Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognition 100 (2006) 131–172
COGNITION
www.elsevier.com/locate/COGNIT

Varieties of musical experience

Jamshed J. Bharucha [*], Meagan Curtis, Kaivon Paroo

Tufts University, Medford, MA 02155, USA

Abstract

In this paper, we argue that music cognition involves the use of acoustic and auditory codes to evoke a variety of conscious experiences. The variety of domains that are encompassed by music is so diverse that it is unclear whether a single domain of structure or experience is defining. Music is best understood as a form of communication in which formal codes (acoustic patterns and their auditory representations) are employed to elicit a variety of conscious experiences. After proposing our theoretical perspective we offer three prominent examples of conscious experiences elicited by the code of music: the recognition of structure itself, affect, and the experience of motion.

Keywords: Music perception; Musical structure; Conscious experience; Emotion; Motion

---

1. Introduction

The minds of the performer and the listener handle an extraordinary variety of domains, some sequentially and some simultaneously. These include domains of musical structure relating to pitch, time, timbre, gesture, rhythm, and meter. They also include domains that are not fundamentally domains of musical structure, such as affect and motion. Some aspects of these domains are available to consciousness and some are not. We highlight two principal distinctions in this paper. One is

---

[*] Corresponding author. Tel.: +617 627 3310.
E-mail address: Jamshed.Bharucha@tufts.edu (J.J. Bharucha).

doi:10.1016/j.cognition.2005.11.008

between the processing of acoustic and musical structure, on the one hand, and processing in domains that do not pertain to acoustic or musical structure – most notably affect and motion – on the other. The former elicits the latter in music perception. Music and the cognitive representations of its structure serve in part to *elicit* experiences of affect and motion.

The second principal distinction is between implicit processes, on the one hand, and conscious experiences on the other. It is commonly acknowledged that we are not conscious or aware of most of the processing that goes on in our brains. We suggest that the conscious experiences that we do have resulted from the allocation of attention resources to selected aspects of underlying processing. Our conscious experiences may be of the recognition of aspects of musical structure itself, or experiences – such as affect and motion – that do not pertain to musical structure.

Fig. 1 depicts schematically the essential distinctions as they apply to the perception of music. The acoustic stimulus is transduced into auditory and cognitive representations of musical structure, shown in the left box entitled "Processing and

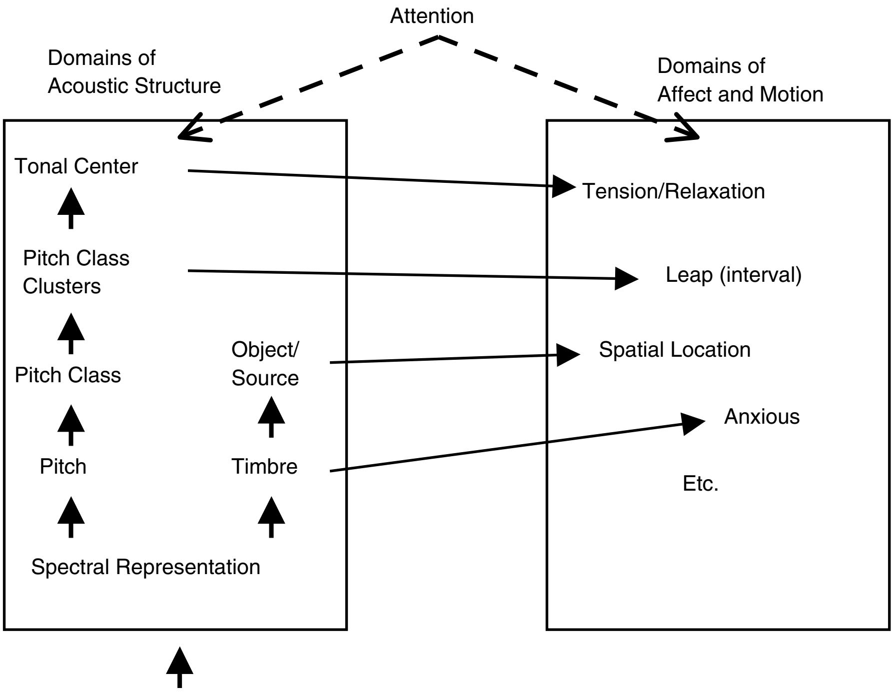

Fig. 1. The acoustic stimulus is transduced into auditory and cognitive representations of acoustic structure, including pitch, timbre, and their derivate structures, such as pitch class, pitch class clusters, and tonal centers. Functions map the acoustic structures to cognitive domains such as affect and motion. Attention to acoustic structure modulates conscious experience of cognitive domains. In the example, two kinds of affect (tension and anxiety) and two kinds of motion (interval leaps and spatial locations) are elicited by mappings from acoustic structures. Attentional resources that are directed selectively to tonality (on the left) may increase attention to the feelings of tension and relaxation.

recognition of sound sources and musical structure''. This box includes many domains of structure, including pitch, time, timbre, and gesture and their derivate structures, including chords, keys, melodies, meter, and rhythm. Cognitive domains that play a role in music (such as affect and motion) but that are not domains of musical structure per se are shown in the right-hand box. These processing domains may receive inputs from sources other than musical structure, for example, facial expressions or language in the case of affect. We have not endeavored to show all possible inputs that can activate these domains. Our purpose here is to show schematically how these domains are activated by music. Shown in dashed arrows are attentional resources that are directed selectively at aspects of the processing in the many domains involved. Attention underlies the conscious experiences we have while listening to music.

The two principal distinctions yield four cognitive categories: (1) implicit processing of musical structure, (2) conscious experience of musical structure, (3) implicit processing of domains other than musical structure, such as affect and motion, and (4) conscious experience in domains other than musical structure. Categories 1 and 3 contain the machinery that drives behavior below the threshold of consciousness. Categories 2 and 4 refer to our conscious experience, as elicited by Categories 1 and 3. We caution the reader against interpreting the conscious categories to be what Dennett (1991) considers the fallacy of the Cartesian stage in which consciousness plays out; these conscious categories refer to the aspects of Categories 1 and 3 to which attention is directed. Category 1 contains what we call *formal eliciting codes*, which do most of the causal work in music perception outside of awareness. They serve to elicit conscious experiences in one of two ways. First, if attention is directed at aspects of formal eliciting codes, we have a conscious experience of musical structure (Category 2). Second, formal eliciting codes can map onto or activate implicit processes for affect and motion (Category 3), which in turn can result in conscious experience of affect and motion (Category 4) if attention is so directed.

1.1. Formal eliciting codes

Formal eliciting codes integrate information from the sound pattern and from memory. We refer to both the acoustic signal of music and the auditory and cognitive representation of its structure as *formal eliciting codes*. The former is an *acoustic* code, and the latter is a *representational* code. An acoustic code is the spectro-temporal pattern of pressure energy that is registered by the peripheral auditory system. The microstructure specifies invariants that evoke the experiences of pitch and timbre; Palmer (1989) has examined the use of microstructure in expressive performance. The macrostructure – such as is codified in music notation – specifies organization on a larger scale. In its most elemental form, the acoustic pattern could be the raw signal emanating from an instrument. At a larger structural scale, the acoustic pattern could be something like a symphony. Ever larger scales of organization are specified by the performer's shaping of phrasing units or by structures within the genre (e.g., Sonata form). Representational codes are the auditory and cognitive representations that parse the acoustic signal, encode musical features, and shape the bottom-up representation according to top-down influences from prior learning.

Formal eliciting codes are *codes* because they can transmit and preserve information. The cognitive representation of an acoustic signal is instantiated in a different medium from the signal itself, yet can in principle be deciphered to reveal the structure of the acoustic signal it represents. Thus, acoustic codes enable information to be transmitted through the air, and auditory or cognitive codes enable information to be transmitted within the brain. Acoustic codes are mapped onto auditory codes because of the causal properties of sensory transducers, producing representations in a domain completely different from the domain of sound itself yet preserving structural information. In turn, auditory codes are mapped onto more abstract cognitive codes, by virtue of the causal properties of neural connectivity. Information can only be transmitted using a code, and codes can only serve a psychological function if they are instantiated in a causal neural system that interfaces with the world through the senses. We call them codes because one of the points we make in this paper is that music seeks to communicate conscious experiences. The acoustic signal and its cognitive representation thus serve as codes in this communicative act.

Codes are *eliciting* because they elicit or evoke conscious experiences. For example, the signal emanating from an instrument, after it has been subject to the necessary implicit processing, elicits the conscious experience of timbre and pitch, perhaps accompanied by conscious affective experiences. A harmonic spectrum, via its ensuing auditory representations, elicits a conscious experience of pitch. The three tones A, C, and E played together, via their cognitive auditory representations, elicit a conscious sense of the minor mode. A subdominant chord followed by a tonic chord elicits the unique conscious experience of the plagal cadence.

We call eliciting codes *formal* because they have at least three key properties. First, they are implicit, as we have mentioned earlier – the causal processes instantiated by the formal codes proceed systematically without our necessarily being conscious. Second, they are syntactic, and are not meaningful in and of themselves. Third, they are modular in Fodor's narrow sense of being informationally encapsulated, cognitively impenetrable and automatic (Fodor, 1983, 2000). For example, a chord automatically generates expectations for chords that typically follow, even if the listener knows that an unexpected chord is going to follow (Justus & Bharucha, 2001). These expectations are driven automatically by a causal mapping from a representation of the context chord onto a set of activations that predict or anticipate the next chord (Bharucha & Stoeckig, 1986; Bharucha, 1987).

While some of the properties we ascribe to formal eliciting codes are consistent with Fodor's characterization of the formal (syntactic) nature of mental representation in his computational theory of mind (Fodor, 1975, 1980, 1983, 2000), others are not. Unlike Fodor, who embraces a strong nativism (see Fodor, 2000), we postulate representational codes within a causal neural system that can learn some of its connectivity, based on some innate constraints (Bharucha, 1991a, 1991b, 1999; Bharucha & Mencl, 1996; Bharucha & Todd, 1991; Tillmann, Bharucha, & Bigand, 2000).

1.2. Conscious experiences

The conscious experiences evoked by formal eliciting codes are what we hear or feel when we listen to music. It is possible to conceive of these experiences as what makes music meaningful, and in the same way that linguistic meaning motivates the study of syntax, these conscious experiences motivate the study of the formal codes (Raffman, 1992). Listeners are not conscious of most features of formal eliciting codes, although they do have conscious access to some of the principal representational units (e.g., pitch and timbre) that are the result of implicit processes. Trained listeners may be conscious of some structural features that may be implicit for untrained listeners. While some conscious experiences can perhaps be evoked directly through sensory stimulation without the mediation of formal eliciting codes, complex and infinitely varied experience is made possible by such mediation. It is only through implicit processes of structural analysis, synthesis and recognition that our conscious experiences can be so systematically varied by manipulating musical structure.

We will not venture to advance a view of conscious experience, and will therefore remain agnostic about the neural basis and philosophical status of conscious experience. We can perhaps operationalize conscious experience as the content of awareness or attention. Attention is a selective processing system of limited capacity (Cherry, 1953; Spieth, Curtis, & Webster, 1954). Attention seems to be necessary for the formation of some perceptual groupings, including stream formation (Carlyon, Cusack, Foxton, & Robertson, 2001) and time varying events (Large & Jones, 1999). It can also enhance detection through frequency selectivity (e.g., Greenberg & Larkin, 1968; Scharf, Quigley, Aoki, Peachey, & Reeves, 1987; Schlauch & Hafter, 1991) and spatial selectivity (e.g., Mondor & Zatorre, 1995). A distinction is sometimes made between exogenous and endogenous attention (see Spence & Driver, 1994), the former being an unconscious early mechanism and the latter a conscious mechanism that functions later in processing. In this paper we use attention or conscious experience synonymously to refer to endogenous attention.

Some conscious experience can be reported verbally, as in ‘I hear a violin’, ‘It sounds dissonant’, ‘It sounds sad’, ‘It takes me back to my childhood’, or ‘It makes me want to dance’. In the case of highly trained musicians, potential verbal reports may be more specific and more focused on structural features, for example, ‘an augmented sixth chord’, ‘modulation to the subdominant’, ‘three against four’, ‘I recognize a motif from the exposition’. However, only a subset of the domains of conscious experience can be equated with explicit knowledge. Much of our conscious experience is ineffable (i.e., we can’t seem to find the words to describe it) because the objects of conscious musical experience are often more nuanced (fine-grained) than the categories for which we have an available lexicon (see Raffman, 1993). Particularly for novices, most conscious musical experience is probably ineffable: there is something of which they are aware but somehow unable to articulate.

Domains of conscious experience (e.g., affect) may have their own structure. Thus the essence of the distinction between formal eliciting codes and conscious experiences is not that one has structure and the other does not, but rather that conscious experience does not directly reveal the structure that elicits it. And yet it is the conscious experience that we report, or unsuccessfully attempt to report, that music lovers cite as the *raison d'etre* of music, and that those of us who study music cognition seek to explain. Cognitive science was a breakthrough precisely because it recognized that the physical causation that enabled cognition could not be discerned by noting regularities in conscious experience (the phenomenological method). Most of the causal processes and neural representations that code information and upon which the processes operate are not available to consciousness. Why some of the outputs of these processes are available to consciousness, or what that means, is beyond the scope of this paper (see Dennett, 1991).

What Dennett (1991) calls the "phenomenological garden" is rich while listening to music, and even richer while performing. It has a fleeting, vacillating quality: I am now aware of this, now of that, as attention switches from one level of processing to another, or from one domain of representation to another. It would be interesting if future research reveals a better understanding of how and why attention – and thus our conscious experience – samples selectively the vast array of information being processed as we listen. For the time being, the body of research in music cognition would seem to suggest that the eliciting codes do their work reliably, and our conscious experience reveals but a fraction of the formal cognitive processing of sound patterns.

