Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognition 100 (2006) 33–72
COGNITION
www.elsevier.com/locate/COGNIT

The capacity for music: What is it, and what's special about it?

Ray Jackendoff [a], Fred Lerdahl [b,*]

[a] *Center for Cognitive Studies, Department of Philosophy, Tufts University, Medford, MA 02155, USA*
[b] *Department of Music, Columbia University, New York, NY 10027, USA*

Available online 27 December 2005

Abstract

We explore the capacity for music in terms of five questions: (1) What cognitive structures are invoked by music? (2) What are the principles that create these structures? (3) How do listeners acquire these principles? (4) What pre-existing resources make such acquisition possible? (5) Which aspects of these resources are specific to music, and which are more general?

We examine these issues by looking at the major components of musical organization: rhythm (an interaction of grouping and meter), tonal organization (the structure of melody and harmony), and affect (the interaction of music with emotion). Each domain reveals a combination of cognitively general phenomena, such as gestalt grouping principles, harmonic roughness, and stream segregation, with phenomena that appear special to music and language, such as metrical organization. These are subtly interwoven with a residue of components that are devoted specifically to music, such as the structure of tonal systems and the contours of melodic tension and relaxation that depend on tonality. In the domain of affect, these components are especially tangled, involving the interaction of such varied factors as general-purpose aesthetic framing, communication of affect by tone of voice, and the musically specific way that tonal pitch contours evoke patterns of posture and gesture.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Rhythm; Meter; Tonality; Melody; Harmony; Affect; Emotion

* Corresponding author.
E-mail addresses: ray.jackendoff@tufts.edu (R. Jackendoff), awl1@columbia.edu (F. Lerdahl).

0010-0277/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2005.11.005

1. What is the capacity for music?

Following the approach of Lerdahl and Jackendoff (1983, hereafter GTTM) and Lerdahl (2001, hereafter TPS), we take the inquiry into the human capacity for music to be shaped by five questions. The first question concerns the character of music perception/cognition: what it means to "hear" a piece of music.

> Q1 (Musical structure): When a listener hears a piece of music in an idiom (or style) with which he/she is familiar, what cognitive structures (or mental representations) does he/she construct in response to the music?

These cognitive structures can be called the listener's *understanding* of the music – the organization that the listener unconsciously constructs in response to the music, beyond hearing it just as a stream of sound. Q1 concentrates on the listener rather than on the performer and/or composer, because the experience of listening is more universal than the other two, and because the acts of performing and composing music require listening as well (including generating and attending to musical imagery in the "mind's ear").

Given that a listener familiar with a musical idiom is capable of understanding novel pieces of music within that idiom, we can characterize the ability to achieve such understanding in terms of a set of principles, or a "musical grammar", which associates strings of auditory events with musical structures. So a second question is:

> Q2 (Musical grammar): For any particular musical idiom MI, what are the unconscious principles by which experienced listeners construct their understanding of pieces of music in MI (i.e., what is the musical grammar of MI)?

Cross-culturally as well as intra-culturally, music takes different forms and idioms. Different listeners are familiar (in differing degrees) with different idioms. Familiarity with a particular idiom is in part a function of exposure to it, and possibly also a function of explicit training. So a third question is:

> Q3 (Acquisition of musical grammar): How does a listener acquire the musical grammar of MI on the basis of whatever sort of exposure it takes to do so?

Q3 in turn leads to the question of what cognitive resources make learning possible:

> Q4 (Innate resources for music acquisition): What pre-existing resources in the human mind/brain make it possible for the acquisition of musical grammar to take place?

These questions are entirely parallel to the familiar questions that underpin the modern inquiry into the language faculty, substituting "music" for "language". The answers might come out differently than they do in language, but the questions

themselves are appropriate ones to ask. In particular, the term "capacity for language" has come to denote the pre-existing resources that the child brings to language acquisition. We propose therefore that the term "capacity for music" be used for the answer to Q4. The musical capacity constitutes the resources in the human mind/brain that make it possible for a human to acquire the ability to understand music in any of the musical idioms of the world, given the appropriate input.

