title (Schumann’s *Träumerei* or Shostakovich's Symphony No. 7 *Leningrad*). Cases of musical iconicity realize similarities between auditory properties and external events. Schlenker (2017) refers to Saint-Saëns's *Carnival of the Animals* with two-note sequences for signalling a cuckoo. Other examples are Beethoven's Symphony No. 6 *Pastoral* with a naturalistic manifestation of a thunder storm.

Even when these examples are consistent with the cognitivist positions, they do not really exclude the emotivist position. Whether Shostakovich's invasion theme[3] of his Symphony No. 7, expresses a particular sad feeling or really induces it may depends on many circumstances and cannot be decided without detailed knowledge of the real performance and individual situation of the listener.

[3] This theme is a 22-bar ostinato that will pervade much of the first movement of Shostakovich's 7th symphony. The theme is inspired by Ravel's *Bolero*. The characterization *invasion theme* is misleading as pointed out by Gojowy (1983), for instance. The official Soviet-Russian interpretation of the theme was that this music characterizes the invading German troops. However, when accepting this kind of intuitive interpretation of the variation theme, it is much more plausible that this chain of variations characterizes the heroically fighting Russian compatriots of Shostakovich (Gojowy 1983, p. 70 ff). I am very sceptic that any external interpretation of this piece of music can really explain the induced or expressed emotions. In this regard, I cannot supress the remark that the Soviet musicology always suffered from the strict obligation to provide these external interpretations of the content of the Soviet composer's music.

Krumhansl has empirically investigated the question whether music elicits emotional responses in listeners or simply expresses emotions that listeners recognize in the music (Krumhansl, 1997). The study compart two groups of students both confronted with different kinds of music representing three kinds of emotion: sad, fear and happy. For one group physiological measurements were made that covered a wide spectrum of cardiac, vascular, electrodermal, and respiratory functions. The subjects of the other group had to indicate dynamic changes in emotions they experienced while listening to the music on one of four scales: sad, fear, happy, and tension. Both physiological measurements and emotion judgments were made on a second-by-second basis.

The comparison between the two main groups strongly supported the emotivist position that music really elicits emotional responses. It was demonstrated that the physiological responses induced by music were very similar to reactions found in non-musical situations. Further, a high agreement was found between the self-reports and the physiological reactions. A recent review article provides further evidence that supports this view (Patrik N. Juslin & Sloboda, 2013).

I already referred to Bierwisch (1978) – one author who warns of constructing a too close analogy between language and music. Another author who argues for a functional difference between music and language is Zbikowski (2008) who extended ideas of Tomasello (1999):

> Michael Tomasello (1999, chapter 5) recently proposed that one of the primary functions of language is to manipulate the attention of another person within a shared referential frame. It could be argued that one of the primary functions of music is to manipulate the emotions of others. Although this argument is hardly new (see, for instance, Meyer, 1956), it has often been advanced within the relatively narrow context of instrumental music produced in western Europe during the late eighteenth and nineteenth centuries. The argument could easily be broadened through the recognition that music can also manipulate the emotions through the way it shapes ritual, dance, and the rendering of a text. If it is that case that language and music have different functions within human culture – that they comprise different domains of experience – it follows that mappings between these domains would yield numerous possibilities for the sort of meaning construction associated with metaphor. (Zbikowski, p. 519)

One of the important conclusions that Zbikowski (2008) draws from the functional differences between music and language is that a mapping between these two domains could be realized via metaphoric connections. When language and music meet each other as in cantatas or songs, this is an interesting mapping mechanism exploiting the thesis that music is a language. Similarly, the opposite thesis that language is a music can be exploited for metered poetry (Lidov, 2005). Both applications are outside the domain of pure music. Therefore, my suggestion for a bare-bone semantics of pure music does not require such structural mapping mechanisms.

Summarizing, people can both express (represent) emotion and they can really feel emotion while listening to music. From linguistic theory, we know that the representation of meaning is the primary idea for formulating semantic theories. However, music is different in several respects from language. Other than language, music is not a communication system where intentions are recognized and processed. Music does not say what it means, but music shows what it means. The induction of emotion is of primary importance for understanding music.

2.2 Several ways how music could induce emotions

Juslin and Sloboda (2013) list several ways how music can arouse emotions. In the BRECVEM framework, these authors list no less than seven mechanisms through which music might induce emotions (see also Juslin & Västfjäll (2008), Juslin et al. (2010)). BREVCEM stands for **B**rainstem reflex, **R**hythmic entrainment, **Ev**aluative conditioning, **V**isual imagery, **C**ontagion, **E**pisodic memory, and **M**usical expectancy.

Brain stem reflex refers to a mechanism inducing emotion because of innate characteristics of musical pattern signalling possibly critical events. This may include sounds that are unexpected, extremely loud or dissonant. Brain stem reflexes are extremely fast and automatic. This mechanism relates to Krumhansl's (2002) claim that music can imitate the sounds of objects or events with emotional connotations. Yet such iconic use of sounds is rather limited when the great variety of music is considered. Interesting examples include the difference between major and minor chords and the role of similarities with the language system of vowels. Obviously, the iconic use of music matters in this connection. In a similar way, tempo and loudness of music has a direct influence on triggered emotions.

Rhythmic entrainment is based on the ability of the Human body to synchronize the heart rate or the breathing to an external rhythm. This can change emotional states in a reflexive, automatic way.

Evaluative conditioning is a mechanism that is based on repeatedly pairing musical episodes with other positive or negative stimuli such as visual scenes or dancing episodes. The underlying mechanism is subconscious, unintentional, and effortless. Juslin and Sloboda (2013) refer to Wagner's Leitmotif strategy as applied "evaluative conditioning".

Visual imagery is a mechanism "whereby an emotion is evoked in the listener because he or she conjures up inner images (e.g., of a beautiful landscape) while listening to the music" (Juslin & Sloboda, 2013, p. 615). Emotional contagion involves an inner imitation of voice characteristics. "The basic notion is that we get aroused by voice like features of music because a brain module responds quickly and automatically to certain stimulus features as if they were coming from a human voice expressing emotions, presumably through some kind of 'mirror-neuron' system involved in empathic reactions." (Juslin & Sloboda, 2013, p. 615).

Episodic memory refers to a mechanism of inducing emotions by certain associations in the listener's personal memory. Hence, music acquires its emotional meaning by association with significant, personal events (*They are playing our tune, Darling*). Spitzer (2002, p. 387) gives a detailed description of this phenomenon what he calls "episodic associations". Without any doubt, this is a very important mechanism for understanding musical emotions. However, if this were the only mechanism, then emotional responses to music would vary greatly from individual to individual. This could not explain the typical observation that listeners agree unusually well with one another in labelling musical emotions.

Musical expectancy is the last factor we have to discuss. It relates to the seminal work of the musicologist Leonard Meyer (1956, 1967). Meyer points out that the principal emotional content of music arises through the composer's arranging of expectations. Musical meaning and emotion depend on how the actual events in the music play against this background of expectations. Meyer argues that music that is too predictable or too surprising is likely to irritate. The secret to composing a pleasant melody is to balance predictability and surprise. Further, Meyer argues that it is the evolutionary benefit of accurate predictions that makes our brains predisposed to predict what will come next and to make us pleased when the expectations are fulfilled.

There are several arguments showing that *musical expectancy* is the most important candidate to be included in a bare-bone semantics of pure music. The factor is the main factor exhibiting systematicity, a factor that does not vary greatly from individual to individual. Further, it does not include extern sources such as personal episodic memories, visual scenes or natural language signals. Very strong emotional responses are chills and tears. Both are not evenly distributed over the time course of musical pieces, but clustered in certain structural locations. Violations of expectancy or the beginning of something new were the most reliable triggers for chills and tears (Patrik N. Juslin & Sloboda, 2013; Mori & Iwanaga, 2017). Hence, I see the factor of musical expectancy as the main pillar of music semantics, and I will ignore the other factors.

2.3 Huron's (2006) ITPRA model of emotional responses

Founded on Meyer's (1956) book, David Huron developed a suspicious and comprehensive theory of the psychology of expectation in music (Huron, 2006). Ball (2010) gives an excellent review of the work of both authors (cf. Chapter 10 of his book), and I will follow his argumentation in many points.

What is the relationship between pitch and the emotional response? Here is an example.

> There is nothing inherently pleasant about the pitch G4. But if a listener predicted the occurrence of G4, then the tone itself is likely to be experienced as pleasant. In effect, the prediction response says "congratulations for predicting the pitch G4". The real prize ought to go exclusively to the mental circuit that succeeded in making the correct response. But misattribution also tacks the warm positive feelings onto the stimulus itself. The result is that G4 "sounds nice". Conversely, if we unexpectedly hear the pitch F4, the prediction response is negatively valenced. We ought simply to feel some element of discomfort for failing to predict the stimulus. But instead, misattribution spreads the blame to include the most salient stimulus in the environment – the tone itself. As a result, the unpredicted F4 "sound bad". (Huron 2006, p.138)

Ball refers to Hindemith (1959), who exactly comes across with the assumption that only perfect coincidence between expectations and what we hear will achieve the ideal aesthetic experience. However, this is not a plausible hypothesis since it would prefer highly primitive music with only a few chords and without musical variabilities. Almost at the same time, Meyer (1956) came up with a much more plausible proposal:

> What Meyer realized is that the emotional effect of music doesn't come from having our expectations met, but from having them more or less thwarted. We don't simply want to feel happy or satisfied by music because it has gone the way we expected; indeed, that would be more likely to leave us only mildly affected by it, if not in fact positively bored. No, we like to listen to music because it sounds exciting, vigorous, poignant, beautiful, ennobled, sexy, and too many other things to list. And it achieves these things, Meyer argued, not in spite of but because of the mismatch between our expectations and the reality (Ball 2010, Chapter 10).

Meyer (1956) illustrates his basis idea with an illuminating example. Obviously, for smokers the availability/non-availability of cigarettes has a positive/negative value. Assume that such a man reaches into his pocket for a cigarette. But he has none. Then he discovers that there are none in the house, and sees that it is late and the shops are closed. This steady frustration of his desire creates first restlessness, then irritation and finally anger. And Ball (2010) continues the story by saying that just then his friend knocks and on the door and turns out to have a packet of cigarettes in his coat. Anger and anxiety give way to joy and relief. Hence, based on matches or mismatches of the listener's expectation, a positive or negative emotional value can be strengthened. For listeners of music, the entities that are evaluated are consonance and dissonance. Consonance has a positive value and dissonance has a negative one.

It follows that a model aiming to describe affective responses has to include several components. Huron's (2006) ITPRA model of emotional responses is based on five functionally distinct components. These five constituents can evoke responses independently. Unfortunately, Huron does not specify how these responses are integrated into an overall feeling. Perhaps, this dynamic mixing of the feelings generated by the five subsystems has to happen within a few seconds only. These are the five subsystems (conforming to five different response types):

1. **I**magination: future-oriented behavioural motivation – feeling the future (daydreaming)
2. **T**ension: mental preparation for an anticipated event
3. **P**rediction: positive or negative reinforcement to encourage the formation of correct expectations
4. **R**eaction: automatic reflex-like response that assumes a worst-case assessment of the outcome and generates an "immediate protective response"
5. **A**ppraisal: conscious response that results in positive or negative reinforcement.

Fig. 1 gives a schematic sketch of the model classifying the responses into pre-outcome and post-outcome responses. It should be mentioned that this classification does not characterize the time course of the response. Rather, it characterizes the kind of inputs required for the system.

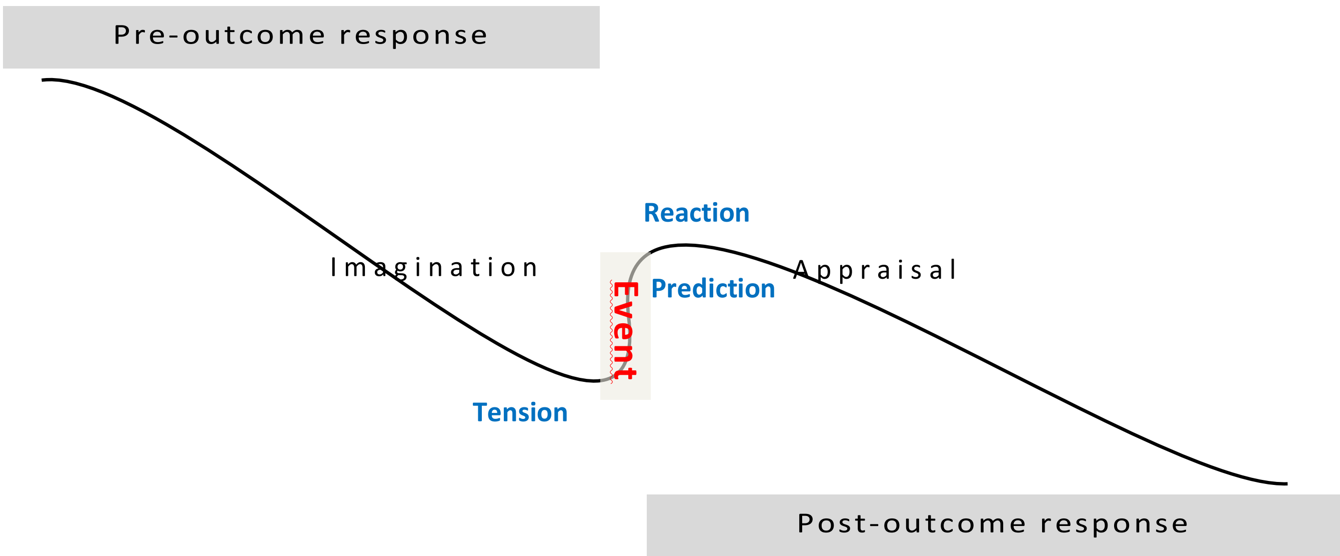

Fig. 1 : Schematic representation of the five types of responses following Huron's ITPRA model (Huron, 2006).

We have two response types that refer only to information that is available before the critical event is occurring (imagination and tension) and three response types that refer to information including information that is only available after the critical event has appeared (appraisal, reaction and prediction). The two types, imagination and appraisal, are conscious responses requiring explicit attention to the music whereas the other three types do not require attention and relate to automatic processing. In many listening situation, we direct no explicit attention to the music, i.e. the listening act is more or less in the automatic mode. Huron (2006, p. 310) calls this mode of listening "stream-of-consciousness listening". It comprises the three response types tension, prediction and reaction. I will focus my discussion on these three primary response types.

The two response types that are involved in explaining the cigarette example are the prediction response and the reaction response. The purpose of the prediction response is to induce positive or negative inducements that encourage the formation of correct expectations. The quale of pleasure origins in the prediction response. Correct expectations get a positive prediction response – even when the expected outcome is bad. In contrast to the prediction response, which does not depend of the quality of the relevant pitch event, the reaction response does.