If we could communicate directly some of the conscious experiences we have while listening to music, without the mediation of air and our auditory systems, eliciting codes would perhaps be unnecessary. They are necessary because the structural properties of some conscious experiential domains do not enable them to function as communicative media in and of themselves. Musical structure and affect are distinct domains, but the former can elicit the latter. We also communicate affect through facial expressions (which serve as formal eliciting structures in the visual domain), even though the domains of facial expressions and of affect are distinct.

1.3. Mapping between domains

The field of psychophysics was originally conceived to discern the functions that map from physical attributes to psychological attributes (Thurstone, 1927): frequency to pitch, spectrograms to timbre, frequency-time-space patterns to stream segregation, etc. A set of psychophysical functions, $\mathbf{f_P}$, maps acoustic structures from the domain of sound, $\mathbf{S}$, to the psychological domain, $\mathbf{P}$:

$$\mathbf{f_P}(\mathbf{S}) \rightarrow \mathbf{P}.$$

With the development of psychoacoustics and then cognitive psychology, $\mathbf{P}$ has come to include not just sensory and perceptual domains, but also increasingly abstract cognitive domains (e.g., expectations, keys, and rhythms). This suggests a hierarchical set of mappings from low-level auditory representations to more abstract cognitive representations. Thus, sound ($\mathbf{S}$) is transduced ($\mathbf{f_T}$) into a set of representations ($\mathbf{R}$):

$$\mathbf{f_T}(\mathbf{S}) \rightarrow \mathbf{R}.$$

Low-level auditory neuroscience is devoted to the articulation of these transduction functions, $\mathbf{f_T}$. Cognitive science and neuroscience are devoted to the mapping of one representational domain, $\mathbf{R_i}$, onto another, $\mathbf{R_j}$, via a set of cognitive mapping functions, $\mathbf{f_C}$:

$$\mathbf{f_C}(\mathbf{R_i}) \rightarrow \mathbf{R_j}.$$

The representational domains, **R**, include the domains of pitch, timing, timbre, motif, emotion, motion, memories and a host of others. When attention is allocated to regions within **R**, we have conscious experience of what is being represented by that region. We can characterize the allocation of attention as yet another mapping function, $\mathbf{f_A}$:

$$\mathbf{f_A}(\mathbf{R_i}) \rightarrow \mathbf{R_c},$$

where $\mathbf{R_i}$ is an implicit representation and $\mathbf{R_c}$ is a conscious one. This last function is thus the eventual function that elicits conscious experience. We wish to make clear that we do not see this mapping as a transduction into a non-neural domain (Dennett, 1998), but simply as a mapping from one neural domain into another.

There may be a many-to-one mapping from some domains of eliciting codes onto some domains of affect. Affect can be elicited by non-auditory codes such as facial expressions and language.

Attention is not just a process that makes us conscious for its own sake. We would suggest that it provides mapping functions that are not available within the modular implicit processing systems. For example, attention may provide enhanced detectability of tones (Greenberg & Larkin, 1968; Schlauch & Hafter, 1991), enhanced fusion into streams (Carlyon et al., 2001; Large & Jones, 1999), and enhanced binding of features into integrated objects or situations (Wrigley & Brown, 2002). We can postulate that these enhanced or newly integrated representations mediated by attention are available to the implicit representational system as another form of top-down processing. We characterize this as a reverse mapping function from conscious to implicit:

$$\mathbf{f_A}(\mathbf{R_c}) \rightarrow \mathbf{R_i}.$$

Mapping functions include mapping from one hierarchical level to another, including at least the following levels: spectral representation to pitch (Terhardt, Stoll, & Seewann, 1982), octave-equivalent pitch class (Bharucha, 1991b; Bharucha & Mencl, 1996), intervals, chords, and keys (Bharucha, 1987; Janata, Tillmann, & Bharucha, 2002; Krumhansl, 1991; Leman, 1995; Lerdahl, 2001; Lerdahl & Jackendoff, 1983; Tillmann et al., 2000). We postulate organizational units such as chords and keys because we are conscious of them. But they are extracted from the spectrum through processes of which we are not conscious.

Mapping functions also include mapping over time (Jackendoff & Lerdahl, in press; Lerdahl & Jackendoff, 1983); mapping from the musical piece to its hierarchical representation occurs over time. Each level of the representational structure is a representational domain in our nomenclature, and the rules to derive one level from the others are the mapping functions.

Mapping over time also includes the expectations generated by a musical context: both schematic expectations (expectations for the most probable continuations) and veridical expectations (expectations for the actual next events in familiar sequences, whether they are schematically likely or not; see Bharucha & Todd, 1991). Mapping functions are thus a form of long term memory – either schematic knowledge or memory for specific musical sequences.

Cognitive mappings are not all sequential bottom-up processes, but rather require the top-down influence of stored representations learned from prior experience, as well as the iterative interaction of top-down with bottom-up processes (Bharucha, 1987). In the case of interactive processes, the cognitive mapping function may need to be unpacked into more local mapping functions that work in ensemble to implement the larger function. For example, in seeking to account for a variety of phenomena in the perception of harmony, Bharucha (1987; see also Tillmann et al., 2000) proposed a neural net that maps from a vector of pitch class activations (representing a decaying pitch class memory over a window of time) to a vector of chord activations and a vector of key activations. The chord activations develop as a result of an iterative accumulation of activation driven from the bottom by the pitch class activations and from the top by the key activations. In the first iteration, there is no information at the key level, so the chord activations are driven solely by the pattern of pitch class activations. The activation of each chord unit is set by spatial summation of activations across the 12 pitch class units, weighted by the strengths of the connections from them.

The weight vector is a form of long term schematic memory and enables the chord unit to function as a filter or complex feature detector. The more closely correlated the pitch class pattern of activation is to the weight vector, the more strongly activated that chord unit will be. In subsequent iterations, key units get activated in analogous fashion by the chord units, and the chord units start to be influenced by both the pitch classes and the keys, until a settled state is reached, which manifests the combined influence of bottom-up (stimulus driven) and top-down (memory driven) effects. Empirical evidence in support of the developing pattern of activation over time comes from priming experiments (Tekman & Bharucha, 1998), and the final settled activation patterns account for data from a range of cognitive tasks (Tillmann et al., 2000). In addition, Bharucha (1991b) and Tillmann et al. (2000) demonstrated how the weight matrices that map the pitch class vector to the chord and key vectors on any given iteration cycle are learned through self-organization. The mapping, $\mathbf{f}_\mathbf{C}$, of interest here is from the pitch class pattern to the settled chord and key patterns of activation, after learning has taken place.

Cognitive mapping functions have been articulated within a variety of modeling paradigms, including grammars (e.g., Lerdahl & Jackendoff, 1983; this volume; Narmour, 1990), spatial models (e.g., Krumhansl, 1991; Lerdahl, 2001), and neural nets (e.g., Bharucha, 1987; Tillmann et al., 2000). In grammars, mapping functions are rules, and the representations are rule-governed strings of symbols. In spatial models, the mapping functions and the representations are spatial configurations. In neural nets, the mapping functions are connection weights between neuron-like units; representations are patterns of activation across vectors of elements that typically function as feature detectors.

1.4. Innate versus learned mappings and representations

In the domain of harmony, there are strong correlations between experienced relationships and acoustic relationships resulting from the physical structure of sound and its transduction. However, it is clear that cultural learning does takes place. Tekman and Bharucha (1998) demonstrated this by pitting cultural convention against acoustic structure in a priming paradigm. In the Western musical environment, the C major chord is more likely to be followed by the D major chord than by the E major chord, because C–D is IV–V in F Major, whereas C–E is not a typical chord transition within any given key. Yet the C major chord shares more harmonics with the E major chord than it does with the D major chord. Thus, C and D are acoustically more closely related, but C and E are culturally more closely related. We found that the cultural relationship dominates the acoustic one: the C major chord primes the D major chord more strongly than it primes the E major chord. This effect cannot be explained by physical constraints (the harmonic structure of pitch producing sources) or known psychophysical phenomena (including both spatial and temporal processes in the auditory system). It must therefore be a result of cultural learning. Differences between listening to culturally familiar versus unfamiliar music also support an effect of cultural learning (e.g., Castellano, Bharucha, & Krumhansl, 1984). There is thus clear evidence against any extreme form of nativism. Not all cognitive mapping functions, $\mathbf{f_C}$, are innately specified, although some may be. A constraint on widespread appreciation of a musical genre is whether these culturally internalized mappings are shared in order to successfully achieve music's ability to communicate conscious experience.

The transduction functions $\mathbf{f_T}$ presumably are innate. They would include the properties of inner hair cells, which transduce mechanical energy of the basilar membrane into neural impulses, and the frequency-tuned properties as well as the temporal properties of cochlear neurons (see Gulick, Gescheider, & Frisina, 1989). They would also presumably include the range of known response characteristics of neurons in the ventral cochlear nucleus, including the capacity to distinguish phasic from tonic responses. As one goes further up the nervous system, it is more difficult to discern whether mapping functions are innate. More research will be required before we have a clear sense of which mapping functions are innate and which are learned. However, given that some cultural mapping functions must be learned, it is important for us to have models for how that might occur, and to test the predictions they make.

In our modeling work (Bharucha, 1991a, 1991b, 1999; Bharucha & Mencl, 1996; Bharucha & Todd, 1991; Tillmann et al., 2000), we have shown how cultural learning of chordal expectations might occur through passive perceptual exposure (see also Leman, 1995). Neural net models assume a set of primitive feature detectors that have innate tuning characteristics. Also assumed is the ability of connections between neurons to be altered through Hebbian learning (Grossberg, 1976; Hebb, 1949; Rumelhart & McClelland, 1986). In our work we have shown how, starting with representational units tuned to pitch class, Hebbian learning as it is developed in models of neural self-organization leads inexorably to the formation of representational units for chords and keys, following exposure to the structure of Western music. Thus the features of the representational domains of chord and key may themselves be learned, as well as their relationships. We are conscious of familiar chords as having a unitary quality, while unfamiliar chords (such as some used in jazz that are unfamiliar to many listeners) sound like a cluster of tones, and fuse only with more exposure.

In our model, the culturally learned priming result mentioned above occurs because the tones of the C major chord activate their pitch class representational units, which in turn activate the chord representational units with which they have become connected through self-organization. Initially, the E major chord unit is more strongly activated (expected) than the D major representational unit, because E major shares a component tone with C major and D major shares none. However, the chord units in turn activate key representational units with which they have become connected through self-organization, and the key units activate chord units in a top-down fashion. As the top-down activation asserts itself, the D major chord unit becomes more active than the E major chord unit. Tekman and Bharucha (1998) tested this time-course prediction made earlier (Bharucha, 1987) by varying the stimulus onset asynchrony (SOA) between prime and target. For short SOA's (50 ms) E major is more strongly primed than D major. For longer SOA's, the pattern reverses as the culturally learned mappings take over.

No doubt there are many innate constraints on cultural learning of cognitive mappings. One we wish to note is invariance under transposition. Bharucha and Mencl (1996) suggested a model in which virtual pitch and the tonic of a key can be used as references to map chords or melodies into a pitch-invariant format. To date we are not aware of any model of how this might be learned. Another likely innate constraint is one suggested by Lerdahl and Jackendoff (1983), which is that whatever the culturally specific mapping functions, a generative hierarchical mapping of the sort they propose is likely to be universal. We would add, based on the discussion below, that this would occur only if the eliciting codes adopted by a culture lend themselves to hierarchical combinatorial generativity.

Nowak and Komarova (2001) frame the development and evolution of language as the change in weights in learning matrices representing each of the two levels of patterning in language: one associating sound patterns with lexical meaning, and one associating syntactic patterns with propositional meaning. Variability in the grammars implicit in the weights enables evolution when individuals succeed in communicating. In our framework, sound patterns are associated with a range of experiential states. The association matrices are acquired both ontogenetically and phylogenetically, manifesting themselves in development (learning) and evolution. Some components of the association matrices may evolve as a result of random variability shaped by social payoffs that occur when individuals recognize that similar acoustic codes evoke similar conscious experiences. These social payoffs include the ability to successfully communicate emotions or other feelings, and the social bond that results from synchronization of experiential states. The resulting association weights specify the innate constraints on learning.

Learning co-occurs with cultural development (as in the development of new forms of music or generational differences in music appreciation) and has a viral quality. The constant quest for new sounds or hits, coupled with variability in the association matrices across individuals, may result in social payoffs that are then copied in the form of musical archetypes. These in turn modify cultural regularities and subsequent learning that in turn influence expectations and their associated evoked states. All the while, there is a healthy tension between fulfilling automatic (modular) expectations (priming) induced by the internalization of cultural regularities and the violation of these expectations. The balance between the fulfillment and violation of expectations reflects the countervailing preferences people have for familiarity and novelty. Some sounds (such as familiar timbres, voices, gestures, motifs, pieces, or recordings) are expressed by the producer and resonate with the listener because they are familiar. (The preference for familiarity itself has multiple roots, including predictability and social identity). Other sounds (such as new timbres and voices, violations of familiar motifs, or new interpretations of pieces) are expressed or resonate because they are novel. Crafted music as we know it today is thus the convergence of multiple developments, and cannot be understood as if it were the result of grand design.