The ability to achieve musical competence is more variable among individuals than the universal ability to achieve linguistic competence. The range in musical learning is perhaps comparable to that of adult learning of foreign languages. Some people are strikingly gifted; some are tone-deaf. Most people lie somewhere on a continuum in between and are able to recognize hundreds of tunes, sing along acceptably with a chorus, and so on. This difference from language does not delegitimize the parallels of Q1–4 to questions about language; it just shows that the musical capacity has somewhat different properties than the language capacity.

Here, we approach the musical capacity in terms parallel to those of linguistic theory – that is, we inquire into the formal properties of music as it is understood by human listeners and performers. As in the case of linguistic theory, such inquiry ideally runs in parallel with experimental research on the real-time processing of music, the acquisition of musical competence (as listener or performer), the localization of musical functions in the brain, and the genetic basis of all of this. At the moment, the domain of formal analysis lends itself best to exploring the full richness and complexity of musical understanding. Our current knowledge of relevant brain function, while growing rapidly, is still limited in its ability to address matters of sequential and hierarchical structure. However, we believe that formal analysis and experimental inquiry should complement and constrain one another.[1] We hope the present survey can serve as a benchmark of musical phenomena in terms of which more brain-based approaches to music cognition can be evaluated.

A further important question that arises in the case of music, as in language, is what aspects of the capacity are specific to that faculty, and what aspects are a matter of more general properties of human cognition. For example, the fact that music for the most part lies within a circumscribed pitch range is a consequence of the frequency sensitivity of the human auditory system and of the pitch range of human voices; it has nothing to do specifically with music (if bats had music, they might sing in the pitch range of their sonar). Similarly, perceiving and understanding music requires such general-purpose capacities as attention, working memory, and long-term memory, which may or may not have specialized incarnations for dealing with music.

It is therefore useful to make a terminological distinction between the *broad musical capacity*, which includes any aspect of the mind/brain involved in the acquisition and processing of music, and the *narrow musical capacity*, which includes just those aspects that are specific to music and play no role in other cognitive activities. This distinction

[1] We recognize that there are major sub-communities within linguistics that do not make such a commitment, particularly in the direction from experiment to formal theory. But we take very seriously the potential bearing of experimental evidence on formal analysis. For the case of language, see Jackendoff (2002), especially chapters 6 and 7.

may well be a matter of degree: certain more general abilities may be specially "tuned" for use in music. In some cases it may be impossible to draw a sharp line between special tuning of a more general capacity and something qualitatively different and specialized. Examples will come up later. The overall question can be posed as follows:

> Q5 (Broad vs. narrow musical capacity): What aspects of the musical capacity are consequences of general cognitive capacities, and what aspects are specific to music?

The need to distinguish the narrow from the broad capacity is if anything more pointed in the case of music than in that of language. Both capacities are unique to humans, so in both cases something in the mind/brain had to change in the course of the differentiation of humans from the other great apes during the past five million years or so – either uniquely human innovations in the broad capacity, or innovations that created the narrow capacity from evolutionary precursors, or both. In the case of language it is not hard to imagine selectional pressures that put a premium on expressive, precise, and rapid communication and therefore favored populations with a richer narrow language capacity. To be sure, what one finds easy to imagine is not always correct, and there is considerable dispute in the literature about the existence and richness of a narrow language capacity and the succession of events behind its evolution (compare Hauser, Chomsky, & Fitch, 2002 with Pinker & Jackendoff, 2005; for example). But whatever one may imagine about language, by comparison we find far less compelling the imaginable pressures that would favor the evolution of a narrow musical capacity (not that the literature lacks hypotheses, e.g., Cross, 2003; Huron, 2003; many papers in Wallin, Merker, & Brown, 2000). All else being equal, it is desirable, because it assumes less, to explain as much of the musical capacity as possible in terms of broader capacities, i.e., to treat the music capacity as an only slightly elaborated "spandrel" in the sense of Gould and Lewontin (1979). The difficulty is: a "spandrel" of what?