The tension response acts as an automatic, mental preparation for the anticipated event. This kind of "pre-outcome limbic reaction" results in changes of arousal and attention and prepares some expected event. The probability for an expected event can have a higher or a lower degree. Huron notes that even when positive outcomes are predicted, high uncertainty will lead to unpleasant stress, arousal and vigilance. Delay tend to magnify the tension response – increasing the probability that the event will occur next.

The tension response can be considered as evaluating emotion on the axis of activation (with different levels of arousal). This contrasts with the cumulative outcome of prediction and reaction response. Their collective outcome can be measured on an axis of valence (from pleasure to displeasure). Both axes together allow the description of emotions in terms of a particular model of emotions: Russell's (1979) arousal-valence model (see Section 4.1).

Further, it should be noted that the reaction response and the (conscious) appraisal response may reinforce one another but they also can evoke contrasting emotions (inducing laughter, for example). From a neurological perspective, the reaction response is an extremely fast response that assumes a worst-case assessment of the outcome. This contrasts with the slow appraisal response, which can result in conflicting responses.

Unfortunately, Huron's ITPRA model does not make quantitative predictions. However, the model list a variety of informal insights and opportunities. For instance, Huron (2006) considered a reach system of qualia and discussed it in the context of tonal music. He identified seven clusters of responses in his qualia survey: 1. certainty/uncertainty, 2. tendency, 3. completion, 4. mobility, 5. stability, 6. power, and 7. emotion. For the interpretation of musical pieces, the possibly most interesting cluster of qualia refers to emotion.

> Here the tonic, median and dominant pitches were described using such positive hedonic terms as *pleasure*, *warmth*, *contentment*, *beauty*, and even *love*. Negative hedonic terms like *harsh*, *jarring*, *uncomfortable*, and *anxious* were applied to such tones such as the raised supertonic, the raised subdominant and the raised dominant. (Huron 2006, p. 164).

This picture of emotion representation is obviously one-dimensional with an axis from negative to positive hedonic evaluation – the axis that corresponds with the axis of valence in the arousal-valence model of emotion.

Interestingly, Huron (2006) discusses the phenomenon of *frisson* using his one-dimensional picture of emotion (p. 281 ff). Referring to Sloboda (1991), Huron points out that such extreme emotions as chills can be induced by strongly unexpected modulations. The example is a passage from Schönberg's *Verklärte Nacht* (bars 225-230) where a change from E-flat minor to D-major takes place, accompanied by an unexpected change of dynamics and rhythm. Other elementary phenomena that are discussed by Huron are *anticipation* and *suspensio*n. In Section 4, I resume this discussion using Russel's (1979) representation of emotion and proposing an explicit model of how to integrate the different response types.

2.4 The tension model of Lerdahl and Krumhansl (2007)

A precise definition of musical tension (and relaxation) is difficult to give. Lerdahl and Krumhansl (2007) see a close relationship to other concepts such as (cognitive) stability/instability and (musical) consonance/dissonance. In Huron's (2007) qualia survey, the empirical concepts of tension and relaxation relate to the following types: tendency, completion, and stability.

In order to measure tension/relaxation, subjects are typically asked to respond to degrees of tension and relaxation on a continuous scale with the help of a computer mouse (cf. Lerdahl & Krumhansl 2007). Interestingly, these instructions elicit consistent interpersonal responses. Experimentally, ratings of tensions are measured sequentially in a stop tension task and continuously in a continuous tension task. In the former task, the presentation of the sound events stop at a certain point (first, second, third, etc. sound event) and the subjects have to rate the degree of tension at that point. In the latter task, the presentation is continuous and the subjects have to move a slider of the graphical interface. The continuous task enables real-time judgments of musical parameters such as tension and other proxies for musical affect. Experiments by Bradley W. Vines, Nuzzo, and Levitin (2005) and Lerdahl and Krumhansl (2007) use continuous tension judgment in order to assess a participant's real-time experience of a musical piece.

In their tension model, Lerdahl and Krumhansl (2007) give an explicit account how an arbitrary sequence of events in any passage of tonal music can be mapped upon the affective parameters of tension and relaxation. In fact, they note that for the specification of tension, concepts of four fields have to be combined: 1. tonal pitch space including a static attraction function, 2. musical consonance/dissonance, 3. hierarchical event structure, 4. voice-leading melodic attractions. So, the empirical (observable) concept of tension can be expressed in terms of more elementary, theoretical concepts. Here is an outline of the basic concepts entering the tension model.

First, let us consider the tonal pitch space. In this paper, we assume 12 pitch classes, also called tones. Tones are equivalence classes based on octave equivalence. We will use a numeric notation to define the twelve tones of the system, in ascending order:

(1) 0 = C, 1 = C♯, 2 = D, 3 = D♯, 4 = E, 5 = F, 6 = F♯, 7 = G, 8 = G♯, 9 = A, 10 = B♭, 11 = B.

A numerical representation of Lerdahl's basic space is given in Table 1. It considers five levels of a tonal hierarchy: **A**: root (root tone, represented by the pitch class 0 in the present case), **B**: Fifth, **C**: Chord, **D**: Diatonic (including all diatonic pitches of C major in the present case), **E**: all (including all 12 pitch classes).

Level **A**	0											
Level **B**	0							7				
Level **C**	0				4			7				
Level **D**	0		2		4	5		7		9		11
Level **E**	0	1	2	3	4	5	6	7	8	9	10	11
Embedding distance *c*	0	4	3	4	2	3	4	1	4	3	4	3
Tonal attraction $s = 6.5 - c$	6.5	2.5	3.5	2.5	4.5	3.5	2.5	5.5	2.5	3.5	2.5	3.5

Table 1: The basic tonal pitch space as given in Lerdahl (1988).

For each tone, the *embedding distance c* is calculated by counting the number of levels down that a pitch class first appears. The embedding distance relates to static tonal attraction. It can be measured empirically and relates to the degree of how well a given (octave) tone fits into a tonal scale or tonal key, let it be a major or minor key. In an celebrated study, Carol L. Krumhansl and

Kessler (1982) asked listeners to rate how well each note of the chromatic octave fitted with a preceding context, which consisted of short musical sequences in major or minor keys. The results of this experiment clearly show a kind of hierarchy: the tonic pitch received the highest rating, followed by the pitches completing the tonic triad (third and fifth), followed by the remaining scale degrees, and finally the chromatic, non-scale tones. This finding plays an essential role in Lerdahl's and Jackendoff's generative theory of tonal music (Lerdahl & Jackendoff, 1983) and is one of the main pillars of the structural approach in music theory. Fig. 2 shows the data of Carol L. Krumhansl and Kessler (1982) for the key C-major. Further, it gives the predictions of a simple model based on embedding distance (using the linear fit $s = 6.5 – c$ for expressing static tonal attraction.

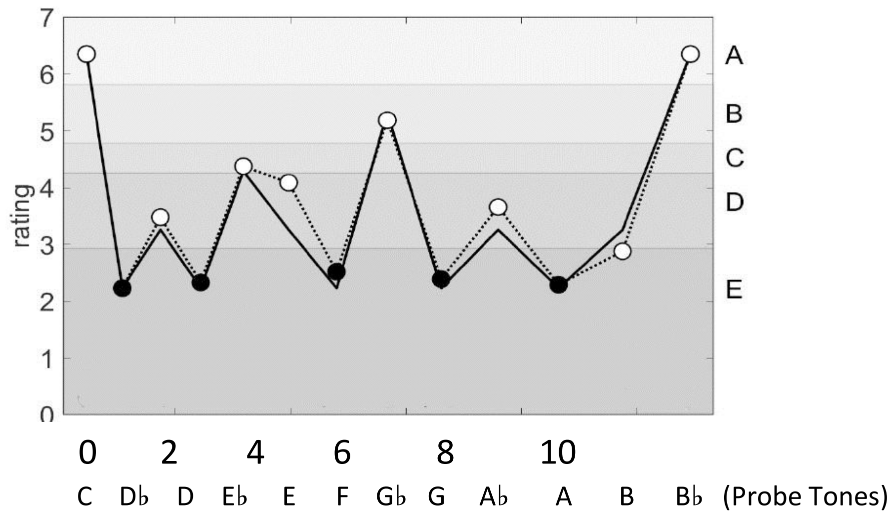

Fig. 2 : Distribution marked by the dotted line shows the data of Krumhansl & Kessler (1982) for the key C major; open bullets indicate scale (diatonic) tones; closed bullets denote non-scale (chromatic) tones, for the C major scale. The black curve presents the fit of the embedding distance *c* relative to the data of Krumhansl & Kessler (1982). This fit is specified by the linear approximation $s = 6.5 – d$. This gives for pitch class 0 the value 6.5 (maximal attraction) and for pitch class 1 (minimal attraction) the value 2.5.

The embedding distance and the attraction function can be used to express the distance between tones and chords (Blutner, 2015). However, static tonal attraction is not enough for expressing musical tension and relaxation.

A second pillar for modelling tension is musical consonance/dissonance. The notion of dissonance (called "surface tension" in the article) could be modelled using established measure of sensory dissonance in the psychoacoustic literature (cf. Section 3.3 for more details). Lerdahl and Krumhansl (2007), however, use a categorical notion based on the simple intuition that nonharmonic tones (tones not belonging to a sounding triad) are causing dissonance – diatonic non-chord tones to a lower degree than chromatic non-chord tones. Using this categorical definition of dissonance (for details, see p.334), two technical concepts of tension are defined, called sequential tension (ST) and hierarchical tension (HT).

(2) a. ST(y) = Diss(y) + $\delta(x, y)$, where x is the chord preceding chord y
b. HT(y) = Diss(y) + $\delta(x_{dom}, y)$, where x_{dom} is the chord that directly dominates y in the prolongation tree.

Hereby, $\delta(x, y)$ is the distance between the chords x and y (relative to a given key). Accordingly, sequential tension of a chord y is increasing with the degree of dissonance and the distance between it and the preceding chord x. In the simplest case, the distance function can be constructed by combining the fifths cycle and the relative and parallel major-minor cycle. With each application of one-step in the circle of fifth, the distance rises by one-step (transforming C to G or C to F, for instance).[4] Note further that the definition of the hierarchical tension concepts the hierarchical event structure plays an important role. This is defined by prolongation trees that are borrowed from Lerdahl and Jackendoff (1983). Unfortunately, both sequential tension and hierarchical tension give very low correlations with empirical data. Using the Grail theme of Wagner's Parsifal, the correlation between the data and sequential tension is close to zero, for hierarchical tension it is $R^2 = .43$. Therefore, a third factor has to be considered.

This third factor is called "voice leading attraction". In the following, I will call it "dynamic attraction" (cf. Parncutt, 2011). In contrast to static attraction, which considers how well a tone fits to a given chord, dynamic attraction considers the expected harmonic development. Importantly, there is a distinction between tonal hierarchies and event hierarchies. The latter are "part of the structure that listeners infer from temporal musical sequences" (Lerdahl 1988: 316). Data that concern the progression of tones or chords should be explained in terms of such event hierarchies. The classical tonal attraction experiments can be modified by asking listeners to rate the degree to which a tone or chord is expected in that context – following a sequence of pitches or chords as the subsequent element. Some of the classical studies (Cuddy & Lunney, 1995; Carol L. Krumhansl, 1995; Schellenberg, 1996; Thompson, Cuddy, & Plaus, 1997) used the probe-tone technique to investigate the tone-to-tone expectancies for continuations of melodies. These studies are important to test the dynamic predictions of models of melodic expectancy, such as Narmour's *implication realization model* (Narmour, 1991, 1992).

I will discuss now how Lerdahl (1996) and Lerdahl and Krumhansl (2007) extend the static attraction model to the level of individual pitch sequences. In order to model the *dynamics* of attraction we need besides a general context (normally established by a certain *key*), a cue tone x (or a sequence of cue tones $\vec{x}$) and a probe pitch y whose attraction value in the environment *key*+x has to be calculated. The attraction algorithm proposed by Lerdahl (1996) gives the following formula to calculate the dynamic attraction from pitch x to pitch y. Note that the context *key* defines the basic tonal space including the tonic triad.

(3) $F_{key}(y|x) = \frac{s(y)}{s(x)} \cdot \frac{1}{n^2}$, with $s(C) = 4, s(E,G) = 3, s(D,F,A,B) = 2,\ s(X\flat) = 1$

Hereby $s(y)$ is the tonal attraction (also called "anchoring strength") of pitch y in the basic tonal pitch space. In the previous section, we have assumed 5 levels of description and 5 as the highest possible number of embedding (following Lerdahl, 1988). Lerdahl (1996) has eliminated the level **B** in his attraction model. As a consequence, the expression $s(y) = 4 - c(y)$ is used for defining the static attraction function (anchoring strength). This gives identical embedding distances (and anchoring

[4] Cf. Lerdahl (1988, 2001) for more details, and Blutner (2015) for alternatives.

strengths) for the pitches E and G, in contrast to the original model including level **B**. In formula (3), the proportion $\frac{s(y)}{s(x)}$ of the two anchoring strengths is multiplied by a factor $\frac{1}{n^2}$, where the number *n* counts the semitones between pitch *x* and pitch *y*. For instance, when calculating the attraction from D to C (relative to key = C major), we get $\mathrm{F(C|D)} = \frac{4}{2} \cdot \frac{1}{2^2} =$ ½. The highest value we get when considering the attraction from B to C: $F(C|B) = \frac{4}{2} \cdot \frac{1}{1^2}$ = 2.

Note that the attraction function is not symmetric. For instance, the attraction from C to B is $\mathrm{F(B|C)} = \frac{2}{4} \cdot \frac{1}{1^2} =$ ½, i.e. only one quart of the attraction $F(C|B)$ from B to C. Obviously, the inverse quadratic distance dependency is borrowed from physics. We find it for forces between electric charges. Another example is the classical formula for calculating the gravitation force between two mass points. What about the pendant of the anchoring strength in physics? In the first case, it relates to the *charge* in electrodynamics, in the second case it relates to the *mass* in gravitation. However, instead of the asymmetric quotient, the (symmetric) product function applies in the physical cases. This makes the physical forces symmetric, in sharp contrast with the musical forces that are always asymmetric.

If more than one melodic line is considered, the sum of the realized voice leading attractions (dynamic attractions) $\sum F_{key}(y|x)$ has to be considered (Lerdahl and Krumhansl 2007). In this way, the calculated dynamic attractions stand for relations between two subsequent chords. When making empirical predictions, the sum value $\sum F_{key}(y|x)$ is calculated and it is taken for the dynamic attraction of the chord consisting of elements *x*. Hence, if these elements define a relatively dissonant chord that resolves into the tonic chord, then the calculated value will be relatively high.