Much has been written about the evolution of music (Wallin, Merker, & Brown, 2000). We would suggest that one candidate that is missing from the discussion is the role music may have played in the evolution of culture, and perhaps in the co-evolution of culture and biology, by facilitating memory. This includes memory for music as well as for declarative knowledge. Music may have facilitated the ability to pass declarative knowledge from one generation to the next. Oral traditions reveal an interesting connection between sound patterns and memory (Rubin, 1995). Musical performance without written notation entails an enormous capacity to recall long sequences. In oral traditions, poetic devices and music have also been used to transmit declarative knowledge. Rubin has studied extensively the role of rhyme, alliteration and assonance in memory for verbal materials. He argues that the expectation of a repetition of sound (in rhyme) cues recall and constrains the search space. Rhythm and meter provide a recurrent temporal framework within which verbal memory can be facilitated. In ballads, for example, linguistic and musical stress tend to coincide (Wallace & Rubin, 1991). In vocal music, as in story telling in the oral tradition, formal musical structure is used in part as a vehicle to elicit linguistic meaning by synchronizing speech with music and leveraging the memory advantages of music. While Rubin's work has shown memory advantages for metrical structure, it remains to be determined whether there are such advantages for melodic or harmonic structures, this hypothesis remains a provocative one, at least for meter.

1.5. Are there necessary conditions for music?

Given the variety of representations and experiences associated with music, we might ask whether any of them is an essential ingredient of music – a necessary condition for calling something music. A definition is a set of necessary and sufficient conditions. There are clearly plenty of sufficient conditions for something being music, as we shall see below. As we address the topic of this volume – the nature of music – it behooves us to consider whether there are any features or conditions that are necessary.

In language, formal phonological, manual, and syntactic structures have the power to elicit lexical and propositional meaning. The syntactic structure of language constitutes a formal code by which meaning can be encoded and communicated. If a code cannot represent and communicate propositional meaning, we would not call it language. If it lacks recursion, or syntactic categories of noun phrase and verb phrase, we would not call it language (although this may now be contested – see Everett, forthcoming). These (and possibly other) universal properties of syntax may not be sufficient conditions for a code being called language, but they are necessary.

It is more difficult to identify necessary conditions for what constitutes music. This point is made not to diminish music as a cognitive capacity but to recognize its varied nature. While the use of pitch categories and pitch patterns is typical of music, it is not a necessary condition: African drumming is clearly music. Conversely, while the use of rhythmic and metrical patterns is typical of music, it is not a necessary condition: the *alap* or opening section of a performance of Indian classical music is a rhythmically free form. There are compositions that are purely timbre-based, and sometimes even isolated timbres can elicit powerful experiences.

Other promising candidates for necessary conditions include hierarchical structure and the existence of a corpus of preference rules governing this structure (e.g., Lerdahl & Jackendoff, 1983). Yet a composer who chooses to eschew such structure is free to do so and may insist on calling the resulting creation music. Furthermore, listening to certain timbres could be considered music by some, even in the absence of pitch-time hierarchical structure; and while timbre hierarchies may exist for some forms of music, timbral variation or patterning per se does not imply or evoke a sense of hierarchical structure. The child, musical novice or Alzheimer's patient who picks away at an individual string or key or a musical instrument and thrills at the raw sound is having a musical experience, rigid definitions of music notwithstanding. Questioning the use of the term 'music' when typical features are missing ("That's not music!") or when audiences do not respond does not have the same weight as questioning the use of the term 'language' when typical features are missing or no one understands it.

While auditory experience may be a necessary condition for calling something music, it does not get us very far in understanding music as a cognitive or brain function, and is not a necessary condition of all experiences evoked by music. For example, emotional experiences evoked by music are not themselves auditory experiences. Some listeners enjoy the recognition of structure and structural manipulation over and above the auditory experience, even though the structures may be built from auditory elements. The experience of music is sometimes characterized spatially (e.g., Johnson & Larson, 2003; Krumhansl, 1991). Music evokes experiences of expectancy, violation, closure, and a host of other mental states that are not specifically auditory, even though they may be triggered by the use of sound. The rhythmic pulse felt in most music is as much a result of a pulsing of attention as it is a perception of period stresses in the sound (Jones & Boltz, 1989; Large & Jones, 1999). Finally, there is an extraordinary variability in the reported conscious experience of music. The plagal cadence is sometimes characterized as warm, and timbres are often described as bright, dark or even sweet. Some claim to experience keys or other musical structures as emotions.

The cognitive activities that we call music are not unified by properties that are necessary, but instead constitute a fuzzy set whose elements are bound together by multiple properties that run through overlapping subsets of instances. A family resemblance structure (Rosch & Mervis, 1975; Wittgenstein, 1958) more accurately describes music than does a set of necessary and sufficient conditions. Some features are more typical than others, but no one feature is necessary. Music is a composite of multiple brain functions, which through cultural and possibly biological evolution and co-evolution have found particular resonance with listeners when implemented together in the ways that have proven most receptive to listeners. Music that eschews one or more of the most typical properties tends to have smaller audiences than does music that leverages these properties in a convergent way. Music that eschews most of the typical properties becomes regarded as experimentation, rebellion, or self-indulgent, resulting in niche audiences.

Pitch and temporal patterning are features most typical of music as we know it, but are not necessary. They have two characteristics that account for their pervasive use: (1) they draw upon neural systems that enable the generative creation of infinitely many hierarchical structures (Lerdahl & Jackendoff, 1983), and (2) they seem to have, through either development or evolution, the capacity to evoke a variety of experiential states. Their generativity enables them to serve as communicative codes that – while not necessary – are pervasive in music because they support the expression and evocation of a varied and infinitely dense space of experiential states in ways that have been either adaptive or desirable. The selection and constrained combination of a small number of pitch classes to form modes or keys and their organization into schematic and event hierarchies (Bharucha, 1984) enables an explosion of possible sequences that – despite their diversity – are recognizable as instances of culturally familiar forms. In the temporal domain, the hierarchical organization of isochronous pulses into metric structures and the ability to represent event onsets in relation to an underlying temporal grid (Povel, 1984) enables further explosions of possible temporal sequences.

The capacity for hierarchical structuring of musical events and of the building blocks of music in the domains of pitch and rhythm has driven the development of musical art forms to their current levels of complexity. Hierarchical representations are of two principal types: event hierarchies and tonal hierarchies (Bharucha, 1984). Event hierarchies represent actual musical events hierarchically in the context of their temporal sequence in a piece of music, and are exemplified by the formal models of Deutsch and Feroe (1981) and Lerdahl and Jackendoff (1983). In the time-span reduction of Lerdahl and Jackendoff (1983), the finest grain of the hierarchical representation consists of metrical pulses which are then combined in successive binary or ternary units to show the subordination of weak beats by neighboring strong beats. At higher levels of the hierarchy unstable pitches are subordinated to stable pitch neighbors, unstable chord functions to stable neighboring chord functions, and so on with longer and more abstract units subordinated to neighboring units. The prolongation reduction represents the evolution of tension over time. While the preference rules that drive the reductions may vary from one musical culture to another, the resulting organization of events is pervasive.

It is interesting that the domain of timbre thus far has not proven to be the basis for pervasive generative hierarchical structure in music as it does in speech. Psychoacoustically, timbres in music are somewhat analogous to phonemes in speech (percussive sounds and sharp attacks correspond to consonants, and steady state timbres correspond to vowels). Like timbres, we identify phonemes by their spectrographic representation, albeit in the context of preceding and succeeding phonemes (Wickelgren, 1969). Potentially infinite numbers of words are generated by combining a limited set of categorically different phonemic units in rule governed ways. One could imagine sequences of phonemes or timbres that are phonologically or 'timbrally' well-formed (but linguistically meaningless) serving as the basis for a musical genre in which timbre is the principle domain of variation. Yet with a few limited exceptions this seems not to have emerged in a pervasive way. This could be in part because of the development of acoustic musical instruments; an acoustic instrument provides an extraordinary pitch range but a comparatively limited timbre range. Electronic music synthesizers expand our timbre range and in theory present the opportunity to manipulate timbre in a generative way at the rate of phonemes in speech, but this application seems not to have taken root yet. Composers have indeed tried to create generative timbre systems using either speech sounds or electronically synthesized timbres, but these compositional systems have not achieved any significant purchase beyond individual composers. Timbre variation was always possible with voice but has developed in only limited ways that have not created combinatorial explosions based on shared constraints on well-formedness. (Limited exceptions include scat singing in jazz and the *Bol* system in Indian drumming, in which drum timbres are named and spoken rhythmically: *dha*, *dhin*, *ta*, *tin*, etc.). It may be that generative timbre variation, such as is found in the sequencing of phonemes, is a modular function linked to language. Nevertheless, a performance of free-form timbre variation without pitch and rhythmic structure would clearly count as music. Thus, while generativity (in the domains of pitch and time) is typical because of its extraordinary power, it is not a necessary condition of music.

While music and language readily share an underlying function that supports generative stress hierarchies of pitch and time (Lerdahl & Jackendoff, 1983), hierarchical generativity of timbre variation seems to be owned by language with little spillover to music. It is intriguing to consider whether hierarchical generativity of pitch is owned by music with little spillover to language. Not enough is known to be definitive about this, but if it were true it might suggest a specialized musical function to which the capacity for generative pitch patterning is yoked.

Indeed, there are findings suggestive of cognitive capacities specialized for music. Studies on congenital amusia (tone deafness) show that some people have severe deficits in pitch processing and even temporal processing that are specific to music (Ayotte, Peretz, & Hyde, 2002; Peretz et al., 2002). Amusics would have difficulty with most music, and would be unable to engage in most communicative acts that we call music, because most music employs variation in pitch and time. However, as we pointed out earlier, free-form timbral variation may count as music, even though it may be more the exception rather than the rule. We do not rule out the possible existence of cognitive capacities specialized for music, but simply argue that the extraordinary diversity of domains that count as music make it difficult to specify necessary conditions.

In language, formal codes (the sound patterns of speech, the manual patterns of sign languages, and syntax) evoke meaning. While meaning may itself have structure, it is the structure of a domain entirely different from that of sound and syntax, as evidenced by the existence of well formed linguistic structures that are meaningless, and of expressions in different languages with much the same meaning. Generative structure in language yields an infinite number of possible propositional meanings. Generative structure in music yields an infinite number of possible experiences.

Musical experiences may be enduring or fleeting, clear or elusive, unambiguous or ambiguous. They may exist as a simultaneous multiplicity. They may be nested or loosely interconnected. They may be easily described or ineffable. They may be emotions or more subtle experiences. They may be auditory or abstract, motoric or synesthetic. Music uses sound to evoke experiential states in a way that goes beyond the distinctive requirements of other forms of auditory expression such as speech and non-speech vocalization. Herein lies the difficulty in developing a semantics of music.

Language utilizes formal codes to communicate meaning, and it is this distinction between an eliciting code and its elicited meaning that leads some to suggest that music too has a semantics. Raffman (1993) argues that the generative structure of music leads the listener to expect meaning to emerge from the structure as it does in language, leading to a sense that the music is meaningful. The experience of meaningfulness coupled with an inability to articulate the meaning may contribute the sense of ineffability and profundity.

1.6. Music as communication of conscious experience

Music is communicative to the extent that it involves an attempt (whether successful or not) to evoke a set of conscious experiences. These experiences may be those of the composer or performer, in which case it is an attempt to align the listener's experiences with those of the composer or performer. Whether or not the evoked experiences are congruent with the intended evocations, it is the attempt to evoke them that distinguishes music from other sources of sound. Thus a natural sound could be called music if it is produced by a person with a communicative intent, but not if it is heard in its natural context without any intention – its creation is intrinsic to its being music and it is not merely generated as a byproduct of another activity. Spontaneous vocal expressions such as crying or wailing are not music, although music may draw upon them. There are cultures in which wailing is used intentionally or in ritualized social contexts, in which case it would count as music.

There are several special cases of musical communication worth enumerating. First is the case in which an originator (typically a composer or performer) seeks to evoke a set of experiences in the minds of listeners. This communicative function may or may not be *expressive*. The expressive case is one in which the originator seeks to evoke his or her own experiences (or memories thereof) in the mind of the listener – to get the listener to feel what the originator has felt. There are social advantages of successfully communicating in this way. Clearly, this is one of the communicative functions language can play. In the simplest case of communicating propositional meaning, a speaker who wishes a listener to understand a proposition uses a linguistic code to cause the listener to represent the same proposition.

There are cases in which communication is not expressive. Here, the originator seeks to evoke a set of experiences, but not necessarily ones the originator wishes to express. An originator may seek to evoke a set of experiences even though the originator is not having the same experiences. This may be called *designative*. The originator believes that by structuring sounds in a certain way, he or she can, by design, evoke a designated set of experiences in listeners. Presumably the originator would also have the same experience evoked while listening, but is not attempting to express a prior experience. For example, a skilled originator may seek to place the listener in a certain mood or motoric state, even though the originator is not in that mood or motoric state. Skillful and experienced originators may have learned devices to do this. This function could be called manipulative rather than designative, but ‘manipulative’ carries a limited set of connotations. Unskilled originators also can do this by playing a recording they believe will place listeners (including themselves) in a set of designated experiential states. The artist on the recording may have adopted either an expressive or designative stance. The communicative function is typically a composite of the composer’s and performer’s intentions, if there are any. Music’s communicative function is often frustrated, because the experiences the originator wishes to communicate are often so inscrutable and dependent upon the individual’s own history, context, and allocation of attentional resources that the listener is not likely to experience the same state. The content of evoked experience may vary across listeners, and may not necessarily correspond to those the performer intended to communicate.

To the extent that language expresses propositional meaning, the mode of communication is transparent or diaphanous. The meaning pops out, and to the extent that registering meaning is an experience, the experience is not auditory (as in spoken language) or visual (as in written language), but rather the comprehension of propositional meaning. Indeed, memory is weaker for the perceptual features of spoken or written language than for the meaning communicated. People remember the meaning of a sentence better than the sentence itself. When we tell the same story repeatedly, we are unlikely to use the same sequence of words. We attempt to preserve the semantics (with some change) but use any number of different syntactic structures to communicate it. In contrast, we tend to perform the same musical sequence with roughly the same structure. There may be variation in repeated performance, but there is not a sense in which we use arbitrarily different structures to communicate the same meaning. Variations on a theme are related to each other by family resemblances, not by transformations in the linguistic sense. Lerdahl and Jackendoff (1983) correctly rejected the transformational analogy to language and instead advanced Schenker's notion of a mapping from the ornamented musical surface to a hierarchically organized set of reductions.