However, the issue is not purely the desirability of accounting for the musical capacity in terms of other, more evolutionarily plausible components of cognition. It is an empirical question to determine what aspects of the musical capacity, if any, are special; evolutionary plausibility is only one among the relevant factors to consider. Another factor is the existence of deficits, either genetic or caused by brain damage, that differentially impinge on music (Peretz, 2003, Peretz, this issue). The factor that we will primarily address in this article is the necessity to account for the details of musical organization in the musical idioms of the world, and to account for how these details reflect cognitive organization, i.e., musical structure and musical grammar. Our hope in doing so is to show how a cognitive approach to musical structure can help inform inquiry into the biological basis of music.

Not all inquiries into the possible evolutionary antecedents of music have addressed this concern. For instance, Hauser and McDermott (2003) frame their discussion in terms of questions parallel to our questions Q3–5. However, they do not pose these questions in the context of also asking what the "mature state of musical knowledge" is, i.e., our questions Q1–2. Without a secure and detailed account of how a competent listener comprehends music, it is difficult to evaluate hypotheses

about innateness and evolutionary history, because it is not clear what the endpoint of the evolutionary process is.

Our discussion is divided into three major parts. Section 2 deals with the rhythmic organization of music, which in turn is divided into grouping and metrical organization. Section 3 discusses pitch structure, first in terms of the construction of scales and harmonic relations, and then in terms of the construction of melody. Section 4 offers an overview of perhaps the most salient issue in musical cognition: the connection of music to emotion or affect. This necessarily follows the review of the components of music structure, because structure is necessary to support everything in musical affect beyond its most superficial aspects.

2. The rhythmic organization of music

The first component of musical structure is what we call the *musical surface*: the array of simultaneous and sequential sounds with pitch, timbre, intensity, and duration. The study of the complex processes by which the brain constructs a heard musical surface from auditory input belongs to the fields of acoustics and psychoacoustics. We will mostly assume these processes here.

The musical surface, basically a sequence of notes, is only the first stage of musical cognition. Beyond the musical surface, structure is built out of the confluence of two independent hierarchical dimensions of organization: rhythm and pitch. In turn, rhythmic organization is the product of two independent hierarchical structures, grouping and meter. The relative independence of rhythmic and pitch structures is indicated by the possibility of dissociating them. Some musical idioms, such as drum music and rap, have rhythmic but not pitch organization (i.e., melody and/or harmony). There are also genres such as recitative and various kinds of chant that have pitch organization and grouping but no metrical organization of any consequence. (Peretz (2001) also presents neuropsychological evidence for their independence, based on cases of brain damage).

By saying that grouping, meter, and pitch organization constitute independent structures, we do not mean to imply that they do not influence each other. Rather, what we mean is that each of these components has its own characteristic units and combinatorial principles. The basic unit of grouping is a *group* of one or more adjacent notes in the musical surface; adjacent groups can be combined into larger groups. The basic unit of metrical structure is a *beat*, a point in time usually associated with the onset of a note in the musical surface. Beats are combined into a *metrical grid*, a hierarchical pattern of beats of different relative strengths. As will be seen in Section 3, the basic unit of pitch structure is a note belonging to a *tonal pitch space* characteristic of the musical idiom; the concatenated notes of a melody are combined hierarchically to form a pattern of tension and relaxation called a *reduction*. The understanding of a piece of music involves all of these structures simultaneously, and many of the principles that assign a piece a structure in each domain interact with the structures in the other domains. Thus the outcome is not unlike language, where the structure of a sentence involves simultaneously the independent dimensions of phonology, syntax, and semantics.

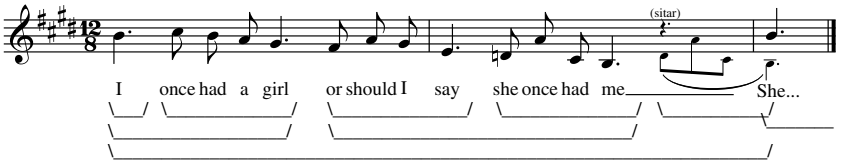

Fig. 1. First phrase of *Norwegian Wood* with its grouping structure.