With the help of dynamic attraction (without sequential or hierarchical tension) the following correlation between data and model has been calculated for the *Grail theme*: R^2 = .35. Combining dynamic attraction with sequential tension (p. 341) gives only a marginal improvement R^2 = .36. For hierarchical tension combined with dynamic attraction, however, the calculated value is R^2 = .75. With further adjustments, finally the authors get a very high correlation of R^2 = .97.

However, even if the experimental fits between model and data are impressive, there remains some theoretical oddity. The point is that a real explanation requires more than data fitting. It requires an independent theoretical motivation of the made assumption; it requires a real grounding of the introduced formulas. This makes the difference between a good theory and a descriptive model. A proper theory should be able to derive particular rules or laws by general principles. These principles should have an independent motivation within the field of exploration. For example, there are plausible cognitive respects that could motivate the asymmetry of tonal forces. Unfortunately, these respects do not play a visible and principled role in deriving formula (3). The cognitive aspects do not have a pendent in the physical domain. Therefore, the analogy between cognition and physics fails.

Some authors have proposed formulas differing from the dynamic attraction formula (3). For an overview of several approaches, the reader is referred to Steve Larson (2012). In my opinion, the creation of such formulas and the extensive efforts of data fitting using such formulas does not really lead to a deeper musical understanding of what goes on. It is important to see that an understanding of music requires more than assuming gravity-like forces:

> If music was simply a matter of following gravity-like attractions from note to note, there would be nothing for the composer to do: a melody would be as inevitable as the path of water rushing

down a mountainside. The key to music is that these pulls can be resisted. *It is the job of the musician to know when and how to do so*. (Ball 2010: 97)

A further criticism concerns the hierarchical model of the basic tonal pitch space, which should give an adequate description for the static attraction profiles. First, this model should explain that for both major and minor profiles, scalar tones have higher values of tonal attraction than non-scalar tones. A second general finding is that all tones of the tonic triad have higher values than the other tones of the scale (Temperley 2007: 84). However, both of these important empirical facts are directly stipulated by the hierarchical model: by assuming a "diatonic space" (level **D**) which includes all scalar notes and by assuming a higher order "triadic space" (level **C**) that includes the tones of the triadic space.[5] The stipulations of the hierarchical model concern the number of levels and the precise content of some levels. For instance, they concern the question of which chords constitute the triadic level. For Western music, the decision is easy to make by assuming that we have a clear distinction between major and minor systems. Non-Western kinds of music need not conform to the major/minor system and can be based on tonal scales quite different from those of Western music. Alternative scales such as Indian ragas or the scales underlying traditional Japanese music are widely used in world music. It is not clear how we can modify or extend the hierarchical model in order to account for the traits of these kinds of music (Klein & Jacobsen, 2012).

Musical tension is a main issue in addressing the domain of meaning for tonal music. In this subsection, I have illustrated that quantitative models of this dimension of meaning are possible – based on research by Krumhansl, Lerdahl and others. Concerning musical tension, it needs a number of theoretical basic concepts that have to be clarified independently – including a metric of tones and chords, a graded conception of consonance and dissonance, a notion of static attraction and a notion of voice leading attraction (dynamic attraction). An important problem is whether these notions are sufficient for constructing a direct mapping between the tonal events and the elicited affective meanings. Is it possible to find a compromise between the quantitative tension model and the qualitative ITPRA model? Is it conceivable in this way eliminate the disadvantages of both models while maintaining their advantages? Before I try to answer this question (in Section 4), I will discuss a recent proposal of Schlenker (2017) endorsing an explicit music semantics.

2.5 Schlenker’s (2017) conception of music semantics

In analogy to the standard semantics for natural language (Montague, 1970), Schlenker (2017) proposes an *external semantics* for music. This contrasts with an internal semantics, an analysis of meaning that does not refer to elements outside the realm of pure music, but simply joins inner-musical events and expectations. Lerdahl (2001) has described this as a journey through tonal pitch space. Authors such as Deutsch and Feroe (1981), Meyer (1956), and many others have developed similar ideas.

Most of Schlenker's examples of semantic effects in music clearly refer to program music, for example Strauss's *Zarathustra*, with a sunrise at the beginning, or episodes from Saint Saëns's *Carnival of Animals*. In all these cases, it is believable that "the meaning of a musical piece is given by

[5] Another finding is that for major scales the tonic pitch has a higher attraction value than the fifth. In the hierarchical model, this suggests the assumption of an open fifth space (level B). Unfortunately, this conflicts with the Kostka-Payne corpus data (Kostka & Payne, 1995). Consequently, the existence of this level of fifth space is questionable, for it does not apply for all assumed tonal scales. We have mentioned already that Lerdahl (1996) has eliminated the level B in his attraction model.

the inferences that one can draw about its 'virtual sources'" (Schlenker 2017, p. 3). Schlenker (2017) gives an illustration by means of the third movement of Mahler's first Symphony. It demonstrates how musical-internal parameters – such as loudness – raise effects for musical-external parameter's – such as the spatial distance of a procession. At the end of the movement, a *decrescendo* may yield the impression that the procession is moving away from the listener. Similarly, in the fourth movement of Beethoven's Symphony No. 6 *Pastoral*, a corresponding *decrescendo* yields the strong impression that the thunder storm is moving away from the pastoral area.

However, in most cases we understand music without knowing such outer sources. This is true even for program music. Hence, I would like to come back to my earlier methodological point. Before we can describe the interaction between music, outer events and possibly language, it is required to investigate the simpler case of *pure music* and to develop a bare-bone semantics for it. On this base, then, we can investigate what further effects outer sources can induce. To start with the opposite strategy is not very useful. The development of Soviet-Russian musicology in the Stalin era illustrates the disaster that can happen when an exclusively external semantics for music is approved (cf. Gojowy 1983).

Another example of an external semantics relates music with physical motion, i.e. the semantic effects of music are analysed in terms of physical motion. This has been called the "metaphoric theory of musical meaning". This account means that acoustical sequences are connected with a "journey through the tonal space" – a journey that triggers inferences about movement in, let's say, physical space as given by the examples before. Proponents of such ideas are Larson (2012), Desain and Honing (1996), Honing (2003) inter alias. An early representative of this view is Jan Broeckx, whose book appeared already in 1981. In this book (Broeckx, 1981), the metaphoric link between musical Gestalts and the expression of emotions was defined as follows:

> "Musical figures evoke moods and feelings by synaesthetic and kinaesthetic analogies of their formal pattern and sonorous qualities with bodily attitudes and movements, and with visual or tactile sensations, integrated into the domain of emotional experience by means of metaphorical transfer." Translated from the original Dutch text. (Broeckx, 1997, p. 266)

In a later article, Broeckx (1997) is now convinced of "the fallacious use of the notion of metaphor" (p. 267) and contrasts his earlier view with his present one. The crucial point is that a metaphoric theory makes sense only when music interacts with non-musical domains (such as domains referred to by natural languages). However, for the study of pure music this is not required. The very same holds for Larson's metaphoric theory that likewise can be avoided in the context of pure music.[6]

Finally, I should add that I do not argue against the possibility of an external music semantics as proposed by Schlenker. My point is that music semantics has also an important internal side that cannot be ignored since it is essential for pure music. Adding semantic elements that draw inferences about external sources can lead to interesting insights. However, the influence and power of the external effects can only be evaluated in relation to the influence and power of the internal effects. Of course, this is not a glorification of pure music (Dahlhaus, 1994) but only a justification of the role of pure music when it comes to clarify the relation between music and its affective content.

[6] It should be noted that this is not an argument against the reality of musical forces that play an essential role in Larson's theory. However, these forces should be treated in a completely different way (beim Graben & Blutner, 2018)

3. The structural grid of tonal music

In the last section, I have argued for the development of an internal semantics for pure music. The semantic account should be the starting point for modeling the intrinsic content of tonal music. It should include the calculation of certain functions, such as (i) static attraction, (ii) musical consonance and dissonance, (iii) musical distance between tones, (iv) dynamic attraction (voice leading attraction), and possibly some more functions that are required for calculating the affective response of tonal music. An important challenge is to find possibilities for deriving these functions from more basic and possibly more abstract entities.

In linguistics, we have the central distinction between semantics and pragmatics. Semantics is concerned with the meaning of natural language expressions. Pragmatics addresses the use of these expressions in concrete contexts. In music, a similar distinction can be made exploiting the general character of semiotics (see the final part of Section 3.5 for more details). In the following chapter, I will argue that static attraction, musical distance and musical consonance are primarily semantic concepts, which are based on the musical meaning of tones and only depend on *key* as a contextual element. The situation is different for dynamic attraction. This notion cannot be properly reduced to tonal meaning. Therefore, it is assigned to the pragmatic component.

3.1 An algebraic description of pitch systems

This subsection starts with an algebraic description of pitch systems. On this base, different representational means of describing the meaning of tones and chords can be deliberated. These ideas form the base of musical semantics. Western music is grounded on 12 pitch classes called tones. This contrasts with other systems that divide the octave in 24 parts (Arabic music) or 31 parts (Indian Music). Subsequently, I will exclusively be concerned with systems based on 12 tones. Musical tuning is important for establishing typical intervals between the 12 tones. I will assume equal temperament, a system of tuning in which the frequency interval between every pair of adjacent tones has the same ratio. This tuning system is important for realizing an important symmetry, called the *principle of translation invariance*. Informally, this principle says that the musical quality of a musical episode is essentially unchanged if it is transposed into a different key (of the same mode – major or minor).

The role of symmetries is important for finding the proper representations of the twelve tones. For the treatment of symmetries in tonal music, I follow Balzano (1980). Mathematically, symmetry is simply a set of transformations applied to given structural states such that the transformations preserve the properties of the states. Precisely, the *principle of translation invariance* states that the crucial musical properties are unchanged when the operations of the cyclic group $\mathbb{Z}_{12}$ are applied. The cyclic group $\mathbb{Z}_{12}$ has exactly 12 elements.For a cyclic group there exists at least one element *g* of the group such that every element in the group can be represented as a composition of *g*'s. The element *g* is called a *generator* of the group.

Let us now enumerating the 12 tones as indicated in (1). Then the group elements can be thought as tonal intervals. Each interval transform the 12 tones into tones increased by the interval. Obviously, there are two different generators of this group. Using modular arithmetic, we can describe the two different generators as follows:

(4) a. $g^C(k) = (k+1)\ mod\ 12$, i.e. $g^C(0) = 1, g^C(1) = 2, \ldots, g^C(11) = 0$
b. $g^F(k) = (k+7)\ mod\ 12$, i.e. $g^F(0) = 7, g^F(7) = 2, \ldots, g^F(5) = 0$

The first operation g^C maps the tones to another tones by adding an interval of one semi-tone (**C**hromatic way of representation). The second operation g^F maps the tones to another tones by adding the interval of **F**ifth (7 semi-tones). Fig. 3 gives a visualization of how each of these two operators (chromatic circle on the left hand side, circle of fifth on the right hand side) generates all 12 pitch classes starting with the tone C. The numbers now indicate how often the generator is applied.

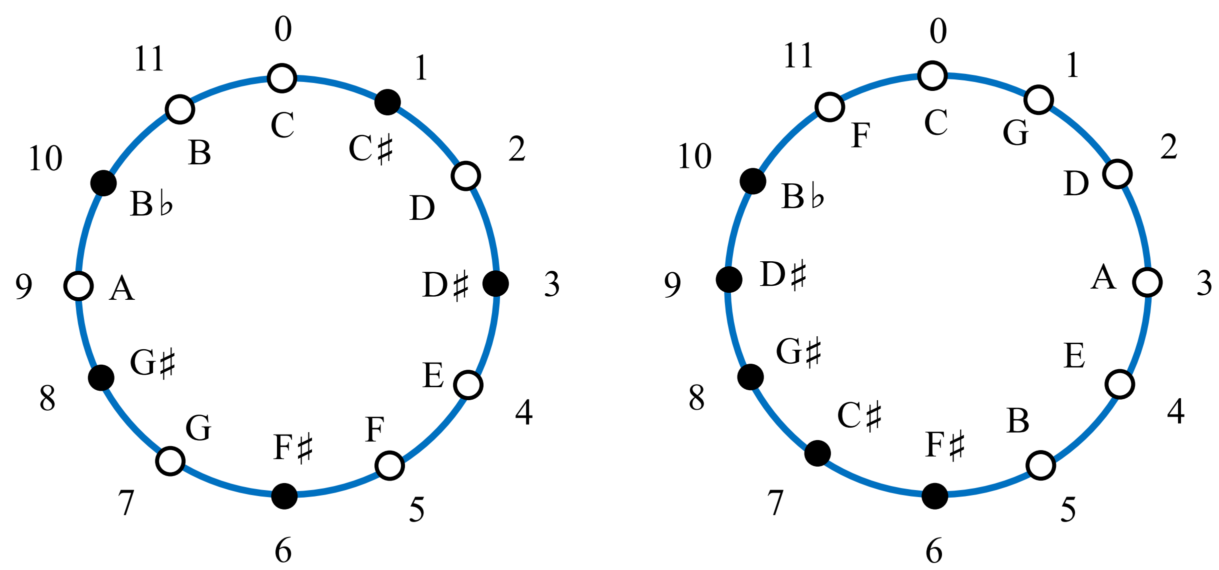

Fig. 3 : Visual representation of $\mathbb{Z}_{12}$. On the left hand side the twelve tones are created by the semi-tone generator g^C. Open bullets indicate scale (diatonic) tones; closed bullets denote non-scale (chromatic) tones, for the C major scale. On the right hand side, the twelve tones are created by the generator g^F that transposes by seven semi-tones (resulting in the circle of fifth). The numbers indicate how often the generator is applied recursively. The tones in the inner circles are the results of application of the corresponding group element to the basic pitch class C.

Interestingly, the tones (or intervals) of a diatonic scale form a connected set if the generator g^F has been used. This property of connectedness will represent an advantage in many applications. Therefore, we will prefer to use the generator g^F corresonding to the circle of fifth.

So far, we have used natural number for representing tones and we have used modular arithmetic for expressing the group operations as in Eq. (4). When we ask for the *meaning* of tones, chords, intervals etc. this kind of representation seems not to be sufficient. In the literature, there are several proposals what kinds of representations are eligable. First, there are spectral representations based on Helmholtz (1863) and his followers. Modern representatives of this approach are Milne, Laney, and Sharp (2015). Second, we have to consider cyclic pulse representations that are useful for implementing the idea of tonal fusion (Stumpf, 1883, 1890). Modern representatives of this approach are Ebeling (2008) and Stolzenburg (2015). The third kind of approach includes template-based models that approach the question how tone-like a sound is. The starting point of this research was set by Terhardt, Stoll, and Seewann (1982) – followed by authors such as Parncutt (1988), Hofmann-Engl (2004), and others. Template-based models are relational. They do not constitute the meaning of tones by themselves. Instead, they build a system of relations with other tones. This can be seen in analogy to language, where likewise the meaning of words cannot be seen in isolation. Rather, the systems of word meanings define fields that have to be considered in a holistic way (Erdmann, 1901; Paul, 1880). Fourth, there are abstract vector representations based on neural networks (Large, 2010; Large, Kim, Flaig, Bharucha, & Krumhansl, 2016) and the framework of quantum cognition (Blutner, 2015). This is not the place to give a detailed desciption and discussion of these four different approaches. However, it should be menti-oned that all four approaches are based on the wisdom of vector-models of cognition and all three apply the representational means of basic linear algebra.