We classify the conscious musical experience into three broad categories: the conscious recognition of structure, affective experience, and the experience of motion. These are elaborated in the following three sections of the paper. There may be other forms of conscious musical experience (e.g., colors), but these three broadly conceived categories designate markedly different qualities of conscious experience that are commonly reported (Pike, 1972). The first section below – on the conscious recognition of structure – is brief because we have covered this in the introduction of this paper, and our past work has been exclusively on aspects of musical structure. We would thus like to draw particular attention to affective experience and the experience of motion.

2. The conscious recognition of structure

Auditory representations exist at multiple levels simultaneously, and we can consciously attend to different levels. For example, we can attend to the timbre, or to the tonal structure, or the rhythmic structure, or the expressive deviations from any of these aspects during performance. We have conscious experience of some of the structural features that are the result of underlying processing, but this awareness is just the tip of the formal structural iceberg.

We can also consciously experience aspects of temporal structure, although as with tonal structure, most remains implicit. Jones and her colleagues (Jones & Boltz, 1989; Large & Jones, 1999) have argued that the isochronous beat that is entrained by a sequence of tones is an oscillation of attention. The periodic focus of attention in time is felt as an accent, even though it might only be implied (i.e., an acoustic event may not occur on each beat). The implied grid provides the basis for metric hierarchies and rhythmic patterns (Povel, 1984). These patterns may be experienced as abstract structures. They may also be experienced as motion, which we address in more detail later.

The conscious perception of structure can be roughly classified into the recognition of sound sources and the recognition of musical structures. The recognition of sound sources is a result of the analysis of the auditory scene and its parsing into segregated streams of recognizable sound sources (see Bregman, 1990). We consciously recognize sound-producing sources: bells, musical instruments, specific techniques for playing an instrument, voices, crumpling paper, etc. We may recognize both individual sources (a specific violin, a specific person's voice) as well as the categories of which they are instances. We may or may not have a name for a source of which we are conscious. For example, I may report hearing something without being able to identify it. The conscious experience of a sound-producing source is typically called a timbre. Timbre descriptors often refer to sources that produce the experiences, e.g., a nasal timbre of an oboe or a particular violin is similar to a nasal voice. We refer to a timbre as metallic when it sounds like what we hear when a metal object is struck. The chief project for a cognitive psychologist studying timbre is to understand how the acoustic patterns that emanate from sources map onto representational codes, and how these elicit the timbres we hear. The timbre is thus the *explanandum*, and the eliciting code the *explanans*. Without being scientists, we recognize the timbre of a violin because our brains are equipped to deliver a conscious experience by undertaking a tremendous amount of information processing of which we are not conscious. As scientists, we seek to understand these processes in order to explain what we know from our conscious experience.

The listener may be conscious of any of a number of structural features as a piece of music unfolds, and will certainly not be conscious of a vast number of them. For example, the listener may be aware of a particular pitch at a particular time but unaware of the individual frequencies that fuse in perception to form that pitch. This indeed is the norm. However, under certain circumstances and with training, one can hear a selected frequency component as a pitch, particularly when cues can be generated to segregate this frequency as a perceptual stream (Bregman, 1990). Psychoacousticians have called this 'hearing out' a partial (see Dennett, 1991, for a discussion of this phenomenon in the context of understanding consciousness). At a more abstract level, we may be aware of a chord but unaware of the pitch classes that comprise it because they have been fused in perception. Again, however, under certain circumstances we might be aware of one or more individual pitch classes. Sometimes we seem to hear the chord as a *gestalt*, as a singular object that is more than just the set of its component pitch classes, and sometimes we may be aware of one or more pitch classes while the overarching chord recedes into the background. When we are aware of the harmony (so-called vertical structure) we are attending to the chord and key levels of representation. When we are aware of melody or counterpoint (so-called horizontal structure) we are attending to patterns of activation at the pitch class level of representation and the chord and key levels are backgrounded. Yet all levels of representation are in causal play. The underlying processing of pitch and harmony goes forward inexorably and automatically (see Justus & Bharucha, 2001), maintaining patterns of activation at each of several levels of abstraction, regardless of whether or not we are aware. Attention can be focused on one level at the expense of another, or switched back and forth, yielding glimpses of recognition of structural features.

The recognition of musical structures includes the recognition of: clusters of frequencies as pitch; collections of octave-equivalent pitches as pitch classes; simultaneous or sequential pairs of pitches as intervals; simultaneous or sequential clusters of pitch classes as chords or chord functions, keys or modes; sequences of pitches as melodies, motifs, themes or gestures; patterns of onsets as meters or rhythms. In all the examples named here, the conscious experience is that of a unitary percept (a simultaneous or broken chord is experienced as a *gestalt*, a unitary psychological object over and above its component pitch classes). Motifs and even melodies, when familiar, sound like continuous, unitary percepts that we can reference as objects integrated over time. The conscious experience is of a series of pitches bound together in some way. The eliciting structures and processes that give rise to this percept are largely unavailable to consciousness. Gjerdingen (1994) proposed an important model of how temporal smearing can give rise to a percept of continuous melodic motion akin to the visual smearing of picture frames in a movie. Even in these spatially and temporally fused forms, we can attend to component features (such as pitch, timbre, melodic intervals, or the words of a song) if we focus or switch our attention, but the contents of our conscious awareness are nevertheless limited to selected outputs of the implicit processes that elicit these percepts.

We may also recognize abstractions of large scale musical structure, such as keys, modes or rhythms. These patterns, while elicited by sound, elicit conscious experiences that are more abstract than the sensory qualities of timbre and pitch. A beginner who may recognize the minor mode has to be taught to attend to the component tones and to realize that the emergent perceptual quality is elicited by a particular set of tones. At its most abstract, the conscious experience may be a vague recognition of culturally familiar schematic features.

We may also be conscious of structural *relationships* without knowing how they are elicited. For example, we consciously experience dissonance (actually, varieties of dissonance) without knowing what produces it. When the partials of one harmonic tone fall within the critical bands of the partials of another harmonic tone, we hear dissonance. The relative structure of the spectra of the two harmonic tones is the code that, when processed by the auditory system, elicits an experience we call dissonance. The recognition of relationships between structures includes the recognition of *similarity* to familiar musical objects (such as variations, or the awareness of similarity to another piece) or schematic cultural features driven by the structure of the eliciting pitch, rhythmic and timbral patterns (as in ‘this sounds Indian’).

Another domain of conscious recognition of structural relationships is that of stability (Krumhansl, 1991) and expectancy (Meyer, 1956; Narmour, 1990). In familiar tonal contexts, some tones are heard as more stable than others. The mechanism by which a tonal context differentiates tones by stability, and that instantiates relative stability, is the object of investigation (see Tillmann et al., 2000) but is not accessible to awareness. However, the stability is accessible, as evidenced by the ability of subjects to report it when probed (Krumhansl, 1991). Tillmann et al. (2000) suggest that stability is instantiated as the activation of representational units. The prior encoding of schematic relationships between tones, when stimulated by the current context, produces patterns of activation that represent relative stability. When subjects are probed with a tone following a musical context and asked to rate how stable the tone sounds (or how closely related it is to the context), they attend to the representational unit corresponding to the probe tone and read off the level of activation.

We are conscious of the fulfillment of an expectation (resolution, closure) and of its violation, but not of the probability-encoding and activating structures that elicit feelings of resolution and violation. In the melodic domain, tones that are unstable are experienced as wanting to resolve to a stable neighbor. This can be explained as follows (see Bharucha, 1996). Each sounded tone draws attention to a frequency band of roughly a quarter of an octave or minor third – called the attention band (Luce, Green, & Weber, 1976; Scharf et al., 1987). Stable tones that fall within this attention band capture attention. This capture of attention is experienced as an anticipation of the stable tone. An Indian classical musician who builds up a strong expectation for resolution to the tonic and then teases the audience by circling around the tonic and delaying the resolution is using motor sequences to produce patterns of sound, all in the service of eliciting the experience of expecting resolution.

3. Affective experience

Emotion often accompanies our listening experiences, and is often the reason why we choose to listen to music. The range of this emotional experience is varied. Sometimes music allows a listener to identify emotion, to proclaim, "This is a sad song", but without actually causing the listener to experience sadness. Sometimes the communication of emotion is accompanied by an emotional response, and other times it is not. Sometimes, music may cause feelings that resemble emotions, but these feelings may be difficult to capture with emotional terminology (see Raffman, 1993), and are perhaps even more difficult to investigate scientifically. Thus, we have opted to focus our discussion of music and emotion on experiences that can be described with emotional terminology, although emotional responses to music are certainly not limited to those that can be captured with such terminology. In this section, we review various theories that have been offered to explain why music induces emotions, and how music may have been mapped to emotions.

3.1. Emotional responses to music

3.1.1. Characterizations

Emotional responses to music begin early in life, and grow increasingly more sophisticated with cognitive development. Infants exhibit an affective response to music by 4 months of age (Zentner & Kagan, 1996, 1998). Identification of emotion in music develops with increasing emotional specificity until it reaches adult-like sophistication, often by the age of six (e.g., Cunningham & Sterling, 1988; Dolgin & Adelson, 1990; Kastner & Crowder, 1990; Kratus, 1993; Terwogt & van Grisven, 1988, 1991).

Listeners find it natural to ascribe emotion to music, generally agree about the emotion conveyed by a particular piece, and can identify specific emotions with remarkable accuracy (Juslin & Laukka, 2003). Listeners identify musical emotions quickly, and can distinguish happy from sad pieces upon hearing as little as a quarter of a second of music (Peretz, Gagnon, & Bouchard, 1998). Human expressions of emotion are processed implicitly (Niedenthal & Showers, 1991), and there is evidence that this implicit processing extends to musical expressions of emotion (Peretz et al., 1998). Listeners may perceive emotion in music without conscious awareness of what has caused the perception.

3.1.2. What is an emotion?

Before further examining emotions in the musical realm, we will briefly review the topic of emotions in general. We begin by differentiating between the various terms that are used to refer to classes of emotional experience.

Affect is a general term that is often used to refer to a wide range of emotional experiences. Affective valence refers to whether an emotional experience is construed as positive or negative, and is thought to precede the ontogenetic development of more nuanced affective states (Harris, 1989). Arousal is a general term that refers to the intensity of an emotion.

Emotions are thought to be adaptive responses that bias action in ways that benefit survival. Many researchers believe that a trigger – an event or object – is a necessary elicitor for emotions to occur. The cognitive appraisal of the object or event induces emotions, which are accompanied by various physiological and phenomenological responses that collectively bias action and behavior. Emotions are often conceptualized according to the 'categorical' approach, which reflects the theory that there are a few universal basic categories of emotion, each of which has a unique adaptive function (see Ekman, 1992; Izzard, 1977; Plutchik, 1994; Tomkins, 1962). For instance, happiness, anger, and sadness can be evoked by the appraisal of events that have significance for the well-being of the individual. Achieving a goal may elicit happiness, having a goal thwarted may elicit anger, and failing to achieve a goal or losing the ability to do so may elicit sadness.

Moods do not seem to require an eliciting event or object, and tend to occur with less intensity than emotions, but for a longer duration. While emotions and moods may both influence behavior (e.g., Scherer, 2000), emotions are thought to bias action, while moods are thought to influence cognitive behavior (Davidson, 1994). Moods can influence memory, decision making, and evaluations (e.g., Bless & Forgas, 2000).

Music can cause mood changes and emotional responses. However, examinations of the response to music usually do not attempt to determine whether a particular response should be referred to as an emotion or mood. In the following, we use emotional terminology to refer to both emotions and moods, since there is usually very little information to allow us to assess whether 'feeling sad' refers to a brief and intense experience of sadness, or a less-intense but lingering mood. We also use less specific terminology, such as affective valence and arousal, to refer to commonalities between different emotions (see Russell, 1980).

3.1.3. Measuring emotion in music

Although we often identify the emotion that is induced by music in terms of the emotions that are induced by non-musical events, it is unclear that these are true emotions. Can the sadness induced by music reasonably be compared to the sadness that is induced by events that have true impact, such as personal loss? In terms of the antecedents and consequences, these two forms of sadness cannot be compared. But, are they nonetheless phenomenologically similar?

Scherer (1993) has proposed that emotions involve a "reaction triad", manifested in one's subjective feelings, expressive behavior, and physiological reactions. There is evidence that music meets these criteria. Evidence of self-reported emotional responses to music has been provided by numerous studies (e.g., Krumhansl, 1997; Pike, 1972; Sloboda, 1991; Sloboda, O'Neill, & Ivaldi, 2001). Crying is an expressive response to music (e.g., Gabrielsson, 1991; Waterman, 1996), and was reported to have occurred in 90% of participants in one survey (Sloboda, 1991). Physiological reactions to music have also been reported (Bartlett, 1996; Krumhansl, 1997; Nyklicek, Thayer, & van Doornen, 1997; Witvliet & Vrana, 1996). These reactions – including changes in heart rate, blood pressure, skin conductance, temperature, and respiration – are similar to those that accompany non-musical emotions. Thus, it does seem that emotional responses to music are similar to emotional responses to goal-relevant stimuli.