2.1. Grouping structure

Grouping structure is the segmentation of the musical surface into motives, phrases, and sections. Fig. 1 shows the grouping structure for the melody at the beginning of the Beatles' *Norwegian Wood*; grouping is represented as bracketing beneath the notated music.[2] At the smallest level of the fragment shown, the first note forms a group on its own, and the four subsequent groups are four-note fragments. The last of these, the little sitar interlude, overlaps with the beginning of the next group. At the next level of grouping, the first four groups pair up, leaving the interlude unpaired. Finally, the whole passage forms a group, the first phrase of the song. At still larger levels, this phrase pairs with the next to form the first section of the song, then the various sections of the song group together to form the entire song. Thus grouping is a hierarchical recursive structure.

The principles that create grouping structures (GTTM, chapter 3) are largely general-purpose gestalt perceptual principles which, as pointed out as long ago as Wertheimer (1923), apply to vision as well as audition. (Recent work on musical grouping within the framework of GTTM includes the experimental research of Deliège, 1987 and the computational modeling of Temperley, 2001; for a more general review of experimental research on grouping and meter, see Handel, 1989, especially chapter 11.) In Fig. 1, the cue for small-scale grouping boundaries is mostly relative proximity: when there are longer distances between note onsets, and especially when there are pauses between notes, one perceives a grouping boundary. But other aspects of the signal can induce the perception of grouping boundaries as well. The notes in Fig. 2a are equally spaced temporally, and one hears grouping boundaries at changes of pitch.

It is easy to create musical surfaces in which various cues of grouping boundaries are pitted against one another. Fig. 2b has the same sequence of notes as Fig. 2a, but the pauses cut across in the changes in pitch. The perceived grouping follows the pauses. However, Fig. 2c has the same rhythm as Fig. 2b but more extreme changes of pitch, and here the perceived grouping may follow the changes in pitch rather than the pauses. Thus, as in visual perception, the principles of grouping are defeasible (overrideable) or gradient rather than absolute, and competition among conflicting principles is a normal feature in the determination of musical structure.

[2] All quotes from Beatles songs are based on text in *The Beatles Complete Scores*, Milwaukee, Hal Leonard Publishing Corporation, 1993. We refer to Beatles songs throughout this article because of their wide familiarity.

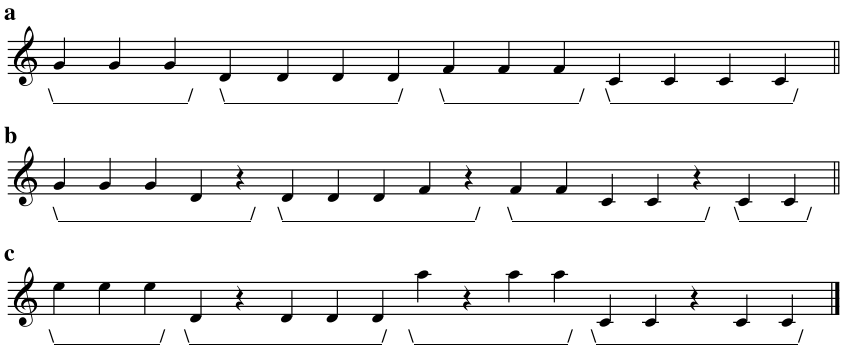

Fig. 2. Application of the gestalt principles of proximity and similarity in the assignment of grouping structure.

(For this reason, GTTM called such principles *preference rules*.) Because the principles have this character, it proves impossible to formulate musical grammar in the fashion of traditional generative grammars, whose architecture is designed to "generate" "all and only" grammatical sentences. Rather, treating rules of musical grammar as defeasible constraints is in line with current constraint-based approaches to linguistic theory such as Head-Driven Phrase Structure Grammar (Pollard & Sag, 1987, 1994) and Optimality Theory (Prince & Smolensky, 1993). Jackendoff (2002, chapter 5) develops an overall architecture for language that is compatible with musical grammar.