Because of the assumed linearity, all of these approaches can be taken to define the relative meanings of tones in terms of a kernel function $\mathrm{K}(k-l)$. This kernel function should be seen as *theoretical conc*ept that does not have a direct empirical impact. Intuitively, the value of the function $\mathrm{K}(k-l)$ describes how well the two tones k and l (using the enumeration of the circle of fifth) fuse into the sound of the interval spanned by the two tones.

In the following subsections, I explain how this semantic kernel function operates in tonal contexts (chords, cadences). In this way, it defines concepts with significant empirical impact such as static attraction (Section 3.2), grades of consonance/dissonances (Section 3.3), underlying scales and musical similarity (Section 3.4). Furthermore, Section 3.5 calculates dynamic attraction profiles based on a modified kernel function that goes beyond pure tonal semantics.

3.2 Static attraction

How well does a given tone fit with the tonic pitch, a tonal chord or another tonal context? What is the probability that it fuses with (or collapses into) the tonal context collapses into a comparison state? In an idealized world, the context could consist of a single tone (the tonic l). In this case, we can take the abstract kernel function $\mathrm{K}(k-l)$ to describe the degree/probability of fusion.

However, in the classical experiments by Krumhansl & Kessler (1982), the context consist of more than a single tone. For example, it is a triad, a cadence, or a whole diatonic scale. In order to calculate the attraction profiles for such richer contexts, the operation of discrete convolution is applied. This operation merges the context with an origin (context-free) attraction profile described by the kernel function. Similar operation have been applied in image processing for constructing particular filters (Kalra & Peleg, 2006) and in distributional semantics for modelling adjectival modification (Blutner, 2010; de Groot, 2013; Mitchell & Lapata, 2010).

Discrete convolution is an operation that takes two functions *f* and K and defines the outcome function $f * \mathrm{K}$ as follows:

(5) $(f * \mathrm{K})[k] = \sum_l f(l) \cdot \mathrm{K}(k-l)$

In the present case, the function f is the characteristic function defining the context. It gives the value $f(k) = 1$ if the tone k is in the context; otherwise it gives 0. The function K is called "kernel" of the convolution.

In Section 2.4, I have described the basic tonal pitch space as given in Lerdahl (1988) and the hierarchic model of tonal attraction. It is an easy exercise to approach the hierarchical model by a symmetric kernel function plus a tonal context defined by a triadic chord.

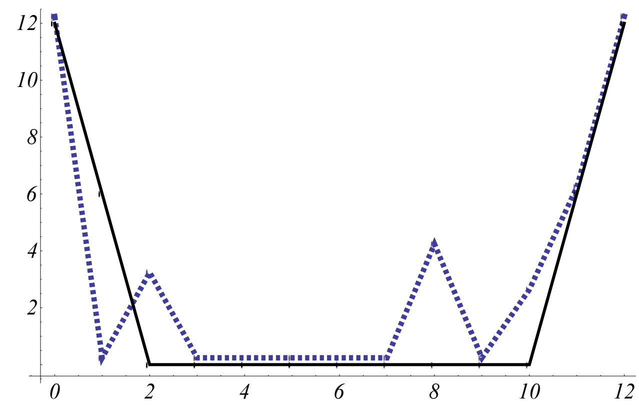

Fig. 4 : Symmetric and asymmetric kernel function. The symmetric kernel function (solid curve) is based on the hierarchical model. The asymmetric kernel function (dashed curve) conforms to Parncutt's (1988) template-based model.

The kernel function (arbitrary scaling) assumes the maximum value (= 12) for the prime and octave intervals and half of it (= 6) for the fifth. All other intervals are damped to the zero level. Then it is straightforward that circular convolution with the major chord {0, 1, 4} generates the highest value for the tonic 0 and the fifth 1. This value is 18 = 12 + 6 + 0. For the third tone (4) of the triad the circular convolution generates the value 12 = 12 + 0 + 0. Since every tone of the diatonic major scale can be accessed via a fifth (upward or downwards) from a tone of the triad, this gives the value 6 = 6 + 0 + 0 for the diatonic tones (with exception of the tones of the triad). The remaining tones are not elements of the diatonic scale and get the value zero, as expected. The correlation between the attraction values calculated with the given symmetric kernel function and the data (Carol L. Krumhansl & Kessler, 1982) is very high: R^2 = .92 for the major data and R^2 = .83 for the minor data.

After performing an appropriate linear transformation, the results are comparable (or even a bit better) than the original hierarchical model (see Fig. 2). An important difference with the original model (Lerdahl 1988) is that the kernel approach conflates the levels **A** and **B** of the hierarchical model. It should be noted that both the original hierarchical model and the present mimicry with a symmetric kernel function have crucial problems when the description of the minor data is considered. The high correlation coefficient found in this case should not deceive us. The point is that the minor chord does not have such a clear root tone as the major chord. The major chord clearly has C as root tone, followed by G. G is followed by E with decreasing attraction. E has an attraction value which is about half of the tonic (Krumhansl & Kessler 1982). This contrasts with the minor chord. Again, C gets the highest attraction closely followed by E♭, and the latter is followed by tone G. Unfortunately, this contrasts with the predictions of the hierarchical model that predicts that G gets a higher attraction than E♭. Some authors even doubt that C is the root tone in the minor case (Hofmann-Engl, 2004).

In order to overcome these difficulties an asymmetric kernel function is required. This point can clearly be seen in connection with template-based models, e.g. the model of Parncutt (1988). Based on Terhardt et al. (1982), Parncutt (1988) developed a model for finding the (best) root tone of a chord or a tonal scale. The basic idea is that a good rote tone contains as many as possible overtones of the chord under discussion. To handle the corresponding optimization problem, Terhardt and his followers consider five supporting intervals (given by the series of overtones up to number 15): unison, fifth, major third, minor third, major second and minor seventh. Parncutt (1988) proposed a non-uniform weighting of these intervals. This can be described by an asymmetric kernel function. This kernel function is shown in Fig. 4 (dashed curve). Using major and minor triads as contexts, a correlation of R^2 = .9 results for both the major and the minor data (applying the discrete convolution operation). In the minor case, E♭ gets a higher attraction value than G in agreement with the data.

3.3 Predicting grades of consonance and dissonance for musical intervals

Musical intervals are simultaneous combinations of two pitches. Why do some intervals sound better (more harmonic) than others? What constitutes the difference between consonant and dissonant intervals in music? Consonance and dissonance play crucial roles in music across cultures (Meyer, 1956). As I have outlined in Section 2.4, dissonance is naturally associated with musical tension whereas consonance is associated with relaxation and stability.

The terms "consonance" and "dissonance" are theoretical terms in a theory of tonal music. Depending on the theoretical context, these terms may have different meanings. The situation is the same as in linguistics, where, for example, there is not only one conception of what a word is. Instead, there are several distinct uses of this term referring to the phonetic word, phonological word, morphological word, graphemic word, syntactic word, and prosodic word. Hence, the same term is used in different ways conforming to different levels of linguistic descriptions. In a similar vein, when we use the terms "consonance" and "dissonance" we should specify at which level of theoretical description we use these terms.

In this paper, we will distinguish two levels – the perceptual and the cognitive level. The perceptual level is investigated in psychoacoustics. It relates the relevant physical properties of sensory stimuli and the psychological responses evoked by them. The cognitive level refers to psychological processes which go beyond the purely sensual processes such as in the musical context of counterpoint. Hence, we will make the distinction of perceptual (or sensory) consonance/ dissonance and musical consonance/dissonance (referring to a particular musical context), following Rasch and Plomp (1999). A similar distinction is made by Carol L. Krumhansl (1991). She uses the labels "tonal consonance" (related on the judgement of two simultaneously presented tones which are contextually isolated) and "musical consonance" (explicitly referring to a musical context).

Mazzola (1990) discusses some problems with the general concepts of consonance and dissonance. For instance, he points out that it is not possible to construct a unique scale that conforms to all aspects of the concepts. The description of the function of counterpoint requires another scale than the description of roughness and pleasure or enjoyment.

Let us now take the perspective of psychoacoustics and look for tonal consonance/dissonance. The classical theoretical idea (going back to Pythagoras) claims that simple frequency ratios of the intervals sound more consonant than complex frequency ratios. There are many different ways to express the intended complexity by number-theoretic functions (Gräf, 2002). The most famous of these functions is Leonhard Euler's *gradus suavitatis* (Euler, 1739).

Fig. 5 makes a comparison between Euler's gradus function and empirical data of a meta-study by Daniel L Bowling and Purves (2015). The correlation is remarkable: R^2 = .64. That means about 2/3 of the overall variance in the data are explained by this model.

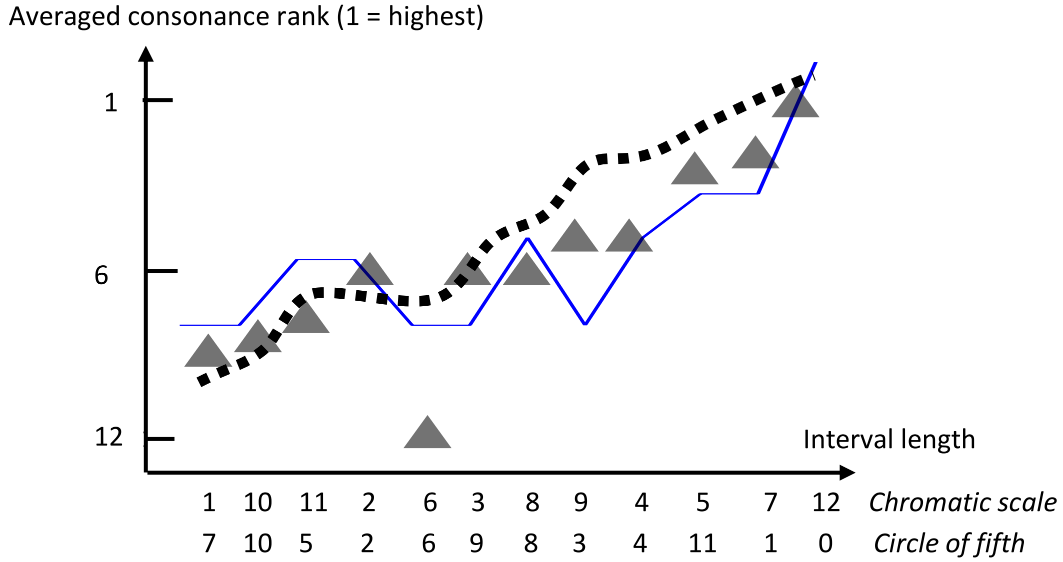

Fig. 5: Averaged consonance rank for the twelve intervals in chromatic notation. The grey triangles show the values of Euler's gradus function (scaled in order to get the value 1 for the octave and 12 for the tritonus interval). The dashed black curve shows the averaged data of a meta-study by Daniel L Bowling and Purves (2015). Finally, the solid curve gives the values calculated by the attraction model.

However, the model has systematic weaknesses. For instance, the triton interval is predicted to be the most dissonant one, which is not the case. Similarly, the major third is predicted to be more dissonant than it really is.

Parncutt (1989) points out that the major/minor distinction is central to the emotional impact of diatonic music at least since the renaissance. Ignoring other aspects such as tempo and timbre, music in major keys is preferred for expressing positive emotions (happiness, confidence, brightness), whereas music in minor keys is preferred for expressing negative emotions (sadness, grief, defeat, darkness). Even when the origins of the emotional connotation of major and minor keys are far from being clear different cognitive and perceptual factors have been listed. Among them, the distinction between (graded) consonance and dissonance seems the most important factor (cf. Parncutt 1989). Fig. 5 shows that the interval of minor third sound more dissonant than the interval of major third. The same hold for the difference between the corresponding major and minor triads (Johnson-Laird, Kang, & Leong, 2012).

Hence, for a theory of tonal meaning that accounts for the emotional impact of pure music, a careful modelling of graded consonance/dissonance is essential. Interestingly, the models I have mentioned in connection with static attraction are likewise relevant for describing graded consonance/dissonance. For instance, the template-based group of models is based on the idea that the degree of consonance increases with a clear and uniquely identifiable *root tone*. Many authors mention that major chords more clearly und undeniably define the tonic of a passage than minor chords. In contrast, minor chords tend to express a feeling of vagueness or ambiguity about the position of tonic in a minor key passage (Parncutt 1989). Unfortunately, the measures that are able to express this idea are not systematic and ad hoc constructions. For example, Hofmann-Engl (2004) proposes a measure called "sonance". The basic idea is that a chord produced a set of virtual pitches (potential root tones). The simpler this set and the higher the strength of the strongest virtual pitch compared to the other virtual pitches, the higher is the amount of sonance. His complicated formula combines (i) the absolute strength of the dominant root with (ii) the relative strength of the dominant root compared to the other roots and (iii) the number of potential roots.

In Fig. 5, the solid curve shows the predictions of a model that is based exclusively on the absolute attraction strength of the dominant root (tone with maximum attraction in the given profile). The attraction profile is calculated from Parncutt's (1988) model. The asymmetric kernel of this model has been shown in Fig. 4 (dashed line). The overall correlation between the model and the data of Bowling and Purves (2005) is not very high (R^2 = .56), i.e. only a bit better than for the hierarchical model (R^2 = .50). However, Parncutt's (1988) model predicts a difference between the major third and the minor third whereas the hierarchical model does not. The same does the model of Hofmann-Engl (2004).

Next, let us investigate the predictions of some models for the consonance ranking of common triads. The hierarchical model predicts equal consonance degrees for major and minor chords as well as for diminished and augmented triads (see Table 2:). This conflicts with empirical data, which predict a clear ordering of triads according to consonances: major triads highest (= rank 1) followed by minor triads followed by the diminished triads followed by augmented triads (lowest) (Johnson-Laird et al., 2012). Better results are provided by Parncutt's (1988) model. Unfortunately, this model ranks diminished triads as the least consonant ones in contrast with the data.

Triad Class	*Empirical Ranking*	*Hierarch. Model*	*Parncutt's 1988 Model*
major	1	.17	.18
minor	2	.17	.15
diminished	3	.11	.11
augmented	4	.11	.13

Table 2: Empirical rankings and model predictions for common triads. The predictions of the models concern the strength of strongest static attraction using normalized attraction profiles.

In Section 4.3, I will come back to the question whether the difference between major and minor modes can be reduced to different grades of consonance. Further, I will discuss the role of major and minor modes in connection with the affective impact of harmonic developments.