3.1.4. Do musical emotions bias adaptive responses?

It is generally accepted that emotions yield adaptive responses, which makes one question the emotional response to music. Music is not seen as a stimulus that should require an adaptive response, nor is it clear that it evokes such a response, so it is difficult to reconcile the emotional response to music with theories of emotion predicated on adaptive responses. A recent model of emotion does not assert that all emotional responses are created equal, and posits that each stimulus is appraised according to its own merit. Under Scherer's (2001) model of appraisal criteria, a stimulus that generates an emotional response does so as a result of its unique appraisal according to several criteria. These criteria include appraising the probability of encountering the stimulus, the cause of the stimulus, its inherent pleasantness, its relevance to one's goals, the most probable outcome, whether one will be able to cope with the outcome, and the urgency of an adaptive response. The unique evaluation of a stimulus according to these (and other) criteria should determine the specific nature of the emotional response, resulting in a unique emotion for each elicitor of emotion (Scherer, 2001). The appraisal theory explains the concept of basic emotions not in terms of basic function, but in terms of similar appraisals. The appraisal theory holds that the response is determined by the appraisal, and that similar stimuli that have similar relevance to one's goals may be appraised similarly, and the emotional responses will be similar, allowing us to categorize emotions sharing a family resemblance structure as basic emotions. This theory offers an appealing explanation for the subtle differences between certain emotions that we tend to lump into the same basic category. For those attempting to account for emotional responses to music with a general theory of emotion, the appraisal theory offers a solution that does not imply that we need to make sense of our emotional responses to music in terms of our emotional responses to non-musical stimuli, given that the unique nature of musical stimuli will elicit unique emotional responses.

3.2. Why do we respond emotionally to music?

When a listener responds emotionally to music, the particular emotion experienced may be influenced by many variables. Personal memories that are attached to a song, as well as one's musical tastes play an important role in the emotional response, but do so in a manner that is extremely variable between listeners. Musical structure, on the other hand, yields identifications of emotion that are consistent between listeners. Although we do not want to overlook the importance of other sources of musical emotion, we have chosen to focus specifically on musical structure as a source of emotion, because it is the only source of emotion that is common to all listening experiences.

3.2.1. Structural sources of emotion

Musical structure functions as a code for communicating emotion. It is generally thought to do so in two ways: through *intrinsic* properties of the music (i.e., how a musical event is construed in relation to its musical context), and through structures that are mapped to the *extrinsic* domain of emotion (Meyer, 1956). We will consider the intrinsic and extrinsic sources in turn.

Intrinsic sources of emotion are specific to the musical structure, without reference to anything outside of the music (see Meyer, 1956). Intrinsic generation of emotion relies on one's schematic knowledge of a musical system. Once a listener has implicitly acquired the rules of the musical system (such as transitional probabilities and tonal hierarchies), one is able to appreciate music according to its adherence to or deviation from the schematic norms. Tonal hierarchies establish a source of stability. The tonic of a key is the most stable musical unit, and is a point of reference by which to measure the stability of other musical units. Departure from this point of stability may create a feeling of musical tension, and a return to this point of stability may feel like the tension is being released (for a formalization of this theory, see Lerdahl & Jackendoff, 1983; Lerdahl, 2001). This intrinsic source of emotion, creating an ebb and flow of musical tension, may not have a particular emotional valence (Sloboda & Juslin, 2001), but can be conceptualized along the dimension of arousal, contributing to the intensity of one's emotional experience.

Each musical unit (such as a note or chord), has a transitional probability of being followed by other musical units. Based on our implicitly acquired knowledge of transitional probabilities, we expect notes and chords to be followed by other notes and chords that – based on our prior experiences – we know have a high probability of occurring next. We calculate these probabilities implicitly as we listen to music, and have an implicit (and sometimes explicit) expectation that specific notes or chords will occur. When our expectations are violated, we may have a conscious experience of knowing that a musical event was not what we expected, even if we were unaware that we had been expecting anything in the first place. Meyer (1956), in his seminal theoretical account of musical meaning, cites these implicit musical expectations – and their fulfillment or violation – as the primary elicitor of emotional responses to music.

Empirical work has validated the implicit generation of musical expectations, showing that listeners cognitively represent tonal hierarchies and transitional probabilities (e.g., Bharucha, 1987;Krumhansl, 1991). Theories of emotion include the assessment of stimulus probability (i.e., conformity to expectations) as part of the appraisal process influencing the emotional response (e.g., Scherer, 2001). Empirical assessment of musical events that elicit emotional responses offers strong support to Meyer's theory of musical expectancy. Sloboda's (1991, 1992) investigations of strong emotional responses, such as weeping or experiencing chills, reveal that aspects of musical structure integral to the generation, maintenance, fulfillment, or violation of musical expectations (such as enharmonic changes and syncopation) are associated with the elicitation of these emotional responses.

Sloboda and Juslin (2001) identify the limitation of intrinsic sources to elicit emotional responses. They regard the types of responses associated with intrinsic sources – surprise, fulfillment, tension, and release – as 'proto-emotions', because they lack emotional specificity, such as valence. The content of our emotional experience is provided by the structural features of the music that have been mapped to concepts extrinsic to the realm of music: specific emotional states.

Musical structure is represented at numerous levels, and many of these individual levels of structure are mapped to the dimensions of valence and arousal. Multiple levels of structure can combine to convey specific emotional states. Many investigators have attempted to determine the qualitative emotional contributions made by specific variables, as well as the type of correlations that exist between the structural variables and the subjective affective assessments. Experiments involving the systematic manipulation of one variable at a time, in an otherwise unchanging musical context, have yielded data reflecting the general contribution of several variables to the conveyed emotion. We will not attempt to review this literature (for a review, see Gabrielsson & Lindström, 2001), nor will we review the influence that a performer may have on conveying emotion (for a review, see Juslin, 2001). Instead, we offer a simplified assessment of some structural variables that are associated with dimensions of affect and arousal.

Arousal denotes the intensity, or energy level, of an emotion, and is most easily conceptualized along a continuum. Anger, happiness, fear, and excitement have higher levels of arousal than sadness, tranquility, pleasantness, and seriousness. Arousal seems to be positively associated with pitch height, dissonance, sound intensity, and tempo (e.g., Costa, Bitti, & Bonfiglioli, 2000; Hevner, 1937; Maher, 1980; Maher & Berlyne, 1982).

Happiness, pleasantness, tenderness, and tranquility have a positive affective valence, while sadness, anger, and fear have a negative valence. In music, valence is positively associated with consonance and harmonic simplicity (Hevner, 1936; Smith & Williams, 1999). The major mode is construed as having a positive valence, and the minor mode as having a negative valence (Kastner & Crowder, 1990).

The structural distinctions between happy and sad music have received much attention. Tempo differentiates these two emotions, as does pitch height. Happiness is associated with faster tempos and higher pitches than sadness. The major and minor modes are strongly associated with happiness and sadness in Western tonal music. The happy-major/sad-minor distinction has been documented in numerous studies (e.g., Crowder, 1985; Gerardi & Gerken, 1995; Gregory, Worrall, & Sarge, 1996; Hevner, 1935a, 1935b, 1936, 1937; Krumhansl, 1997; Nielzén & Cesarec, 1982; Peretz et al., 1998; Wedin, 1972). Mode may be the most important structural feature used in the differentiation of these emotions in adult listeners, as indicated by the weights of these factors relative to others in Hevner's (1937) assessment of the results from several experiments.

3.2.2. What is the origin of the mappings between musical structure and emotion?

Various theories have been proposed to account for the mappings between musical structure and emotions. We focus on two particular theories, since they both regard the perception of emotion in music as inherently functional.

The acquisition of mappings between acoustic structures and emotions may begin with early interactions between mother and infant. Newborn infants show a predisposition to identify and respond to the contours, rhythms and movements of their mothers' expressive vocalizations (Malloch, 1999; Robb, 1999). Mothers speak to their infants in a special manner known as *motherese*. Strikingly different from adult-directed speech, the infant-directed motherese may have a closer semblance to singing, marked by slow, exaggerated pitch contours. Infants respond to subtle changes in their mothers' vocalizations (involving vocal timbre, timing, contour, and volume) by adjusting their own vocalizations to match (Beebe, Jaffe, Feldstein, Mays, & Alson, 1985; Papoušek, 1989; Papoušek & Papoušek, 1991; Trevarthen, 1993, 1999; Trevarthen & Aitken, 1994). These vocalizations have musical qualities, but serve an important communicative function. It is essential that mothers and infants develop a communicative code, so the infant can alert the mother to its biological needs. The mother can, in turn, use this communicative code to regulate the infant's emotional state. The musical quality of these expressive vocalizations indicates, as pointed out by Papoušek (1996), that the origin of music's expressiveness may be traced to the system of emotional communication between mothers and infants.

Motherese is not the only vocal method employed by mothers to communicate with their infants. Caregivers throughout the world sing to infants to regulate their emotion and capture their attention (e.g., Trehub & Trainor, 1998). Infants prefer infant-directed singing to adult-directed singing, and seem to be predisposed to this preference; Masataka (1999) reports this preference in 2-day-old infants who, having been born to congenitally deaf parents, had no previous exposure to singing. However, infants do not confuse infant-directed singing with the song-like motherese. They prefer infant-directed singing, and pay more attention to their mothers' infant-directed singing than to her infant-directed speaking (Trehub, 2001). These findings suggest that music also plays an important role in the emotional communication between mothers and infants, and that infants are biologically predisposed to this form of communication.

Juslin (1997, 2001) offers a theory that also relates music's expressiveness to the vocal communication of emotion. He proposes that the cognitive basis of music's association with expressiveness may lie in neural systems that automatically detect emotion in vocal expression. The acoustic cues that are used to detect vocally expressed emotion may be automatically detected in the acoustic structure of music, and would be identified as expressing the same emotion in music that they would express in human vocalizations.

There are numerous similarities between the vocal and musical expression of emotion. Juslin and Laukka (2003) reviewed the acoustic parameters that are used in each domain to express emotions, and found that tempo (speed of articulation), intensity, and pitch height have similar communicative functions in both domains. Other correspondences were also noted. These similarities lend plausibility to the hypothesis that music is perceived as expressive due to its similarities with vocal expression. Juslin (2001) suggests that vocal expression of emotion is processed automatically by brain modules that detect certain stimulus features (e.g., Fodor, 1983).

These modules may not differentiate between different classes of acoustic stimuli (such as vocalizations and music), but respond automatically to acoustic features that have emotional relevance.

Neuroimaging studies have the potential to test the validity of this hypothesis. Experiments that use vocal and musical expressions of emotion as stimuli within the same subjects could test whether the same neural regions respond to acoustic parameters that are thought to convey emotion in both domains. Parametric modulation of intensity, tempo, and pitch height in both domains may correlate with activity in neural regions involved in the automatic detection of emotionally relevant stimuli. Such findings would support Juslin's (1997, 2001) theory of music's expressiveness.

3.2.3. How do we go from recognizing an emotional expression to experiencing the emotion?

When hearing a musical expression of emotion, the listener often assimilates the expressed emotion into their own emotional state or mood. This response had been explained under social contagion theory (e.g., Juslin, 2001; Sloboda & Juslin, 2001); people seem to catch the emotions that are observed in others. Although emotions do seem to be contagious in social situations (e.g., Schacter & Singer, 1962; Sorce, Emde, Campos, & Klinnert, 1985), it is curious that musical emotions are so easy to catch. Unlike other types of contagions, the germs of emotion transmitted by music seem to require no social interaction – musical emotions are airborne contagions.

The social contagion of emotions is thought to stem from the tendency to automatically mimic the social cues of others, such as body posture, movement, facial expressions, and vocal expressions. It is perhaps the latter that leads to social contagion in music. If music is perceived as expressive due to its similarity with vocal expressions, listeners may be moved to mimic these vocal expressions, perhaps subliminally, or by actually singing or humming along. This motor mimicry may elicit emotions, perhaps through proprioceptive feedback, spreading the emotion associated with the motor response to other response-components associated with the emotion, as suggested by Scherer and Zentner (2001).

Expressive music can induce facial expressions in listeners, even if they only exist as subliminal manifestations, as indicated by studies using facial electromyography (Witvliet & Vrana, 1996; Witvliet, Vrana, & Webb-Talmadge, 1998). There is evidence that facial expressions, even when produced in the absence of the emotions they express, may actually induce subjective and physiological emotional responses (Ekman, Levenson, & Friesen, 1983). These physical responses may drive our emotional responses to music.

3.3. Do musically induced emotions have functionality?

Given that emotions are thought to be adaptive responses that increase one's chances of survival, it seems relevant to wonder if musically induced emotions are functional. It has been suggested that music serves no evolutionary function, and is the equivalent of 'auditory cheesecake' – a concoction that appeals to various mental faculties, but is superfluous to our biological needs (Pinker, 1997, p. 534). However, music evokes emotional responses, and emotions are adaptive. Could music's ability to elicit emotions be adaptive?

Any functionality that music may have is probably reflected in the very reasons people feel drawn to music, and it seems that the experience of emotion is particularly compelling. Young adults and adolescents cite the experience of emotion as one of the central reasons that they listen to music (Panksepp, 1995; Wells & Hakanen, 1991). Adults intentionally use music for mood regulation, often to change a mood induced by a prior context, induce a mood to prepare for a future event, or to vent an emotion (DeNora, 1999; Sloboda, 1999). These data support the theory that music serves the function of mood optimization, as suggested by Zillman (1988). While it would perhaps be premature to assert that this function is evolutionarily adaptive, we can nonetheless claim that mood optimization may confer survival benefits. Mood influences aspects of cognition (see Bless & Forgas, 2000). Negative moods have been shown to negatively affect cognitive performance and learning (Koester & Farley, 1982; Kovacs & Beck, 1977), but positive moods have been shown to positively influence cognitive performance (Ashby, Isen, & Turken, 1999; Isen, 1999). Music clearly has the power to influence mood, and musical mood manipulations have been shown to influence performance on cognitive tasks (e.g., Kenealy, 1988; Thompson, Schellenberg, & Husain, 2001). When a listener uses music to turn a negative mood into a positive mood, their cognitive performance may benefit as a result. Music has the power to induce happiness that – unlike prototypical experiences of happiness – is not contingent on making progress towards a goal. However, music may reverse the typical direction of this causal relationship. Musically induced happiness may help one progress towards a goal, by virtue of its potential to positively influence cognition.