3.4 Underlying scales and musical similarity judgements

Pitches, and, on a more abstract level, tones, are objects of our acoustic perception. Basically, perception is based on similarity. Consequently, similarity can order tones. Geometric models simulate the perceived similarity of tones by geometric distances in a spatial model (Gärdenfors, 2000). A standard example is the *tonnetz* first proposed by Euler (1739), later adapted by Riemann, Longuet-Higgins and many others. Balzano (1980) proposed another kind of *tonnetz* grounded on group theory.

One significant problem with both kinds of tonnetz approach is that they generate absolute, context-independent distances (or similarities) between tones. However, the similarity of tones is not absolute – it is dependent of a given scale that underlies the tonal system. In Western systems of tonal music, for instance, a common scale is a diatonic scale based on a certain root tone. If C is the root tone, the diatonic scale (C major) consists of the seven tones C, D, E, F, G, A, B. Based on the C-major scale, the perceived distances between E-F on the one hand and C-D on the other hand are equal. However, this is not reflected by equal distances in the *tonnetz*. Similar observation can be based on acoustic considerations by comparing the fundament frequency quotients (assuming the octave is equally divided into twelve parts, i.e. the tuning is equal temperament). For E-F the quotient is about 16/15. In contrast, for C-D the corresponding quotient is 9/8.

Another example may illustrate a related point. Assume again that the C-major scale defines the contextual setting. Then the distances between E-F♯ on the one hand and C-D on the other hand are perceived differently even when the corresponding distances in the *tonnetz* are the same. Likewise, the frequency quotients are identically in both cases (9/8).

However, when the underlying scale is changed to G-major we get the opposite pattern even though the distances within the *tonnetz* (and the corresponding frequency quotients) are not changed. The perceived distances between E-F and C-D are different but the perceived distances between E-F♯ and C-D are identical. The situation is comparable to the situation in natural language phonology. For instance, the similarity relations between different phonemes depend on the underlying language – assuming the considered phonemes do really appear in the considered language.

An important theoretical insight of this and similar findings can be stated as follows:

> Our experience is determined by internal schemata as well as by external inputs. Thus musical tones, though physically variable along a continuum of frequency, tend to be interpreted categorically as the discrete notes (named *do, re, mi, fa, sol, la, ti, do)* of an internalized musical scale." (Shepard & Jordan, 1984).

The internal schema acts as a prototype model that maps the unequally spaced physical tones onto the discrete units of the schema, unifying the distances between the tonal neighbours. Given a selection of tones, we can ask for an underlying scale that considers these tones as tones of the scale. A concrete mechanism for selecting the underlying scale was proposed by Temperley (2007). He described a key finding process based on a Bayesian procedure.

The present implementation of the internal schema is based on a (probabilistic) kernel function and a mechanism of contextual merging (discrete convolution). This approach allows the realization of related ideas for modelling key finding. This enables the model to account for the context dependence of similarity judgement.

In the same vein, the present model solves another problem with the *tonnetz* that was first pointed out by Lerdahl (1988). The problem is that the tonnetz approach does not provide a consistent theory of similarity considering the level of tones, chords, and tonal regions. As shown in Blutner (2015), for the calculation of the similarity between tonal regions we can use a measure that is commonly used in quantum information science: the Kullback-Leibler distance (also called *relative entropy*). It is defined as follows, where p and q denote two probability distributions:

(6) $\mathrm{KL}(p/q) = \Sigma_k\, p_k \log_2 (p_k / q_k)$

Hereby, the index k ranges over all events of a given partition of the sampling space. Typically, one of the distributions represents empirical observations, the other an approximating model. Intuitively, the Kullback-Leibler distance is the expected number of bits required to code samples for *p* when using a code optimized to code samples for *q*.

Using this measure, an excellent description of similarity relations between tonal regions was found (Blutner 2015). Equally, the description of the similarity between chords as found by Huron's (2007) corpus analysis is possible. In both cases, the theoretical distributions are calculated as static attraction profiles based on discrete convolution. Notably, the convolution operation merges an abstract kernel function (based on ideas presented in the previous subsection) with the triad representing the tonal region or the chord, respectively.

The aim of the present work is not to find the model that best fits empirical data on tonal attraction, graded consonance and musical similarity. Rather, it is the elaboration of a general framework that is based on principles of parsimony, universality, and mathematical rigorosity. We are looking for a uniform, abstract concept that can represent the meaning of tonal music. We have found such a concept in an abstract kernel function (representing tonal relations) that merged with a context (key) via the operation of discrete convolution. With the help of these concepts, a fair approximation of static attraction, grades of consonance and musical similarity is possible. The next subsection argues for a pragmatic component extending this framework.

3.5 Dynamic attraction

Dynamic attraction (voice leading attraction) has some similarities with static attraction. Both concepts are graded notions and both concept are relative to a given musical context (normally established by a key). From the theoretical point of view, both concepts are modelled by using the mechanism of discrete convolution – see Eq. (5). Both concepts crucially make use of context (key or tonal scale).

In recent research, Matthew Woolhouse has proposed to explain dynamic tonal attraction in terms of interval cycles (Woolhouse, 2009, 2010, 2012; Woolhouse & Cross, 2010). The basic idea is that the (dynamic) attraction between two pitches is proportional to the number of times the interval spanned by the two pitches must be multiplied by itself to produce some whole number of octaves. Assuming twelve-tone equal temperament, the *interval-cycle proximity* (ICP) of the interval can be defined as the smallest positive number *icp* such that the product with the interval length (i.e. the number of half tone steps spanned by the interval) is a multiple of 12 (maximal interval length). The following table lists the *icp*s for all intervals spanned by a given interval length. For example, you see that the *icp* for the triton is 2 and the *icp* for the fifth is 12. This has the plausible consequence that, relative to a root tone, the fifth has higher tonal attraction than the triton. The ICP model exhibits two symmetry principles: firstly, a mirror symmetry that pairs an interval of *n* semitones with the same *icp* as an interval of 12–*n* semitones; second, a symmetry that gives the elements of a chromatic scale the same value as the numerically corresponding elements of the circle of fifth (i.e. the elements k and $7k \ mod \ 12$ get identical values).[7]

An important question concerns the theoretical status of the ICP model. What is the motivation behind a kernel function realizing ICP? According to Woolhouse and Cross (2011) the nature of ICP is based on Gestalt principles of tonal perception:

[7] This symmetry principle guarantee that the half-tone interval (1 and 11) and the fifth (7 and 5) get the same value.

> With respect to originality, while we agree that, to our knowledge, ICP is novel, its emphasis on grouping clearly places it within the ambit of gestalt psychology, a tradition that has heavily influenced music cognition research. (Woolhouse and Cross 2001, fn. 6).

Since general Gestalt principles are not learned but form part of the innate prerequisites of perception and cognition, this view would suggest, that ICP and similar organization principles are probably innate. I have serious doubts concerning this point.

Besides the mentioned similarities between the static and dynamic attraction, there are important differences. The first point concerns the experimental procedure triggering the two attraction types. Obviously, the two attraction types are connected with different kind of instructions – suggesting a simultaneous integration of all tones in the first case and the stepwise procession of adjacent tones and the idea of resolution in the second case. My second point concerns the nature of static and dynamic attraction. In my opinion, the static notion of attraction has a universal nature whereas the dynamic notion is culture-dependent. It prefers small tonal steps in Western cultures. Contrasting with these cultures, the dynamic notion seems to prefer bigger tonal steps in Indian and Chinese cultures. Hence, whereas the static conception has the character of an instinct, the dynamic conception is culture-specific and has to be learned.

There is ample evidence suggesting the universal character of static attraction. Carol L. Krumhansl (1991) reviews data for Indian and other non-Western cultures suggesting the universality of tonal hierarchies and static attraction (given the proper key). On the other side, culture-dependencies were found for dynamic attraction. For example, Carol L. Krumhansl et al. (2000) found the judgments of cultural outsiders revealed some biases of their home culture and a lack of sensitivity to some cultural conventions. In fact, they could derive melodic expectations from minimal exposure to another culture's music. Further evidence for the culture-dependency of dynamic attraction comes from a study by Curtis and Bharucha (2009).

I will demonstrate now how the ICP model can be modified in order to model the presumptive culture-dependency of dynamic attraction. Assume that Western cultures are based on a kernel function preferring small tonal steps. That means, this culture is based on the assumption that small tonal steps are preferred – for instance by assigning 12 points to a half-tone step and 6 point to a single tone step. Further, the two symmetry principles of the ICP model are assumed to be valid, and the value 0 is assigned to all elements not bound so far. The upper line of Table 2 shows the ICP kernel; the next line shows a kernel function preferring small tonal steps. This contrast with Chinese music that prefers bigger tonal steps. Assuming that the symmetry principles apply likewise, we can construct another kernel function preferring big steps. The lower line of Table 3: shows such a kernel function.

interval length	0	1	2	3	4	5	6	7	8	9	10	11
interval-cycle proximity (icp)	1	12	6	4	3	12	2	12	3	4	6	12
dynamic attraction kernel preferring small steps	0	12	6	0	0	12	0	12	0	0	6	12
dynamic attraction kernel preferring big steps	0	10	3	6	8	10	12	10	8	6	3	10

Table 3: Comparison between the kernel function of Woolhouse's (2009) ICP model, and the dynamic attraction kernel preferring small and big tonal steps

Woolhouse's ICP model gives excellent predictions when considered for dynamic attraction data. Such data were collected in a probe-tone experiment (Woolhouse, 2009). This experiment limits analysis to only the first new element after the presentation of the context chord. In the original experiments, five different context chords are considered: major triad {C, E, G}, minor triad {C, E♭, G}, dominant seventh {C, E, G, B♭}, French sixth {C, E, G♭, B♭}, or half-diminished seventh {C, E♭, G♭, B♭}. Probe tones are all twelve tones of the chromatic scale. Both the context chord and the probe tone each lasted two seconds. There was no temporal gap between context chord and probe tone. The subjects had to decide (on a 7-point Likert scale) "the level of attraction and/or resolution they felt from the chord to the probe tone: seven for a high level of attraction, one for a low level of attraction" (Woolhouse, 2009). Table 4: shows the correlation coefficients between ICP model and Woolhouse's (2009) data. It also includes a column for the correlation between the same data and the present dynamic attraction model that prefers small tonal steps.

context chord	correlation with ICP model	correlation with dynamic attraction model
{C, E, G}	0.57	0.56
{C, E♭, G}	0.79	0.78
{C, E, G, B♭}	0.75	0.76
{C, E, G♭, B♭},	0.79	0.87
{C, E♭, G♭, B♭}	0.89	0.89

Table 4: Comparison between Woolhouse's (2009) data, the ICP model, and the dynamic attraction model

Obviously, there is no real difference between the predictions of both models. Unfortunately, Woolhouse's procedure was not applied to collect data for Chinese or other non-Western musical cultures. For that reason, a direct test of the third kernel function presented in Table 3: is not possible at the moment. However, Pearce and Wiggins (2006) tested Narmour's implication realization model (Narmour, 1991, 1992) and found important differences concerning the tone-to-tone expectancies for continuations of melodies for Western melodic music and Chinese folk songs. The constraints the made the differences between the two musical styles mainly concerned the intervallic differences. Presently, it is an open issue whether these differences are reflected by the dissimilarities of the two kernel functions shown in Table 3:.

Next, let me make a sign-theoretic remark concerning the status of static and dynamic attraction. Sign theory or Semiotic was founded by the American philosophers Charles Peirce and Charles Morris (Rochberg-Halton & McMurtrey, 1983). The ideas of the two philosophers are not equivalent though related. Peirce gave the original formulation but Morris modified it in a way that became accepted by many present researchers in the field of language and beyond. From him we got the famous distinction between syntax, semantics and pragmatics:

> *Pragmatics* is that portion of semiotic which deals with the origin, uses and effects of signs within the behavior in which they occur; *semantics* deals with the signification of signs in all modes of signifying; *syntactics* deals with combinations of signs without regard for their specific significations or their relation to the behavior in which they occur. (Morris, 1946, p. 219)

This is not the place to go into a discussion of the different variants of pragmatisms and of the different ways it has been applied to modern linguistics and musicology.[8] Concerning the present distinction between static and dynamic attraction, however, it can be stated that the former conception relates to the semantics of music and the latter to its pragmatics. Musical semantics starts with the definition of the meaning of tones whereas musical pragmatics concerns the use of tones in constructing sequences of tones with certain event hierarchies and varying degrees of expectation for subsequent tones.

It goes without mentioning the enormous literature that goes beyond Woolhouse's approach and includes, inter alias, Bayesian probabilistic modelling (Temperley, 2008). A careful discussion of such alternative approaches and a detailed comparison has to wait for another occasion.

4. The arousal-valence model of emotions and how the bare-bone semantics describes the elicited emotions

The last section has suggested a semantic approach to tonal music based on a semantic kernel and a contextual element (key). With the help of a particular semantic merging mechanism (discrete convolution), it has been achieved to approximate simple static attraction profiles and graded values of tonal consonance. Based on the same assumptions it is likewise possible to account for musical similarity judgements. Even when the fit with the empirical data was not thoroughgoing, the model can be seen as a solid fundament of the present bare-bone semantics. Further, I have demonstrated that the modelling of dynamic attraction requires musical pragmatics. Technically, it is defined by a culture-dependent kernel function that has to be learned. In this section, I will demonstrate how the proposed bare-bone semantics (extended by musical pragmatics) can account for the elicited emot-ions.

4.1 The arousal-valence model of emotions

The arousal–valence theory or circumplex model has been developed by James Russell (Russell, 1979, 1997, 2003). There are forerunners of this model, dominantly developed in the field of linguistic semantics. For example, Osgood developed a behavioural conception for measuring the affective meaning of words (Charles E. Osgood, 1952). In a related book, he and his colleagues classified the emotional meaning of words along three dimensions: valence, arousal and dominance (Charles Egerton Osgood, Suci, & Tannenbaum, 1964). Russell's model can be seen as a simplification of this model. It treats all emotions as if they fall into a two-dimensional representation defined by (i) valence (positive to negative) and (ii) arousal (very awake to asleep). Within the arousal–valence model, appetitive and aversive experiences occur at opposite ends of the same bi-polar dimension. That means the experience of positive affect excludes the experience of negative affect. This contrast with other studies, that explicitly allow multi-valent emotional states (cf. Bradley W Vines,

[8] Note that Peirce followed scholastics and proposed a three-part division of semiotic into speculative grammar, logic proper, and speculative (= theoretic) rhetoric. Morris (accompanied by Carnap), promoted the distinction of linguistics into syntactics, semantics and pragmatics. Later, Grice (1967, 1989) continued the work of Peirce and Morris.