4. Experience of motion

Music often elicits an experience of motion. This includes rhythmic movements one seems compelled to make while listening, but also bodily sensations of motion and more abstract conceptions of motion. Here we focus on internal experiences of motion and we identify four primary facets of such experience: a fundamental sense of self-motion through space; a perception of the motion of bodies (though not necessarily one's own), including imitative percepts; a sense of movement and causality derived from more abstract structural elements such as tonality; and a conceptual sense that utilizes the other senses to dramatize the music through metaphor. For these various kinds of motional experience, there are likely different aspects of the music affecting different neural systems. In fact, the first facet is an example (proposed in the Introduction) of the acoustic code largely bypassing the normal musical mechanism for eliciting experiences, namely implicit structural knowledge, while the latter three are coupled with an increasing structural hierarchy in the music. A complete account of motional experience must therefore offer a description of cognitive neuroscientific research on these phenomena as well as a correlation between the psychology and the neuroscience. In what follows, we lay the foundation for these phenomena.

4.1. Vestibular self-motion

Truslit (see Repp, 1993) argues that ‘inner motion’ is a core experience of music. Its two basic components are a basic bodily sense of motion, such as movement of the head or trunk, and a gestural sense related to limb activity. He distinguishes inner motion from other types of motion that can overwhelm this sometimes ineffable sense, such as overt physical movements used to perform music, and a conceptualization of motion one might forge in the imagination. The task of the composer, he argues, is to communicate his own sense of inner motion to an audience through the music, and this is accomplished largely through the pitch and rhythmic components that constitute melodic contour. The motion of which Truslit speaks is emphatically not metaphorical but rather physiologically induced, and he attributes it to the vestibular system, which is indeed involved in the perception of self-motion.

Conventionally it is thought that the cochlea, as an exteroceptive sense organ, is the point of transduction for acoustic information in the environment. By contrast the vestibulum is a proprioceptive organ that detects changes in the spatial state of the organism’s body. However, research on a variety of animals has shown that parts of the vestibulum, namely the saccule and lagena, are also sensitive to auditory stimuli (Todd, 1993; and see Todd, 2001). These tiny organs in the inner ear are sensitive to forces of acceleration and help animals discern translations in spatial position. Humans do not have lagenae, but nerves from our saccule project through intermediaries to spinal motoneurons, establishing a pathway by which acoustic stimuli could influence the spine and thus create a compulsion to move to music in addition producing a sense of inner motion.

Some literature on humans bears out this possibility. Colebatch, Halmagyi, and Skuse (1994) and Todd (1999, 2001; Todd and Cody, 1999) have demonstrated the vestibular role in an evoked myogenic response to loud clicks. The critical response occurs in the target muscle even in patients with bilateral sensorineural deafness, but not in patients with ipsilateral vestibular deficits, suggesting the response is vestibular and not cochlear. Loud percussive sounds evoked very similar (though somewhat more complex) myogenic responses, and these sounds were of intensities similar to those found in rock concerts and dance clubs: by 120 dB, 90% of the subjects had reached threshold for the EMG components. This research has interesting implications for the nature of musical experience. If true, the saccule’s ability to respond to a basic aspect of acoustic structure is being used by composers and performers to stimulate listeners in a way that does not rely on the listeners’ implicit structural knowledge of music. And while the purpose of this stimulation is not entirely clear, one might speculate, as Todd and Cody (1999) do, that a sense of self-motion is pleasurable, just as swings or amusement park rides induce pleasure partly by virtue of their effect on the vestibular system.

However, other literature on the structure and function of the saccule calls into question its ability to respond to acoustic information. Highstein (2004) argues that the vestibular system has a complicated evolutionary history, and that though this system in many fauna has a similar organization, the actual function of the organs can vary across species. Thus while the saccule and lagena respond to acoustic information in fish, the saccule in humans responds only to forces of linear acceleration. Furthermore, the structure of the human saccule does not appear to be well-suited to transducing sound energy (Rabbitt, Damiano, & Grant, 2004). As yet there is no account of how the human saccule might transduce acoustic energy, allowing composers and performers to elicit experiences of self-movement.

Clearly, more research must be done on this appealing hypothesis to bear out in full the effect the vestibular system, and in particular the saccule, has on our experience of music. For instance, it is necessary to describe a mechanism by which the human saccule could respond to sound. One important study would be to determine if the self-motion purportedly induced by music is primarily in the plane to which the human saccule is responsive: the saccule is a flat organ that responds to linear acceleration in the transverse plane; the utricle, another flat otolithic organ, responds to linear acceleration in the sagittal plane (Rabbitt et al., 2004). Another strategy is to examine effects elsewhere in the brain, perhaps in the cerebellum and the basal ganglia, which are involved in corporeal dynamics and their timing attributes. Because these brain regions are connected with the saccule (Rabbitt et al., 2004), they may be active in response to vestibular acoustic stimulation. A study might look at the timing of cerebellar or basal ganglia activation in relation to the saccular myogenic evoked responses. Furthermore, it would be useful to determine whether this sense of motion can be induced from simply imagining music, when no stimulus is directly activating the saccule. Until results from such studies are reported, the hypothesis that vestibular activation contributes to a sense of self motion during music listening will continue to be speculative but promising in the study of musical experience.

4.2. Sound sources: Perceptual specification and imitation

The motion one experiences in music is not limited to the constraints imposed by the properties of the saccule, so other mechanisms are required to explain other experiences. One of these experiences is the sense of a piece of music as sounding like a moving object. Research on music dynamics (intensity modulation) and agogics (durational modulation) has shown that these properties conform to the nature of physical force mechanics and thus may account for that sense of movement. For example, models for tempo change (i.e., accelerando and ritardando), one aspect of agogics, have been proposed for decades. The equations used in the models are typically quadratic, each one uses different equations on different parameters of different musical excerpts, and they describe musical tempo change as accurately as they describe linear acceleration (Sundberg & Verillo, 1980; Kronman & Sundberg, 1987; Todd, 1992, 1995; Feldman, Epstein, & Richards, 1992).

For instance, Kronman and Sundberg (1987) proposed a model for describing deceleration (ritard). They used the inverse of tone duration, or *instant tempo*, to achieve a “generalized retardation function”. This quadratic equation describes the transition toward a final beat, but its parabolic curve models gravity equally well, with velocity corresponding to the number of musical meters per second. Alternatively, Todd (1995) conceives of music as mapping onto a two dimensional grid of pitch space and metrical position, and his subsequent regression analyses plot the latter against beat onset times. In this way he develops acceleration and velocity variables that describe estimated tempo quadratically, while allowing identical metrical times to yield different performance times, an experience familiar to most music listeners. Using another measurement for tempo change, Feldman et al. (1992) describe parabolic force analogs in ritardando that appear between two sections of constant tempo instead of at the end of a phrase. Their regression analysis of beat duration against beat number yielded excellent fits (r^2 of .975 in one case) of cubic equations for the five musical excerpts they studied. While it remains unclear what specific temporal and musical parameters the brain uses to hear phenomena like accelerando and ritardando, it is clear that the brain is using *some* mechanism to extract information about dynamics from the incoming sound stream, because our experience of those dynamics is well-modeled by these equations using different forms of this information.

As mentioned previously, these quadratic and cubic equations accurately describe not only our experience of musical dynamics, but also the mechanics of linear and quadratic acceleration, respectively. This suggests that the brain detects perceptual invariances in the music that reliably convey information about physical movement, so that hearing a ritardando as a slowing object is not merely an exercise in metaphor. This is an ecological perspective, first described in the visual domain by Gibson (1979) and applied to audition by Bregman (1990). Physical properties of sounds streams, such as pitch height, event density, timbre, or intensity, as well as changes in any of these, specify objects in space by virtue of what one has learned about the world perceptually. One literally hears the direction or weight of an object based on the sound energy it produces in the same way one would see the direction or feel the weight based on its light reflectance or mass properties, respectively.

Clarke (2001) follows in this tradition, arguing that musical sounds are not fundamentally different than sounds in the natural world. In fact, precisely because they do originate in the natural world, they are subject to the same scene-analysis processes as ordinary sounds (Bregman, 1990; Clarke, 2001), and we form similar expectations about them. When these same general perceptual principles are applied to music, one hears the music as specifying objects as its source. This occurs in two different forms. According to Shove and Repp (1995), the source specified by music is the literal source, the performer. The object specified is thus the human and instrument, and the movement one hears in the music is the movement undertaken by the performer. Alternatively, Clarke argues that music can specify other, illusory objects in a virtual environment analogous to the virtual environment that a painting creates. If the streams segregated by the auditory system (in the case of music, the different lines of music) cohere and are well-coordinated, one is aware of a single body specified, and typically that body is oneself moving in relation to the sound sources. (This illusion is still distinct from the feeling of movement described in the previous section.) On the other hand, complex lines of music that remain distinct are perceived as external objects with components moving in relation to each other (Clarke, 2001). Thus, physical properties of the acoustic stimulus determine the type of motion experienced (self or other) and the qualities of that motion. And sudden changes in tempo or dynamics violate expectations because one implicitly accepts that smooth motion should persist, as it does in the real world.

An alternative possibility is that certain external sounds are experienced in terms of self-produced sounds. Not all listeners understand the performance movements required to produce those musical sounds, and neither must motional experience be of an object moving through (virtual) space. Cox (2001) proposes a mimetic hypothesis: that listeners experience music not as the movements of performers nor as virtual objects, but rather in terms of the behaviors one believes one would have to execute to create those sounds. This process of imitation involves comparing the sounds of music with the same or similar sounds one has produced previously. Cox argues that some types of musical motion are more abstract than imitative behaviors and concludes that conceptual metaphor is the mechanism for translating music into mimesis. However, we argue that there are indeed aspects of the physical stimulus that can specify motion in, for example, melodic gesture, as well as physical mechanisms for translating those aspects into motion. The more abstract level of musical motion to which he refers can be accounted for by a mechanism intermediate to perceptual specification and conceptual metaphor, which we explore in the next section. To leap all the way to conceptual metaphor to explain this phenomenon would be to ignore an important mechanism by which music moves us: its relationship to natural sounds.

One example of mimesis is the subliminal facial gesturing people exhibit in response to music (Witvliet & Vrana, 1996; Witvliet et al., 1998). This research is discussed in Section 3.2, "*How do we go from recognizing an emotional expression to experiencing the emotion*?" in the context of emotions. Facial responses can help induce emotions (Ekman et al., 1983), perhaps through proprioceptive feedback (Scherer & Zentner, 2001). Expanding on this idea, these facial expressions also have motional and spatial properties – they can change quickly or slowly, be tense or loose, round or flat. Because the facial expressions occur in response to the music, the listener perceives the music as the cause of both the expressions and their attendant sensations of movement.

For example, a pair of notes that goes from a moderately low pitch of moderate volume to a high pitch of high volume might convey a sense of tension and restriction precisely because the face, including the vocal apparatus, must go from relaxed and positioned low to tense, constricted, and positioned high. The face is effectively translating the music into corporeal motion. This phenomenon would correspond to low level aspects of music because these are the parts of the stimulus that could be vocally imitated; vocalizing a chord progression is impossible because of the pitch limitations of the apparatus. The forgoing account, though complex, lends itself to empirical investigation better than an account based only on metaphor.

This aspect of motional experience requires learning over many years, including experience in singing, and having some degree of motor coordination is a minimal requirement. One must also have produced enough sounds to be able to correlate the movement that produces those self-sounds with external sounds, so as to establish a concept of self movement that would produce that sound. Of course, music must be discovered and learned culturally too, and its contemporary property of perceptual specification is no different. It could not always specify the approach or acceleration of a body; that capacity had to be invented, just as linear perspective was invented long after humans had begun to draw and paint. And the techniques used to convey perceptual effects may be different than those effects themselves. Clarke (2001) gives the example of different pigment hues being used to convey an effect that would be caused by differential reflectance in reality, and there are surely examples of this in music as well. But it is important not to mistake as metaphor the correlation between the physical technique and the effect it seeks to replicate simply because the two might be different, and because the technique had to be created whereas the original stimulus is given. The perceptual systems rely on structural regularities in the physical world – natural eliciting codes, if you will – to pick up the percepts we experience. Close approximations of those features can be produced by more than one source.

4.3. Musical objects and abstract motion

Often the motion we hear in music cannot directly be attributed to an external source. Many listeners hear a certain naturalness or inevitability in the course of a piece of music, changes in the structure that seem to have the reality of actual spatial movement. Yet a transition from a leading tone to the tonic has no sensible correlation in the physical world: no sound source produces a leading tone just prior to its goal and a tonic once its goal is realized. So is talk of this kind of attraction, or of a piece of music as going on a journey, just a metaphor? That is, are we only mapping concepts from one domain (physical space) to another?

Research in music theory and cognition over the past twenty years suggests that we are not only using metaphors. While perceptual specification relies on low-level aspects of the eliciting code, it is the most paradigmatic properties of the code, those which truly characterize music as a code, that elicit the experiences of motion now under discussion. Music *structure*, especially melody, harmony, and tonality, has the properties of a mathematical space, but it also maps onto a psychological space (Krumhansl, 1991; Lerdahl, 2003). Moving through this abstract space means exploring its implications. While it has no counterpart in a physical space, neither is it just a metaphor for physical space: it is a conscious representation that functionally captures spatial relationships and movement.

There are two prominent examples of this kind of abstract motion in the field of music psychology. One type of melodic motion is explored by Gjerdingen (1994). Trying to answer the question of why listeners hear a melodic line as a single object unfolding in time, he designed a neural net to mimic the temporal integration that occurs in human audition, and presented it with melodies. What he found was that the output of the model was not a succession of distinct pitches; rather, the net had traces of activity in the intervals between pitches as if it were trying to connect them across time. The model seemed to conform to the human intuition of a melody as a single, connected object.

Another significant facet of abstract movement is the sense of compelled movement through tonal space (Lerdahl, 2001). An acoustic code establishes a location in tonal space, with less proximal events creating tension and more proximal events resulting in satisfaction (Lerdahl, 2001, 2003). Because this representational space is so stable across listeners, it can reliably be used by composers to elicit tension as well as movement through the space. What some may experience affectively as expectation–satisfaction (Meyer, 1956), implication–realization (Narmour, 1990), or yearning (Bharucha, 1996), may also be experienced as a desire to move to a different location in psychological space (Lerdahl, 2001). This space, however, is not metaphoric: it is psychological, and the experience of movement is real.

4.4. Conceptual metaphor

The embodied experience of Cox (2001) and the perceptual specification of Clarke (2001), as well as the spatial properties created by the formal eliciting system itself (Lerdahl, 2003), explain many of the motional aspects of our experience of music. But there is still a gap between the level of detail in our experience of motion accounted for by those theories, and the degree of elaboration, whether of complex scenes or characterizations of personae, many people often ascribe to music. Can the acoustic array specify objects with as much precision as we often seem to experience in music? And how elaborate is our experience of making sounds such that we can directly compare it, for example, to the intricate parts of a fugue?

In describing a passage of music as a sunny stroll through a park, one is applying a metaphorizing process that dramatizes the sound stimuli beyond the percept of the sources it may specify (passing sounds that convey a sense of forward motion) and beyond the capabilities of human auditory mimicry (the sounds we make when relaxed and cheerful). This process relies on concepts that have become associated with sounds through cultural and individual experience. For example, paint on a canvas may perceptually specify light spreading from overhead, but only the conceptual metaphor of 'light is spiritual' allows one to see those paint markings as indicating a divine presence within the scene of the painting. Similarly, a crescendo may specify the approach of an object emitting sound, but only a conceptual metaphor could specify that object in detail. One example is the rich description Clarke (2001) ascribes to certain crescendi in *Wozzeck*; interestingly, even with the context of the opera to supply and constrain metaphors, he is uncertain as to what exactly the crescendi might refer to. Sometimes the full appreciation of music means that metaphors must be even less constrained by the formal code's specifications and even by the context in which one hears it, reinforcing the prominent role of imagination.

This is the approach to musical motion taken by Johnson and Larson (2003), who base their work on the more substantial treatise by Lakoff and Johnson (1999). Their claim, however, is much stronger than for a mere metaphorizing process on top of more physically based ones. For Johnson and Larson (2003), *any* sense of motion in music is entirely metaphorical. Conceptual metaphors such as 'music as landscape' define the experience of a musical passage. In this case, rather than hearing elements of the stimulus that would convey movement through a landscape in the real world, Johnson and Larson argue that listeners use a concept of 'landscape' and transfer to the music the logic of one's physical interaction with a landscape. Thus the music is conceived of as being in some place, and the listener is conceived of as moving through and observing that place, but only as an act of the imagination. As discussed above, Clarke (2001) has argued persuasively that there are fundamental perceptual reasons why one might experience music as a landscape through which one travels. However, it is possible that the conceptualization process begins quickly after perceptual processes have specified some degree of motion. In that case, to the listener it might seem as if the conceptual metaphor were the only component of the musical motion, but one can still distinguish between early and late stages of the motional experience.

4.5. Relations among different types of motion

We have identified four primary components of the experience of motion in music: a fundamental sense of self-motion; a sense of motion derived perceptually from the acoustic stimulus, including an ecological 'pick-up' of acoustic movement information and a tacit mimicry of sounds; a manipulation of formal rules that create a sense of movement through logical or symbolic space; and an imaginative process of applying conceptual metaphors to music. These components are not mutually exclusive; in fact, they are probably hierarchically nested. The vestibular sense of self-motion, if born out, would likely be the most fundamental. It is an uncomplicated sensation elicited by basic aspects of the stimulus, though it may be subtle for many people and overwhelmed by more complex experiences of motion. The perceptual specification and imitative sources of motion may rely on the sense of self-movement to assist in determining whether the subject or object is moving, and in what manner. Both of these experiences may help to draw attention to different aspects of the structure, which influences what experiences of motion the acoustic code actually elicits. Consistency among different conceptual metaphors formulated about music might arise because these other more physiologically and acoustically driven experiences of motion serve to shape and constrain the kinds of metaphors one can possibly apply to a given passage of music.

Not only are the different types of motion related to each other, they are related to other conscious experiences as well. The feeling of a dominant chord moving to a tonic can be simultaneously experienced as movement (of type three), as affect (satisfaction or relaxation), and as structure in itself. If one becomes agitated by a passage of music rapidly increasing in pitch height, density, and intensity, the experience of the musical movement – that of an object quickly approaching (type two motion) – is nearly inseparable from the emotion of fear or alarm. Though not necessarily commingled, many experiences of music can occur together. The acoustic code is complex, with many variables impinging on the auditory (and perhaps vestibular) system at once. Different facets of that code can elicit experiences in distinct domains simultaneously.

Finally, does the conscious experience of motion in music help to explain the near universal enjoyment of orchestrated sound? From a social perspective, collective participation in music or dance can forge or maintain social cohesion of an in-group by synchronizing behavior and the experiences the music and movement evoke. Even when listeners are not active participants, their conscious experiences may still be synchronized. Equally important is the individualistic aspect of music; the demonstration of skill and the ability to influence others through music can set apart group members and mark them as special and distinctive. From a hedonistic perspective, it is clear that humans enjoy deliberate manipulation of their sense of balance and acceleration through space; from the mildest of playground swings to the most terrifying rollercoaster, these devices serve little other purpose, and they are constantly in demand (Todd, 1999). Music may also be popular for the same reason art and optical illusions are: by conforming to some properties of natural sound stimuli and violating others, musical pieces can convey impossible worlds. This can be taken as a weak hypothesis, namely that the sounds of music do things we would never hear in nature, or as a stronger hypothesis, which, following Clarke (2001), is based on the idea that music specifies objects and that these objects might be illusory and behave in physically impossible ways. The question remains what objects are actually moving in music, as Todd (1995) has argued, though this indeterminacy is likely part of music's appeal as well. There is no such thing as motion divorced from an object in the physical world, but in music we may hear motion per se, acknowledging a source but never fully comprehending it.

5. Conclusion

No one aspect of formal musical structure or of musical experience seems to qualify as a necessary condition for calling something music. Music is inherently varied, and the range of things we call music is tied together loosely in a network of family resemblances. In this paper we provide a characterization of music not in terms of purported necessary conditions but in terms of the distinction between formal codes and the variety of experiences they evoke.

The acoustic structure of music, and its auditory and cognitive representations, are formal codes that elicit a variety of conscious experiences. Like formal linguistic structure, formal musical structure has a communicative function – to elicit conscious experiences. However, unlike language, in which the content afforded by formal structure is lexical and propositional meaning, the content derived from formal musical structure is a variety of conscious experiences that may include conscious experiences of structure, affect, and motion. Domains of conscious experience may also have structure, but the domains are often distinct from those of the eliciting codes, except when one is experiencing aspects of the eliciting structure itself. Formal eliciting structures are thus communicative vehicles or media that elicit a variety of domains we report as conscious experience.

We have described a variety of conscious experiences common to music listeners, and we have proposed mechanisms by which aspects of musical structure elicit these experiences. The conscious experience of structure is familiar to most musicians. Non-musicians may struggle to report their experiences of structure, and may be more inclined to describe the emotions and thoughts of movement that are conveyed by the structure.

Music often evokes feelings that we cannot describe with emotional terminology, but sometimes words like anger, fear, happiness, and sadness do seem to capture the essence of the musical expression. The musical codes that convey these emotions are processed implicitly, and are decoded with a high degree of accuracy and between-listener agreement. Although the origin of the mappings between musical structure and the affective domain is uncertain, vocal expressions of emotion seem to be a probable mapping source. Early emotional communication between mothers and pre-linguistic infants has a musical quality, and the pitch contours and temporal parameters of these emotional vocalizations may form the basis of the mappings that convey emotion in music. Listeners may automatically perceive these acoustic patterns as expressions of emotion in both speech and music.

And finally, a sense of motion may not be as prominent an experience as affect for most people, but it is still a salient phenomenon. The first kind of motional experience we enumerate is termed self-motion because it stimulates the vestibular system. The second kind of motional experience falls into two categories: perceptual specification and imitation. In the former case, music seems to be caused by an external source the properties of which we can identify as if a real object were moving in the real world. Our attempt to imitate the sounds we hear in music, and the sensations we experience during such imitation, account for the latter category. Third, the rules and preferences of music constitute a psychological space in which more or less proximal events are situated, and changes within this abstract space can create a sense of moving musical objects. Lastly, few things in life stimulate the imagination like music, and the metaphors we apply to it add much detail and richness.

Acknowledgments

This work was supported in part by a NIH Grant 2P50 NS17778-18. The authors thank the guest editor Isabelle Peretz for her editorial work, and Eric Clarke, Daniel Dennett, Ray Jackendoff, and Fred Lerdahl for valuable comments on earlier drafts.

References

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529–550.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125*, 238–251.

Bartlett, D. L. (1996). Physiological reactions to music and acoustic stimuli. In D. A. Hodges (Ed.), *Handbook of music psychology* (2nd ed., pp. 285–342). San Antonio, TX: IMR Press.

Beebe, B., Jaffe, J., Feldstein, S., Mays, K., & Alson, D. (1985). Interpersonal timing: The application of an adult dialogue model to mother–infant vocal and kinesthetic interactions. In T. Field & N. Fox (Eds.), *Social perception in infants* (pp. 217–247). Norwood, NJ: Ablex.

Bharucha, J. J. (1984). Event hierarchies, tonal hierarchies, and assimilation: A reply to Deutsch and Dowling. *Journal of Experimental Psychology: General, 113*, 421–425.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5*, 1–30.

Bharucha, J. J. (1991a). Cognitive and brain mechanisms in perceptual learning. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain*. London: Macmillan.

Bharucha, J. J. (1991b). Pitch, harmony and neural nets: A psychological approach. In P. Todd & G. Loy (Eds.), *Music and connectionism* (pp. 84–99). Cambridge: MIT Press.

Bharucha, J. J. (1996). Melodic anchoring. *Music Perception, 13*, 383–400.

Bharucha, J. J. (1999). Neural nets, temporal composites, and tonality. In D. Deutsch (Ed.), *The psychology of music*. San Diego: Academic Press.

Bharucha, J. J., & Mencl, W. E. (1996). Two issues in auditory cognition: Self-organization of categories and pitch-invariant pattern recognition. *Psychological Science, 7*, 142–149.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 403–410.

Bharucha, J. J., & Todd, P. (1991). Modeling the perception of tonal structure with neural nets. *Computer Music Journal, 13*, 44–53 (Reprinted in P. Todd & G. Loy (Eds.), *Music and connectionism* (pp. 128–137). Cambridge: MIT Press).

Bless, H., & Forgas, J. P. (Eds.). (2000). *The message within. The role of subjective experience in social cognition and behavior*. Hove, UK: Psychology Press.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge: MIT Press.

Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 115–127.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General, 113*, 394–412.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America, 93*, 3271–3290.

Clarke, E. (2001). Meaning and the specification of motion in music. *Musicae Scientiae, 5*(2), 213–234.

Colebatch, J. G., Halmagyi, G. M., & Skuse, N. F. (1994). Myogenic potentials generated by a click-evoked vestibulocollic reflex. *Journal of Neurology, Neurosurgery, and Psychiatry, 57*, 190–197.

Costa, M., Bitti, P. E. R., & Bonfiglioli, L. (2000). Psychological connotations of harmonic musical intervals. *Psychology of Music, 28*, 4–22.

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae Scientiae, V*(2), 195–212.

Crowder, R. G. (1985). Perception of the major/minor distinction: III. Hedonic, musical, and affective discriminations. *Bulletin of the Psychonomic Society, 23*, 314–316.

Cunningham, J. G., & Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. *Motivation and Emotion, 12*, 399–413.

Davidson, R. J. (1994). On emotion, mood, and related affective constructs. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 51–55). New York: Oxford University Press.

Dennett, D. (1991). *Consciousness explained*. Boston: Little Brown and Co..

Dennett, D. (1998). The myth of double transduction. In S. Hameroff, A. W. Kaszniak, & A. C. Scott (Eds.), *The international consciousness conference: Toward a science of consciousness II, the second Tucson discussions and debates* (pp. 97–107). Cambridge: MIT Press.

DeNora, T. (1999). Music as a technology of the self. *Poetics, 27*, 31–56.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 85*, 503–522.

Dolgin, K. G., & Adelson, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music, 18*, 87–98.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*, 169–200.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221*, 1208–1210.

Everett, D. (forthcoming). Cultural constraints on grammar and cognition in Piraha. Another look at the design features of human language. *Current Anthropology*.

Feldman, J., Epstein, D., & Richards, W. (1992). Force dynamics of tempo change in music. *Music Perception, 10*(2), 185–204.

Fodor, J. (1975). *The language of thought*. New York: Crowell.

Fodor, J. (1980). Methodological solipsism considered as a research strategy in cognitive science. *Behavioral and Brain Sciences, 3*, 63–109.

Fodor, J. (1983). *The modularity of mind*. Cambridge: MIT Press.

Fodor, J. (2000). *The mind doesn't work that way*. Cambridge: MIT Press.

Gabrielsson, A. (1991). Experiencing music. *Canadian Journal of Research in Music Education, 33*, 21–26.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 223–248). New York: Oxford University Press.

Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception, 12*, 279–290.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gjerdingen, R. O. (1994). Apparent motion in music?. *Music Perception 11*, 335–370.

Greenberg, G. Z., & Larkin, W. D. (1968). Frequency-response characteristics of auditory observers detecting signals of a single frequency in noise: The probe-signal method. *Journal of the Acoustical Society of America, 44*, 1513–1523.