Krumhansl, Wanderley, Dalca, & Levitin, 2011). As pointed out in the article by Vines et al. (2011), several studies support the arousal–valence model (Faith & Thayer, 2001; Green, Goldman, & Salovey, 1993; Green, Salovey, & Truax, 1999). Moreover, other studies employ the arousal–valence model to study real-time emotional responses to musical stimuli (e.g., Schubert, 1999).

Fig. 6 gives a simplified illustration of the arousal-valence model following Russell (1979).

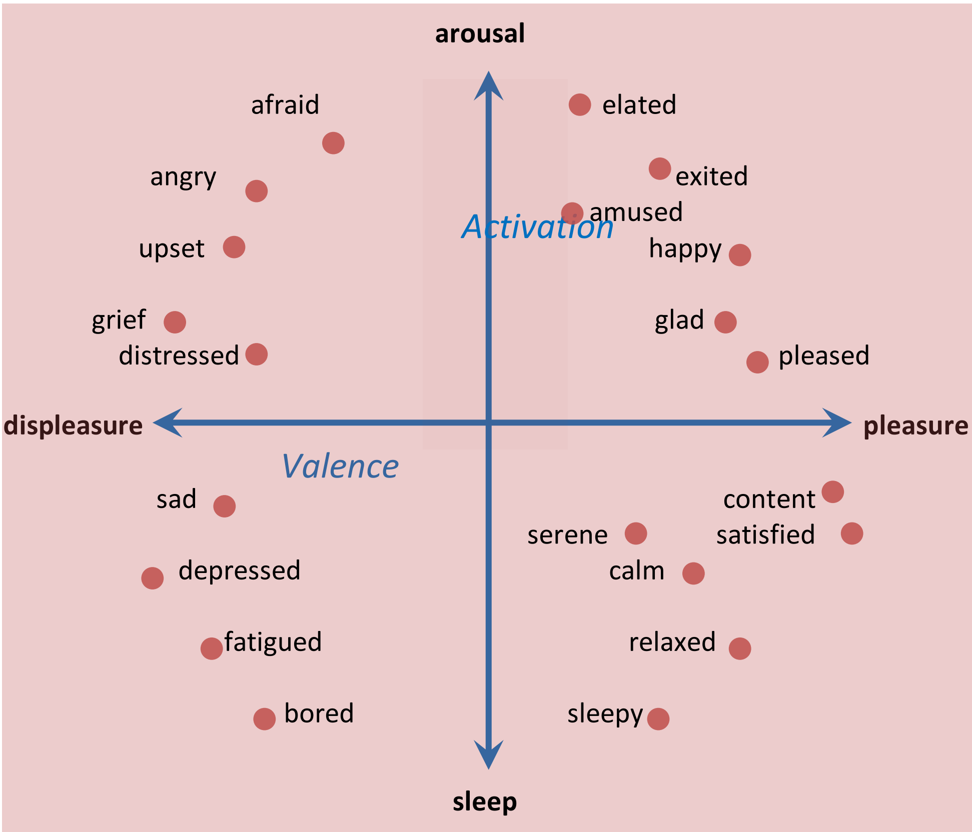

Fig. 6 : Pleasure-arousal judgment space for affective feelings according to James Russell. This model of core emotions has two dimensions: valence (pleasure–displeasure) and activation (arousal–sleep)

In a certain approximation, this picture describes what philosophers call the qualia – the subjective experiences that accompany sensory experiences. Accepting this picture as a fair description of the possible qualia (also called emotions) in the domain of tonal music, a theory of evoking emotions will be developed in the following subsection. It achieves to correlate musical forms of pitch events with the appropriate region in the emotional space.

4.2 Bare-bone semantics and the emotional response

In probabilistic frameworks, entropy is perhaps the most basic probabilistic concepts. For any probability distribution p_k the entropy $E(p)$ is a non-negative real number defined as

(7) $E(p) = -\Sigma_k \, p_k \log_2 (p_k)$

Informally, entropy can be described as the amount of uncertainty, degree of surprise, or information that is connected with a source generating the probability distribution p_k. A related measure is *redundancy*, which is the relative distance between the maximum entropy and the observed entropy. This is always a number between 0 and 1. The definition is as follows:

(8) $Rdd(p) = 1 - E(p)/\log(n)$, where n is the number of elementary events

If the distribution p_k describes a source with maximum entropy, then the redundancy of the source is 0. With decreasing uncertainty, the redundancy is rising (up to its maximum value of 1).

Temperley (2007) describes the application of this concepts in musicology and identifies Youngblood (1958) as the pioneers of applying entropy. Youngblood investigated small corpora of Romantic composers and made an analysis of their styles by calculating the entropy of the key-related pitch classes (scale degrees) of a piece. In the present context, I apply the analysis of entropy to the distribution provided by the (normalized) dynamic attraction profile (relatively to a key). In this way, I propose to measure uncertainty/confidence concerning future musical elements during the melodic and harmonic development of a piece of music – following the basic idea of Meyer (1956).

Following common wisdom, I propose to identify the dimension of valence (pleasure/displeasure) with the dimension of harmonicity (consonance/dissonance). For simplicity, let us define valence as the attraction strength of the root tone (based on the normalized, static attraction profile). The second dimension of activation (arousal/sleep) will be identified with the measure of redundancy (cf. formula (8)). The higher the redundancy value, the bigger the confidence in the expected event (dynamic attraction profile). Hence, in order to estimate the emotional response of a piece of music it needs two calculation: (i) from the semantic kernel and the context (key) we can determine the static attraction potential. It yields the dimension of valence; (ii) from the pragmatic kernel plus context we can calculate the dynamic attraction profile. It gives the dimension of activation. In this way, musical signals can be paired with points in the pleasure-arousal space. This points indicate the elicited emotions.

4.3 The major-minor distinction and its emotional content

Almost 90 years ago, the striking contrast in the psychological effects of major vs. minor modes earned immense attention (Heinlein, 1928). Helmholtz (1863) saw one difference in the dissimilar degrees of consonance/dissonance connected with major and minor triads. Helmholtz explained this difference by diverging effects connected with the harmonic series generated by major and minor triads. Others see the differences grounded in principles of fusion (Stumpf, 1883, 1890). Obviously, both major and minor chords consist of a minor and a major third. The only difference between the two modes is that in minor triads the minor third has a deeper position than in major triads. Consequently, the fusion mechanism is assumed to operate differently in both cases (Pear, 1911). Heinlein (1928) and others, however, have demonstrated that a reversal of the constituent intervals from lower to higher position is not sufficient to explain the differences which tend to exist between the two triads. I think it is correct that the minor triads are perceived as more dissonant than the major triads. A simple explanation of this observation based on Parncutt's (1988) model has been given in Section 3.3. It should be noted that some authors prefer an alternative explanation based on musical pragmatics. This explanation is grounded on the predominance of the major mode in the Western culture and effects of statistical learning (Huron, 2006). As consequence, "the major mode appears to be a 'default' listening mode for Western-enculturated listeners" (Ladinig & Huron, 2013, p. 117).

The distinction made between major and minor modes is normally based on one dimension only. In Section 3.3, the dimension of consonance/dissonance was identified and different models were proposed to capture this dimension quantitatively – for instance, template-based models (Parncutt 1988, Hofmann-Engl 2004). I suggest here a more differentiated picture by taking the arousal-valence model as background. In this case, two dimensions matter – one related to the classical consonance-dissonance distinction, the other related to the physiological dimension of arousal. In this connection, Huron (2012) gives an instructive example: the difference between sadness and grief.

> Sadness is an affective state characterized by low physiological arousal. When sad, a person typically exhibits slow heart rate, shallow respiration, slumped posture, loss of appetite, sleep, reduced engagement with the world, a tendency to avoid conversation (i.e., mute), and rumination (thinking sad thoughts). Grief, by contrast, is an affective state characterized by *high* physiological arousal. When in a state of grief, a person typically exhibits fast heart rate, erratic respiration, flushed face, tears, nasal congestion … (Huron 2012, p. 476)

In the pleasure-arousal judgment space of Fig. 6, sadness and grief are localized both on the displeasure side, however with different levels of arousal: a higher level for grief and a lower level for sadness. In music, both emotions are typically connected with minor triads. However, the sadness-type arises in contexts where the listener clearly expects a certain chord whereas the grief-type arises in contexts with a rather vague expectation pattern. Examples are Chopin's *marche funebre* (B-flat minor Sonata, op. 35) with definite and confident expectations of the minor chords. This recruits the sadness type typical for funeral music. On the other hand, the *Adagio, ma non troppo* of Beethoven's op. 110 contains many segments where the expectations for minor (or major) chords are rather unclear. Consequently, the appearance of the minor chord raises the grief-type of displeasure.[9]

4.4 Anticipation, suspension and ostinato

In this subsection, I will reconstruct some examples Huron gave in his book (Huron 2006) by applying the proposed bare-bone semantics. In Section 2.3, I have concluded that it is the tension response that evaluates emotion on the axis of activation (with different levels of arousal). On the other hand, we have the cumulative outcome of prediction and reaction response. Their collective outcome evaluates the dimension of valence (from pleasure to displeasure).

Anticipation

As an illustrative example of the emotional effects evoked by certain pitch event, let us consider the case of prototypical anticipation as shown in Fig. 7. The numbers identify three moments that we will analyse separately. The moments are (1) pre-anticipation, (2) anticipation, and (3) post-anticipation (Huron 2006, p. 306). The difference between the two variants shown in Fig. 7 is that a small rhythm-ic delay of the anticipation step (2) is realized on the right hand side but not on the left hand side.

[9] Of course, there are many more dimensions that are appropriate to signal grief or sadness including timbre, tempo, breaking voice, erratic breathing etc. (see Huron 2006). Further, knowing the story of Beethoven's "unsterbliche Geliebte", which the Beethoven sonata seems to address, may add an additional, extrinsic framework for generating the affective state of grief.

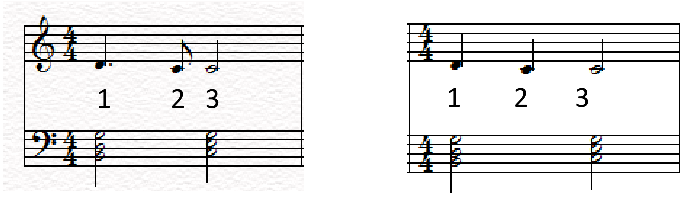

Fig. 7 : Prototypical example of anticipation (V-I cadence assuming an established C-major key). On the left hand side, the anticipation tone refers to a quarter note falling on beat 2 whereas on the right hand side the anticipation tone refers to an 1/8 note which is delayed by 1/8.

In the following, I discuss the route in the emotion space realized by the proposed mapping from tonal events to the affective responses. I give a qualitative explanation even when the proposed model is explicit enough to make quantitative prediction.

(1) We assume the first chord appears. It is perceived as dominant chord (V) since we assume that the tonic chord (I) has been established before. At the beginning of a new piece, the tension response is important. In the present case, the V chord has a low probability to be followed by silence (in order to closure the piece) and a high probability of leading to the tonic. That means there is little of the stress that comes with uncertainty. Hence, the tension response comes with a relatively neutral value on the arousal scale. Further, the sonorous element (dominant chord V) is likely to evoke a positive valence on the pleasure scale.

(2) We consider now the moment when the anticipation tone appears. Reaction response: With this tone, the sonority becomes more dissonant. That is, the reaction response has a comparatively negative valence on the pleasure scale. Prediction response: Since the previous moment (1) lead the listener to make a prediction, we can now consider the successfulness of this prediction. It is positively valenced. This positive value on the pleasure scale has to be added to the negative vale of the reaction response. Hence, the resulting valence at moment (2) is less negative than without the prediction response. Now consider the difference between the two variants of anticipation shown in Fig. 7. On the left hand side, the anticipation tone falls on a more predictable beat than on the right hand side where it is delayed by an eight. Consequently, the prediction response (concerning the "when") is stronger in the first case. This results in more pleasure in case of predictable beats than in case of delayed beats. Tension response: Compared with the pre-anticipation sonority, the anticipation note significantly reduced the uncertainty for the events predicted for moment (3). The reduction of uncertainty is higher in cases where the anticipation tone is delayed. That is, the presence of the eighth note significantly reduces the uncertainty as to whether the tonic chord will appear at beat three, or wait until the next measure.

(3) Reaction response: The reaction response is highly positive: the chord has high sensory consonance. Prediction response: The listener's confident prediction of this moment is realized, and so there is a high positively valenced prediction response. Taken together, a high degree of pleasure is predicted. Tension response: The closure associated with this moment creates a highly certain expectation that the current moment will be sustained for two or more beats, and perhaps followed by silence. Hence, uncertainty is reduced further and this leads to a lower level of arousal.

Fig. 8 shows the hypothetical change of the affective feelings in the two-dimensional emotional space. The picture also illustrates the "higher dynamics" that is involved when the anticipation tone is delayed a bit.

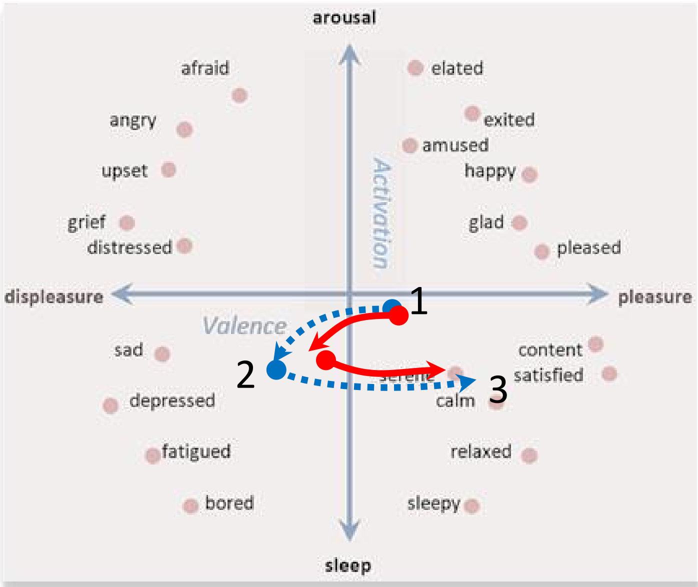

Fig. 8 : Two routes in the emotional space. Solid line: induced by the cadence given on the left hand side of Fig. 7; dashed line: induced by the cadence given on the left hand side of in Fig. 7.

A careful explanation of the difference between the two variants shown in Fig. 7 requires a modelling of the interaction between rhythm and harmony/voice leading. This interaction is presently not realized in the proposed bare-bone semantics.

Suspension

Next, consider the case of prototypical suspension. Assume that via a key the tonic chord is established (step 1). It follows a dominant chord with the tonic pitch suspended (step 2). Finally, the suspended pitch resolves downward by step (step 3). Resolving the suspended tone is not the end of the matter. In step 4, we find an ensuing tonic resolution (see Huron 2006, p. 309 ff.)

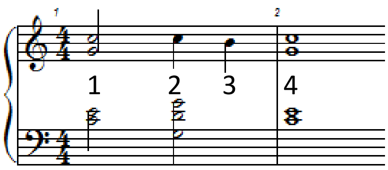

Fig. 9 : Example of prototypical suspension

The following discussion considers the four moments shown in Fig. 9.

(1) Assume again that the first chord appears in an established key context and it is perceived as a tonic chord. As a tonic chord, it is quite stable and there is little feeling of anticipation. Hence, there is less arousal and the chord has a positive sonority.

(2) Consider now the moment when the suspended sonority appears. The sonority is now more dissonant. So, the reactions response has a comparatively negative value on the pleasure scale. There is a low tension response since the suspended pitch creates a high expectation that the melody will move to note B. Low uncertainty will induce a low amount of arousal and vigilance.

(3) At step 3, we will find a large positive prediction response since our expectations from step 2 are fulfilled. Likewise, the reaction response will be positive since the formerly dissonant sonority (step 2) has been replaces by a more consonant sonority in step 3. Further, there is a relatively low tension response.

(4) This step realizes the tonic resolution. Both the reaction and the prediction response are positive. Further, the closure associated with this moment creates a very low tension response.

Fig. 10 shows the tentative route in the emotional space induced by the suspension cadence shown in Fig. 9.

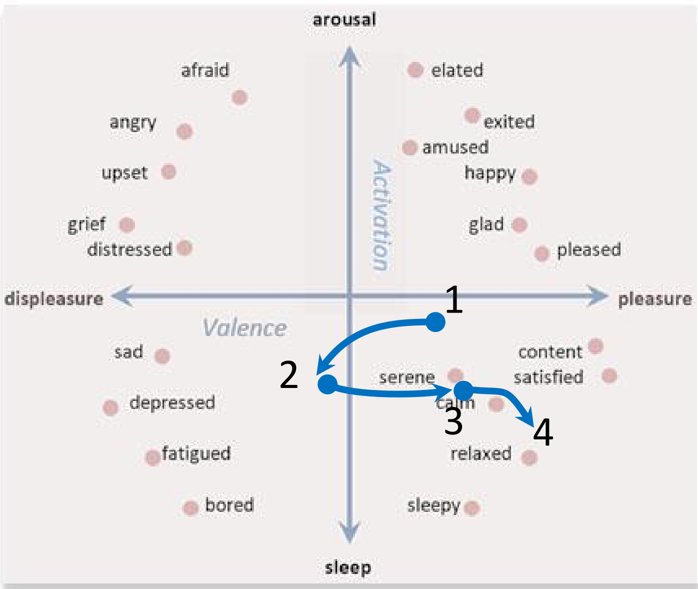

Fig. 10 : Route in the emotional space considering the suspension cadence of Fig. 9

Harmonic ostinato

The Italian word 'Ostinato' denotes the idea of "persistence". In music, it refers to any melodic, rhythmic or chordal phrase that is repeated continuously through a section or through the whole piece. Passacaglias (or chaconnes) are based on ostinatos. Minimalism is sometimes described as "the triumph of the ostinato". The emotional effect of such repetitive, circular pattern created by any ostinato figure is not to define uniquely. It can be tension; it can be relaxation. Essentially, the persistence of the repetitions (harmonic, melodic and/or rhythmic) can strengthen the emotional effects.

Ostinate repetitions, harmonic, melodic and rhythmic, trigger a pathway through the affective space. However, repetition of a musical phrase does not lead to repeatedly running the same route. Instead, we suppose that permanent repetition (that are open to modulations and modifications of rhythm) can induce reinforcement effects. In this way, we can possibly reach extreme points conforming to chills, tears, up to thrills and ecstasies (Patrik N. Juslin & Sloboda, 2013; Mori &

Iwanaga, 2017). Obstinate repetitions, run throughout Ravel's *Bolero*, Shostakovich's *Passacalia* of his violin concerto or Bach's *Toccatas*, are able to generate huge physical excitement.

Summarizing, I have considered some simple examples that demonstrate how tonal pieces can elicit emotional effects and aesthetic emotions. The latter are the conscious outcomes of the former. In the present simple model, aesthetic emotions are represented within a two-dimensional emotional space (Russel 1979, 2003) by using the axes activation and valence. Simplifying Huron' (2006) model, I assume that the tension response determines the quantitative value of the activation axis and the cumulative outcome of prediction and reaction response determines valence. On the on hand, the present model is a simplification of Huron (2006); on the other hand, it is more explicit than Huron's model and allows quantitative descriptions based on the calculation of graded consonance and arousal (entropy). This might be important for a future empirical verification of the model.

5. Towards an overall picture: Some preliminary conclusions

This article has been devoted to the relation between music and affective feelings. It contains an explicit proposal for a bare-bone semantics of *pure* or *absolute* music. Making the distinction between extrinsic versus intrinsic meanings, I have stressed the importance for starting with an intrinsic semantics. An intrinsic semantics ignores factors that are outside the domain of music.

The basic idea of the proposed intrinsic semantics is that musical meaning and emotion depend on how the actual events in the music play against the background of (internal) musical expectations (Meyer 1956). Of course, this idea does not exclude or even forbid the development of an extrinsic semantics – accounting for an external programme of music and for a combination of music and language/dance. However, such a programme should likewise base on an intrinsic semantics. I think it cannot work the other way around by starting with an extrinsic semantics and augmenting it by intrinsic factors (as suggested by Schlenker 2016; 2017).

Following the lead of Meyer (1956), both the qualitative character of chords (consonant/ dissonant) and the dynamic generation of expectations are relevant (in tandem with rhythmical structures). This paper introduces novel ideas concerning the mapping of musical structures onto a two-dimensional emotional space. The induction of emotion is of primary importance for understanding music.

My proposal for a bare-bone semantics of pure music is grounded on an abstract meaning component. It consists of a musical kernel expressing the relational meaning of tones (static attraction). The kernel merges with the context (key, previous chords). The abstract meaning component is sufficient to account for several components of the structural grid of tonal music including static attraction, grades of consonance/dissonance, underlying scales and their similarities.

Further, I have proposed to combine static semantics with musical pragmatics in order to describe how music elicits emotional responses and aesthetic feelings in a rather direct way. Accepting the basic idea of the *doctrine of affections* (German: *Affektenlehre*), the present intrinsic account explicates a direct mapping between music and the elicited emotions. Aesthetic emotions are the conscious effect of the mapping – arising from a series of affective responses that accumulate over time. In the terminology of Peirce (1991), this account corresponds to the idea of indexicality that gives a causal explanation for the mapping between musical signs and their affective denotates. For an interesting discussion of this issue the reader is referred to Poller (2005).

Notably, the terms semantics and pragmatics suggest some similarity with the linguistic enterprise. I have stressed already that these similarities cannot be formulated in terms of intentionality, conventionality or communicative function. Music is – in contrast to language – not a communication system recognizing and processing intentions. Music shows what it means. It does not say it. Rather, the similarities between music and language result from (i) similar cognitive functions such as the common merging operation (discrete convolution), (ii) similar learning mechanisms based on universal, innate prerequisites, and (iii) similar mechanisms of contextual dependencies and contextual enrichment.[10]

A permanent idea of this writing is to apply the methodology of natural sciences to the field of musicology. This includes the idea of science as *solving riddles*, given an appropriate scientific paradigm (Kuhn, 1996). Section 3 of this article is mainly devoted to this topic. The riddles are numerous. They concern the best ways of how to model static attraction, degrees of consonance, musical similarity judgements, and dynamic attraction (melodic development). It goes without mentioning that the present approach – based on audacious-oversimplification – did not solve any riddle completely but most riddles partially. This includes the discussion of consonance in connection with major and minor mode. Presently, it is debatable whether an improvement of the kernel function itself or an improvement of the evaluation mechanism based on it is the critical issue. Further, it would be very helpful if an integration of alternative approaches (Ebeling, 2008; Milne et al., 2015; Stolzenburg, 2015) could be mastered. Another topic that has been widely ignored so far is the integration of rhythm and melody/harmony.

The issue of inborn or learnt parts of musical similarity, consonance, and tonal attraction is a hot problem for cognitive musicology. Recent work by Daniel L. Bowling, Hoeschele, Gill, and Fitch (2017) and McDermott, Schultz, Undurraga, and Godoy (2016) shows that the issue is far from being solved. The debate also shows that such deep questions can be solved in the context of a powerful theory only. Data alone cannot answer the question which phenomena of music cognition are dominantly explained by learning and which are best explained by psychological nativism. The present account suggest taking the semantic kernel as innate. This contrasts with the pragmatic kernel that can be assumed being learnt. Other learnt mechanisms relate to the variance of the contextual elements (key selection). Further, there are good arguments to assume that the merging mechanism (discrete convolution) is not learnt but innate, since it appears in many independent cognitive domains (including language and perception).

Finally, let me make a remark concerning recent approaches of applying the field of quantum cognition to cognitive musicology (Blutner, 2015, 2016; beim Graben & Blutner, 2017, 2018). Based on the lead of Guerino Mazzola (Mazzola, 1990, 2002), who was the first to see the analogy between physics and music in connection with the existence of symmetries and musical forces, this approach develops an account of tonal attraction in particular and of musical forces in general. This account sharply contrasts with authors who see musical forces as conceptual metaphors in the sense of

[10] Bierwisch (1978, p. 52 ff.) also introduces an abstract meaning function of musical signals which he calls "gestic form" (*gestische Form*). Contrasting with natural language utterances (especially sentences) that express propositions and say something about the world, musical utterances express gestures and thus show emotional pattern. A formal account of ideas similar to those of Bierwisch was recently given by Mannone (2018). Gestic forms describe the rhythmic and melodic components of music and ignore the harmonic part. The present bare-bone semantics is complementary to these accounts by considering the harmonic part and ignoring the rhythmic and melodic part.

Lakoff and Johnson (1980). These authors structure our musical thinking per analogy with falling, inert and attracting physical bodies (Steve Larson, 1997-98, 2004; Steve Larson, 2012). It should be mentioned that the quantum approach gives a deeper justification of many assumptions made in the present article. Further, it realizes the almost 400 years old dream of unifying parts of musicology with parts of physics (Kepler, 1619). An important methodological insight of this approach has to do with the idea of symmetry breaking. The modelling of a class of phenomena starts with the assumption of as many symmetries as possible with the stepwise breaking of particular symmetries (Brading & Castellani, 2013). However, a breaking of the initial symmetry cannot happen without a reason. In this way, particular causes can be created, for instance in terms of physical forces. Obviously, this story of symmetry breakings for developing powerful explanatory theories is a fresh application of the idea of abstractions and idealizations. In the present case of music semantics, it could encourage new ways for solving some of the puzzles we have left open so far, for instance the modelling of graded consonance for triadic chords, a more advanced modelling of voice leading, and ways to go beyond octave equivalence.

Acknowledgement

I would like to thank the Barenboim-Said Academy and their members Pamela Rosenberg, Linda Reisch, Michael Naumann, and Carsten Siebert for organizing a series of lectures about the functioning of music (since 2015). The presentation of Philippe Schlenker about the semantics of music on June 12, 2017 was the intellectual and emotional trigger for the present writing. I thank Philippe for sending me several useful comments relating to issues raised in the present paper. Further thanks are devoted to Manfred Bierwisch and Peter beim Graben for a productive cooperation and helpful comments on this article. My son Stefan is acknowledged for a critical reading of the whole manuscript.

References

Ball, P. (2010). *The Music Instinct: How music works and why we can't do without it*. London: Bodley Head.

Balzano, G. J. (1980). The group-theoretic description of 12-fold and microtonal pitch systems. *Computer Music Journal, 4*(4), 66-84.

Bierwisch, M. (1978). Musik und Sprache. In E. Klemm (Ed.), *Jahrbuch Peters 1978. Aufsätze zur Musik* (pp. 9-100). Leipzig: Edition Peters.

Blutner, R. (2010). *Geometrical models of meaning and compositionality*. Paper presented at the Approaches to the Lexicon, Invited talk, Copenhagen Business School. Copenhagen. Available from https://conference.cbs.dk/index.php/lexicon/lexicon/paper/viewFile/867/484.

Blutner, R. (2015). Modelling tonal attraction: Tonal hierarchies, interval cycles, and quantum probabilities. DOI 10.1007/s00500-015-1801-7. In M. L. Dalla Chiara, R. Giuntini, E. Negri & S. Smets (Eds.), *Logic, Music and Quantum Information*. Berlin: Springer.

Blutner, R. (2016). *Musical Forces and Quantum Probabilities*. ILLC, University of Amsterdam. Available from https://www.researchgate.net/publication/301289738_Musical_Forces_and_Quantum_Probabilities. Amsterdam.

Bowling, D. L., Hoeschele, M., Gill, K. Z., & Fitch, W. T. (2017). The nature and nurture of musical consonance. *Music Perception, 35*(1), 118–121.

Bowling, D. L., & Purves, D. (2015). A biological rationale for musical consonance. *Proceedings of the National Academy of Sciences, 112*(36), 11155-11160.

Brading, K., & Castellani, E. (2013). Symmetry and Symmetry Breaking. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy Stanford: Stanford University. Retrieved from https://plato.stanford.edu/archives/spr2013/entries/symmetry-breaking/.

Broeckx, J. L. (1981). *Muziek, Ratio en Affect*. Antwerpen: Metropolis.