Gregory, A. H., Worrall, L., & Sarge, A. (1996). The development of emotional responses to music in young children. *Motivation and Emotion, 20*, 341–348.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: Part I. Parallel development and codings of neural feature detectors. *Biological Cybernetics, 23*, 121–134.

Gulick, W. L., Gescheider, G. A., & Frisina, R. D. (1989). *Hearing*. Oxford: Oxford University Press.

Harris, P. L. (1989). *Children and emotion*. Oxford, UK: Basil Blackwell.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hevner, K. (1935a). The affective character of the major and minor modes in music. *American Journal of Psychology, 47*, 103–118.

Hevner, K. (1935b). Expression in music: A discussion of experimental studies and theories. *Psychological Review, 47*, 186–204.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48*, 246–268.

Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology, 49*, 621–630.

Highstein, S. M. (2004). Anatomy and physiology of the central and peripheral vestibular system: Overview. In S. M. Highstein, R. R. Fay, & A. N. Popper (Eds.), *The vestibular system* (pp. 1–10). New York: Springer.

Isen, A. M. (1999). Positive affect. In T. Dalgleish & M. Power (Eds.), *The handbook of cognition and emotion* (pp. 521–539). New York: Wiley.

Izzard, C. E. (1977). *The emotions*. New York: Plenum Press.

Jackendoff, R., & Lerdahl, F. (in press). The capacity for music: What is it, and what's special about it? *Cognition*, this issue, doi:10.1016/j.cognition.2005.11.005.

Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective, and Behavioral Neuroscience, 2*, 121–140.

Johnson, M. L., & Larson, S. (2003). "Something in the way she moves" – Metaphors of musical motion. *Metaphor and Symbol, 18*(2), 63–84.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459–491.

Juslin, P. N. (1997). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception, 14*, 383–418.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). New York: Oxford University Press.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological Bulletin 129*(5), 770–814.

Justus, T. C., & Bharucha, J. J. (2001). Modularity in musical processing: The automaticity of harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 1000–1011.

Kastner, M. P., & Crowder, R. G. (1990). Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception, 8*, 189–202.

Kenealy, P. (1988). Validation of a musical mood induction procedure: Some preliminary findings. *Cognition and Emotion, 2*, 41–48.

Koester, L. S., & Farley, F. H. (1982). Psychological characteristics and school performance of children in open and traditional classrooms. *Journal of Educational Psychology, 74*, 254–263.

Kovacs, M., & Beck, A. T. (1977). An empirical clinical approach toward a definition of childhood depression. In J. G. Schulterbrandt & A. Raskin (Eds.), *Depression in childhood: Diagnosis, treatment, and conceptual models* (pp. 1–25). New York: Raven Press.

Kratus, J. (1993). A developmental study of children's interpretation of emotion in music. *Psychology of Music, 21*, 3–19.

Kronman, U., & Sundberg, J. (1987). Is the musical ritard an allusion to physical motion. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 57–68). Stockholm: Royal Swedish Academy of Music.

Krumhansl, C. L. (1991). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51*, 336–352.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How we track time-varying events. *Psychological Review, 106*, 119–159.

Leman, M. (1995). *Music and schema theory – Cognitive foundations of systematic musicology*. Berlin: Springer.

Lerdahl, F. (2001). *Tonal pitch space*. Oxford: Oxford University Press.

Lerdahl, F. (2003). Two ways in which music relates to the world. *Music Theory Spectrum, 5*, 367–373.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge: MIT Press.

Luce, R. D., Green, D. M., & Weber, D. L. (1976). Attention bands in absolute identification. *Perception and Psychophysics, 20*, 49–54.

Maher, T. F. (1980). A rigorous test of the proposition that musical intervals have different psychological effects. *American Journal of Psychology, 93*, 309–327.

Maher, T. F., & Berlyne, D. E. (1982). Verbal and exploratory responses to melodic musical intervals. *Psychology of Music, 10*, 11–27.

Malloch, S. N. (1999). Mothers and infants and communicative musicality. *Musicae Scientiae*, 29–54 (Special Issue 1999–2000).

Masataka, N. (1999). Preference for infant-directed singing in 2-day-old hearing infants of deaf parents. *Developmental Psychology, 35*, 1001–1005.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Mondor, T. A., & Zatorre, R. J. (1995). Shifting and focusing auditory spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 66–79.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication–realization model*. Chicago: University of Chicago Press.

Niedenthal, P. M., & Showers, C. (1991). The perception and processing of affective information and its influences on social judgment. In J. P. Forgas (Ed.), *Emotion and social judgments* (pp. 125–143). Oxford, UK: Pergamon Press.

Nielzén, S., & Cesarec, Z. (1982). Emotional experience of music as a function of musical structure. *Psychology of Music, 10*, 7–17.

Nowak, M. A., & Komarova, N. L. (2001). Toward an evolutionary theory of language. *Trends in cognitive science, 5*, 288–295.

Nyklicek, I., Thayer, J. F., & van Doornen, L. J. P. (1997). Cardiorespiratory differentiation of musically-induced emotions. *Journal of Psychophysiology, 11*, 304–321.

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

Panksepp, J. (1995). The emotional sources of "chills" induced by music. *Music Perception, 13*(2), 171–207.

Papoušek, M. (1989). Determinants of responsiveness to infant vocal expression of emotional state. *Infant Behavior and Development, 12*, 505–522.

Papoušek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège & J. A. Sloboda (Eds.), *Musical beginnings* (pp. 88–112). Oxford, UK: Oxford University Press.

Papoušek, H., & Papoušek, M. (1991). Innate and cultural guidance of infants' integrative competencies: China, the United States and Germany. In M. H. Bornstein (Ed.), *Cultural approaches to parenting* (pp. 23–44). Hillsdale, NJ: Erlbaum.

Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhume, V., et al. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*, 185–191.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition, 68*, 111–141.

Pike, A. (1972). A phenomenological analysis of emotional experience in music. *Journal of Research in Music Education, 20*, 262–267.

Pinker, S. (1997). *How the mind works*. New York: Norton.

Plutchik, R. (1994). *The psychology and biology of emotion*. New York: Harper Collins.

Povel, D. J. (1984). A theoretical framework for rhythm perception. *Psychological Research, 45*, 315–337.

Rabbitt, R. D., Damiano, E. R., & Grant, J. W. (2004). Biomechanics of the semicircular canals and otolith organs. In S. M. Highstein, R. R. Fay, & A. N. Popper (Eds.), *The vestibular system* (pp. 153–201). New York: Springer.

Raffman, D. (1992). Proposal for a Musical Semantics. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication*. Washington, DC: APA Press.

Raffman, D. (1993). *Language, music and mind*. Cambridge, MA: MIT Press.

Repp, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's (1938) Gestaltung und Bewegung in der Musik. *Psychology of Music, 21*, 48–72.

Robb, L. (1999). Emotional musicality in mother–infant vocal affect, and an acoustic study of postnatal depression. *Musicae Scientiae*, 123–151 (Special Issue 1999–2000).

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structures of categories. *Cognitive Psychology, 7*, 573–605.

Rubin, D. C. (1995). *Memory in oral traditions*. New York: Oxford University Press.

Rumelhart, D.E., & McClelland, J.L. (1986). *Parallel distributed processing, Vol. 1*. Cambridge: MIT Press.

Russell, J. A. (1980). A circumplex model of emotion. *Journal of Personality and Social Psychology, 39*(6), 1161–1178.

Schacter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional states. *Psychological Review, 69*, 379–399.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Perception and Psychophysics, 42*, 215–233.

Scherer, K. R. (1993). Neuroscience projections to current debates of emotion psychology. *Cognition and Emotion, 7*, 1–41.

Scherer, K. R. (2000). Psychological models of emotion. In J. Borod (Ed.), *The neuropsychology of emotion* (pp. 137–162). New York: Oxford University Press.

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92–120). New York: Oxford University Press.

Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361–392). New York: Oxford University Press.

Schlauch, R. S., & Hafter, E. R. (1991). Listening bandwidths and frequency uncertainty in pure-tone signal detection. *Journal of the Acoustical Society of America, 90*, 1332–1339.

Shove, P., & Repp, B. H. (1995). Musical motion and performance: Theoretical and empirical perspectives. In J. Rink (Ed.), *The practice of performance* (pp. 55–83). Cambridge, UK: Cambridge University Press.

Sloboda, J. A. (1991). Music structure and emotional response: Some empirical findings. *Psychology of Music, 19*(1), 110–120.

Sloboda, J. A. (1992). Empirical studies of emotional response to music. In M. Riess-Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 33–46). Washington, DC: American Psychological Association.

Sloboda, J. A. (1999). Everyday uses of music listening: A preliminary study. In S. W. Yi (Ed.), *Music, mind and science* (pp. 354–369). Seoul, Korea: Western Music Research Institute.

Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In P. Juslin & J. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 71–104). New York: Oxford University Press.

Sloboda, J. A., O'Neill, S. A., & Ivaldi, A. (2001). Functions of music in everyday listening: An exploratory study using the experience sample method. *Musicae Scientiae, 5*, 9–32.

Smith, L. D., & Williams, R. N. (1999). Children's artistic responses to musical intervals. *American Journal of Psychology, 112*, 383–410.

Sorce, J., Emde, R., Campos, J., & Klinnert, M. (1985). Maternal emotional signaling: Its effect on the visual cliff barrier of one-year olds. *Developmental Psychology, 21*, 195–200.

Spence, C. J., & Driver, J. (1994). Covert spatial orientation in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 555–574.

Spieth, W., Curtis, J. F., & Webster, J. C. (1954). Responding to one of two simultaneous messages. *Journal of the Acoustical Society of America, 26*, 391–396.

Sundberg, J., & Verillo, V. (1980). On the anatomy of the retard: A study of timing in music. *Journal of the Acoustical Society of America, 68*, 772–779.

Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 252–260.

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.

Terwogt, M. M., & van Grisven, F. (1988). Recognition of emotions in music by children and adults. *Perceptual and Motor Skills, 67*, 697–698.

Terwogt, M. M., & van Grisven, F. (1991). Musical expression of mood states. *Psychology of Music, 19*, 99–109.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological Science, 12*, 248–251.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review, 34*, 273–286.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107*, 885–913.

Todd, N. P. M. (1992). The dynamics of dynamics: A model of musical expression. *Journal of the Acoustic Society of America, 91*(6), 3540–3550.

Todd, N. P. M. (1993)Sundberg & Verillo (Eds.). Vestibular feedback in musical performance: Response to Somatosensory Feedback in Musical Performance. *Music Perception, 10*(3), 379–382.

Todd, N. P. M. (1995). The kinematics of musical expression. *Journal of the Acoustic Society of America, 97*(3), 1940–1949.

Todd, N. P. M. (1999). Motion in music: A neurobiological perspective. *Music Perception, 17*(1), 115–126.

Todd, N. (2001). Evidence for a behavioral significance of saccular acoustic sensitivity in humans. *Journal of the Acoustic Society of America, 110*(1), 380–390.

Todd, N. P. M., & Cody, F. W. (1999). Vestibular responses to loud dance music: A physiological basis of the "rock and roll threshold". *Journal of the Acoustic Society of America, 107*(1), 496–500.

Tomkins, S. (1962). *Affect, imagery, and consciousness: The positive affects.* New York: Springer.

Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences, 930*, 1–16.

Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research, 12*, 43–77.

Trevarthen, C. (1993). The self born in intersubjectivity: An infant communicating. In U. Neisser (Ed.), *Ecological and interpersonal knowledge of the self* (pp. 121–173). New York: Cambridge University Press.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae*, 155–211 (Special Issue 1999–2000).

Trevarthen, C., & Aitken, K. J. (1994). Brain development, infant communication, and empathy disorders: Intrinsic factors in child mental health. *Development and Psychopathology, 6*, 599–635.

Wallace, W. T., & Rubin, D. C. (1991). Characteristics and constraints in ballads and their effects on memory. *Discourse Processes, 14*, 181–202.

Wallin, N. L., Merker, B., & Brown, S. (2000). *The origins of music*. Cambridge: MIT Press.

Waterman, M. (1996). Emotional responses to music: Implicit and explicit effects in listeners and performers. *Psychology of Music, 24*, 53–67.

Wedin, L. (1972). Multidimensional study of perceptual-emotional qualities in music. *Scandinavian Journal of Psychology, 13*, 241–257.

Wells, A., & Hakanen, E. A. (1991). The emotional use of popular music by adolescents. *Journalism Quarterly, 68*, 445–454.

Wickelgren, W. A. (1969). Context sensitive coding, associative memory, and serial order in (speech) behavior. *Psychological Review, 76*.

Wittgenstein, L. (1958). *Philosophical investigations*. New York: Macmilan.

Witvliet, C. V., & Vrana, S. R. (1996). The emotional impact of instrumental music on affect ratings, facial EMG, autonomic response, and the startle reflex: Effects of valence and arousal. *Psychophysiology Supplement*, 91.

Witvliet, C. V., Vrana, S. R., & Webb-Talmadge, N. (1998). In the mood: Emotion and facial expressions during and after instrumental music, and during an emotional inhibition task. *Psychophysiology Supplement*, 88.

Wrigley, S.N., & Brown, G.J. (2002). A neural oscillator model of auditory selective attention. In Dietterich, T.G., Becker, S., & Ghahramani, Z. (Eds.), *Advances in neural information processing systems, 14*.

Zentner, M. R., & Kagan, J. (1996). Perception of music by infants. *Nature, 383*, 29.

Zentner, M. R., & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior and Development, 21*, 483–492.

Zillman, D. (1988). Mood management: Using entertainment to full advantage. In L. Donohew, H. E. Sypher, & E. T. Higgens (Eds.), *Communication, social cognition, and affect* (pp. 147–171). Hillsdale, NJ: Earlbaum.