Broeckx, J. L. (1997). Beyond metaphor: Musical figures or "Gestalte" as expressive icons. *Journal of New Music Research, 26*(3), 266-276.
Crickmore, L. (2017). The measurement of aesthetic emotion in music. *Frontiers in Psychology, 8*.
Cuddy, L. L., & Lunney, C. A. (1995). Expectancies generated by melodic intervals: Perceptual judgments of melodic continuity. *Perception & Psychophysics, 57*(4), 451-462.
Curtis, M. E., & Bharucha, J. J. (2009). Memory and musical expectation for tones in cultural context. *Music Perception, 26*(4), 365-375.
Dahlhaus, C. (1994). *Die Idee der absoluten Musik*. Kassel: Bärenreiter.
de Groot, M. L. (2013). *Two computational methods of attributive modification in natural language semantics compared.* (Bachelor Thesis, https://esc.fnwi.uva.nl/thesis/centraal/files/f218739155.pdf), Universiteit van Amsterdam, Amsterdam.
Desain, P., & Honing, H. (1996). *Physical motion as a metaphor for timing in music: the final ritard.* Paper presented at the Proceedings of the International Computer Music Conference.
Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88*, 503-522.
Ebeling, M. (2008). Neuronal periodicity detection as a basis for the perception of consonance: A mathematical model of tonal fusion. *Journal of the Acoustical Society of America, 124*(4), 2320–2329.
Erdmann, K. O. (1901). *Die Bedeutung des Wortes*. Leipzig: Haessel.
Euler, L. (1739). *Tentamen novae theoriae musicae*. St. Petersburg. Stuttgart 1926: Teubner.
Faith, M. L., & Thayer, J. F. (2001). A dynamical systems interpretation of a dimensional model of emotion. *Scandinavian Journal of Psychology, 42*(2), 121-133.
Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. Cambridge, Mass.: The MIT Press.
Gojowy, D. (1983). *Schostakovitsch*. Reinbek bei Hamburg: Rowohlt Taschenbuch Verlag.
beim Graben, P., & Blutner, R. (2017). Toward a Gauge Theory of Musical Forces. In J. A. de Barros, B. Coecke & E. Pothos (Eds.), *Quantum Interaction. 10th International Conference (QI 2016), LNCS 10106* (pp. 99-111). Berlin, Heidelberg, New York: Springer.
beim Graben, P., & Blutner, R. (2018). Quantum approaches to music cognition. Submitted to the Journal of Mathematical Psychology. Retrieved from https://arxiv.org/abs/1712.07417 website:
Gräf, A. (2002). Musical scale rationalization – a graph-theoretic approach. Mainz: Johannes Gutenberg-University Mainz.
Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of personality and Social Psychology, 64*(6), 1029-1041.
Green, D. P., Salovey, P., & Truax, K. M. (1999). Static, dynamic, and causative bipolarity of affect. *Journal of personality and Social Psychology, 76*(5), 856-867.
Grice, P. (1967). *Logic and conversation. William James Lectures, reprinted in Grice 1989*.
Grice, P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, 3: Speech Acts* (pp. 41-58). New York: Academic Press.
Grice, P. (1989). *Studies in the way of words*. Cambridge Mass.: Harvard University Press.
Hanslick, E. (1891). *The beautiful in music*. London: Novello, Ewer & Co.
Heinlein, C. P. (1928). The affective characters of the major and minor modes in music. *Journal of comparative Psychology, 8*(2), 101-142.
Helmholtz, H. v. (1863). *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Friedrich Vieweg und Sohn.
Hindemith, P. (1959). *Komponist in seiner Welt*: Schott.
Hofmann-Engl, L. J. (2004). Virtual pitch and its application to contemporary harmonic analysis. from Chameleon Group Online Publication. Available from: http://www.chameleongroup.org.uk/research/virtual_analysis.html
Honing, H. (2003). The final ritard: On music, motion, and kinematic models. *Computer Music Journal, 27*(3), 66-72.
Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, Mass.: The MIT Press.
Huron, D. (2012). *Understanding music-related emotion: lessons from ethology*. Paper presented at the 12th Int. Conf. on Music Perception and Cognition, Thessaloniki, Greece.
Johnson-Laird, P. N., Kang, O. E., & Leong, Y. C. (2012). On musical dissonance. *Music Perception, 30*(1), 19-35.
Juslin, P. N., Liljeström, S., Västfjäll, D., & Lundqvist, L.-O. (2010). How does music evoke emotions? Exploring the underlying mechanisms. In P. N. Juslin & J. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 605-642). Oxford; UK: Oxford University Press.

Juslin, P. N., & Sloboda, J. (2013). Music and Emotion. In D. Deutsch (Ed.), *The psychology of music. Third Edition* (pp. 583-645). Amsterdam etc.: Elsevier.
Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31*(5), 559-575.
Kalra, P., & Peleg, S. (Eds.). (2006). *Computer Vision, Graphics and Image Processing*. Berlin, Heidelberg: Springer.
Kania, A. (2016). The philosophy of music. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy (Winter 2016)*. Retrieved from https://plato.stanford.edu/archives/win2016/entries/music.
Kant, I. (1790). *Kritik der Urteilskraft*. Berlin und Libau: Lagard und Friedrich.
Katz, J. (1972). *Semantic theory*. New York: Harper & Row.
Katz, J., & Fodor, J. A. (1963). The structure of semantic theory. *Language, 39*, 170-210.
Kepler, J. (1619). *Harmonices mundi*. Linz (Austria): Edited by Gottfried Tampach, printed by Johannes Plank.
Klein, J., & Jacobsen, T. (2012). *Music is not a Language. Available from http://www.artistic-research.de/publikationen*. Institut für künstlerische Forschung. Berlin.
Kostka, S., & Payne, D. (1995). *Tonal Harmony*. New York: McGraw-Hill.
Krumhansl, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual Review of Psychology, 42*(1), 277-303.
Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum, 17*, 53-80.
Krumhansl, C. L. (1997). An Exploratory Study of Musical Emotions and Psychophysiology. *Canadian Journal of Experimental Psychology, 51*(4), 336-352.
Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the Dynamic Changes in Perceived Tonal Organization in a Spatial Representation of Musical Keys. *Psychological Review of General Psychology, 89*, 334-368.
Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition, 76*(1), 13-58.
Kuhn, T. S. (1996). *The structure of scientific revolutions*. Chicago: University of Chicago press.
Ladinig, O., & Huron, D. (2013). Minor mode cuing: Do composers signal minor mode sooner than major mode?." 23.2 (2013): 117. *Psychomusicology: Music, Mind, and Brain, 23*(2), 117-122.
Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
Large, E. W. (2010). Neurodynamics of music. In M. R. Jones, R. R. Fay & A. N. Popper (Eds.), *Music Perception* (pp. 201-231). New York: Springer.
Large, E. W., Kim, J. C., Flaig, N. K., Bharucha, J. J., & Krumhansl, C. L. (2016). A neurodynamic account of musical tonality. *Music Perception: An Interdisciplinary Journal, 33*(3), 319-331.
Larson, S. (1997-98). Musical Forces and Melodic Patterns. *Theory and Practice, 22/23*(8), 55-71.
Larson, S. (2004). Musical Forces and Melodic Expectations: Comparing Computer Models and Experimental Results. *Music Perception, 21*(4), 457-498.
Larson, S. (2012). *Musical Forces: Motion, Metaphor, and Meaning in Music*: Indiana University Press.
Leman, M. (2012). *Music and schema theory: Cognitive foundations of systematic musicology* (Vol. 31). Berlin, Heidelberg: Springer Science & Business Media.
Lerdahl, F. (1988). Tonal pitch space. *Music Perception, 5*(3), 315-350.
Lerdahl, F. (2001). *Tonal Pitch Space*. NewYork: Oxford University Press.
Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, Mass.: The MIT Press.
Lerdahl, F., & Krumhansl, C. L. (2007). Modeling Tonal Tension. *Music Perception, 24*(4), 329-366.
Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences, 8*(4), 529-539.
Lidov, D. (2005). *Is language a music? Writings on musical form and signification*. Bloomington: Indiana University Press.
London, J. (2008). Third Party Uses of Music and Musical Pragmatics. *The Journal of Aesthetics and Art Criticism, 66*(3), 253-264.
Mannone, M. (2018). Knots, music, and DNA. *Journal of Creative Music Systems, 2*(2), 1-23.
Mazzola, G. (1990). *Geometrie der Töne*. Basel: Birkhäuser.
Mazzola, G. (2002). *The topos of music: geometric logic of concepts, theory, and performance*. Basel: Birkhauser.
McDermott, J. H., Schultz, A., Undurraga, E., & Godoy, R. (2016). Consonance preferences are not universal: Indifference to dissonance among native Amazonians. *The Journal of the Acoustical Society of America, 139*(4).
Meyer, L. B. (1956). *Emotions and Meaning in Music*. Chicago: Chicago University Press.

Michaelis, C. F. (1795). *Ueber den Geist der Tonkunst: Mit Rücksicht auf Kants Kritik der aesthetischen Urtheilskraft. Ein ästhetischer Versuch*. Leipzig: In der Schäferischen Buchhandlung.

Milne, A. J., Laney, R., & Sharp, D. B. (2015). A spectral pitch class model of the probe tone data and scalic tonality. *Music Perception, 32*(4), 364-393.

Mitchell, J., & Lapata, M. (2010). Composition in distributional models of semantics. *Cognitive Science, 34*(8), 1388-1429.

Mohr, G. (2007). Die Musik ist eine Kunst des innern Sinnes und der Einbildungskraft Affekt, Form und Reflexion bei Christian Friedrich Michaelis. In U. Tadday (Ed.), *Musik-konzepte: Neue Folge. Die Reihe über Komponisten. Sonderband Musikphilosophie* (pp. 137-151). München edition text+kritik.

Montague, R. (1970). Universal grammar. *Theoria, 36*, 373-398.

Mori, K., & Iwanaga, M. (2017). Two types of peak emotional responses to music: The psychophysiology of chills and tears. *Scientific Reports, 7*, 1-13.

Morris, C. W. (1946). *Signs, Language, and Behavior*. New York: Braziller.

Narmour, E. (1991). The top-down and bottom-up systems of musical implication: Building on Meyer's theory of emotional syntax. *Music Perception, 9*, 1-26.

Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model*. Chicago: University of Chicago Press.

Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological bulletin, 49*(3), 197-237.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1964). *The measurement of meaning*. Urbana: University of Illinois Press.

Parncutt, R. (1988). Revision of Terhardt's psychoacoustical model of the root (s) of a musical chord. *Music Perception: An Interdisciplinary Journal, 6*(1), 65-93.

Parncutt, R. (1989). *Harmony: A psychoacoustical approach*. Berlin, Heidelberg, New York: Springer.

Parncutt, R. (2011). The tonic as triad: Key profiles as pitch salience profiles of tonic triads. *Music Perception: An Interdisciplinary Journal, 28*(4), 333-366.

Paul, H. (1880). *Prinzipien der Sprachgeschichte. Studienausgabe 1975*. Tübingen: Niemeyer.

Pear, T. (1911). The experimental examination of some differences between the major and the minor chord. *British Journal of Psychology, 4*(1), 56-88.

Pearce, M. T., & Wiggins, G. A. (2006). Expectation in melody: The influence of context and learning. *Music Perception: An Interdisciplinary Journal, 23*(5), 377-405.

Peirce, C. S. (1867). On a new list of categories. *Proceedings of the American Academy of Arts and Sciences, 7*, 287-298.

Peirce, C. S. (1991). *Peirce on signs: Writings on semiotic*. Chapel Hill and London: UNC Press Books.

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

Poller, T. R. (2005). *Musik und Emotion*. Berlin. Available from http://trpoller.de/Texte.html.

Rasch, R., & Plomp, R. (1999). The perception of musical tones. In D. Deutsch (Ed.), *The psychology of music* (Vol. 2, pp. 89-112). London: Academic Press.

Rochberg-Halton, E., & McMurtrey, K. (1983). The Foundations of Modern Semiotic: Charles Peirce and Charles Morris. *American Journal of Semiotics, 2*(1), 129-156.

Russell, J. A. (1979). Affective space is bipolar. *Journal of personality and Social Psychology, 37*, 345-356.

Russell, J. A. (1997). The psychology of facial expression. In J. A. Russell & J. M. Fernández-Dols (Eds.), *Reading emotion from and into faces.* (pp. 295-320). Cambridge: Cambridge University Press.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172.

Schellenberg, E. G. (1996). Expectancy in melody: tests of the implication-realization model. *Cognition, 58*, 75-125.

Scherer, K. R., & Coutinho, E. (2013). How music creates emotion: a multifactorial process approach. In T. Cochrane, B. Fantini & K. R. Scherer (Eds.), *The emotional power of music* (pp. 121-145). Oxford, UK: Oxford University Press.

Schlenker, P. (2016). Prolegomena to Music Semantics. *Manuscript, Institut Jean-Nicod and New York University*.

Schlenker, P. (2017). Outline of music semantics. *Music Perception, 35*(1), 3-37.

Schopenhauer, A. (1819). *die Welt als Wille und Vorstellung, erster Band*. Leipzig: Bibliographisches Institut & F. A. Brockhaus.

Schubert, E. (1999). Measuring emotion continuously: Validity and reliability of the two-dimensional emotion-space. *Australian Journal of Psychology, 51*(3), 154-165.

Shepard, R. N., & Jordan, D. S. (1984). Auditory illusions demonstrating that tones are assimilated to an internalized musical scale *Science, New Series, 226*(4680), 1333-1334

Sloboda, J. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music, 19*(2), 110-120.

Smolensky, P. (2000). *What I learned from Dave Rumelhart — the fundamentals of PDP: Principles of Dave's Philosophy — a personal view. Rough transcript of a contribution to the David Rumelhart Celebration held at Carnegie-Mellon University, October 15−17, 1999*. Carnegie-Mellon University. Pittsburg, PA.

Spitzer, M. (2002). *Musik im Kopf*. Stuttgart: Schattauer.

Stokhof, M., & van Lambalgen, M. (2011). Abstractions and idealisations: The construction of modern linguistics. *Theoretical Linguistics, 37*(1-2), 1-26.

Stolzenburg, F. (2015). Harmony perception by periodicity detection. *Journal of Mathematics and Music, 9*(3), 215-238.

Stumpf, C. (1883). *Tonpsychologie, Band 1*. Leipzig: Hirzel.

Stumpf, C. (1890). *Tonpsychologie, Band 2*. Leipzig: Hirzel.

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

Temperley, D. (2008). A probabilistic model of melody perception. *Cognitive Science, 32*(2), 418-444.

Terhardt, E., Stoll, G., & Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America, 71*, 679-688.

Thompson, W. F., Cuddy, L. L., & Plaus, C. (1997). Expectancies generated by melodic intervals: Evaluation of principles of melodic implication in a melody-completion task. *Perception & Psychophysics, 59*(7), 1069-1076.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, London: Harvard University Press.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: Cross-modal interactions in the perception of emotions in musical performance. *Cognition, 118*(2), 157-170.

Vines, B. W., Nuzzo, R. L., & Levitin, D. J. (2005). Analyzing temporal dynamics in music: Differential calculus, physics, and functional data analysis techniques. *Music Perception, 23*(2), 137-152.

Woolhouse, M. (2009). Modelling Tonal Attraction Between Adjacent Musical Elements. *Journal of New Music Research, 38*(4), 357–379.

Woolhouse, M. (2010). Modes on the Move: Interval Cycles and the Emergence of Major-Minor Tonality. *Empirical Musicology Review, 5*(3), 62-83.

Woolhouse, M. (2012). *Wagner in the Round: Using Interval Cycles to Model Chromatic Harmony.* Paper presented at the Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, Thessaloniki.

Woolhouse, M., & Cross, I. (2010). Using Interval Cycles to Model Krumhansl's Tonal Hierarchies. *Music Theory Spectrum, 32*(1), 60-78.

Woolhouse, M., & Cross, I. (2011). To The Editor: A Response To Ian Quinn. *Music Theory Spectrum, 33*(1), 99-105.

Youngblood, J. E. (1958). Style as information. *Journal of Music Theory, 2*(1), 24-35.

Zbikowski, L. M. (2008). Metaphor and music. In R. W. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 502-524). Cambridge Cambridge University Press.