Returning to *Norwegian Wood* (Fig. 1), there are two converging cues for its larger grouping boundaries. One is symmetry, in which groups pair up to form approximately equal-length larger groups, which again pair up recursively. The other is thematic parallelism, which favors groups that begin in the same way. In particular, parallelism is what motivates the grouping at the end of Fig. 1: the second phrase should begin the same way the first one does. The cost in this case is the overlapping boundaries between phrases, a situation disfavored by the rule of proximity. However, since the group that ends at the overlap is played by the sitar and the group that begins there is sung, the overlap is not hard to resolve perceptually. Grouping overlap is parallel to the situation in visual perception where a line serves simultaneously as the boundary of two abutting shapes.

2.2. Metrical structure

The second component of rhythmic organization is the *metrical grid*, an ongoing hierarchical temporal framework of *beats* aligned with the musical surface. Fig. 3 shows the metrical grid associated with the chorus of *Yellow Submarine*. Each vertical column of x's represents a beat; the height of the column indicates the relative strength of the beat. Reading horizontally, each row of x's represents a temporal regularity at a different time-scale. The bottom row encodes local regularities, and the higher rows encode successively larger-scale regularities among sequences of successively stronger beats. Typically, the lowest row of beats is isochronous (at least cognitively – chrono-

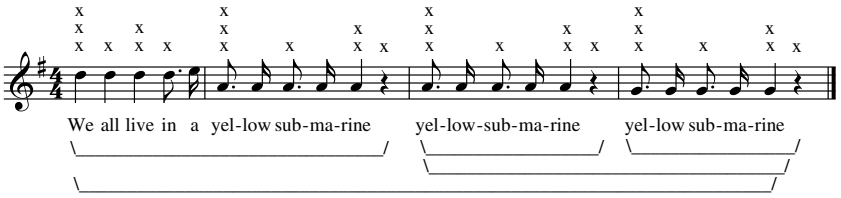

Fig. 3. The first phrase of *Yellow Submarine* with its metrical and grouping structures.

metrically it may be slightly variable), and higher rows are uniform multiples (double or triple) of the row immediately below. For instance, in Fig. 3, the lowest row of beats corresponds to the quarter-note regularity in the musical surface, the next row corresponds to two quarter notes, and the top row corresponds to a full measure. Bar lines in musical notation normally precede the strongest beat in the measure.

A beat is conceived of as a point in time (by contrast with groups, which have duration). Typically, beats are associated temporally with the attack (or onset) of a note, or with a point in time where one claps one's hands or taps one's foot. But this is not invariably the case. For instance, in Fig. 3, the fourth beat of the second measure is not associated with the beginning of a note. (In the recorded performance, the guitars and drums do play on this beat, but they are not necessary for the perception of the metrical structure.) Moreover, the association of an attack with a beat is not rigid, in that interpretive flexibility can accelerate or delay attacks without disrupting perceived metrical structure. Careful attention to the recorded performance of *Norwegian Wood* reveals many such details. For instance, the sitar begins its little interlude not exactly on the beat, but a tiny bit before it. More generally, such anticipations and delays are characteristic of jazz and rock performance (Ashley, 2002; Temperley, 2001) and expressive classical performance (Palmer, 1996; Repp, 1998, 2000; Sloboda & Lehmann, 2001; Windsor & Clarke, 1997). We return to their role in music in Section 4.3.

In Fig. 3, the beginnings of groups line up with strong beats. It is also common, however, for a group to be misaligned with the metrical grid, in which case the phrase begins with an upbeat (or anacrusis). Fig. 4, a phrase from *Taxman*, shows a situation in which the first two-bar group begins three eighth notes before the strong beat and the second two-bar group begins a full four beats before its strongest beat. The second group also illustrates a rather radical misalignment of note onsets with the metrical grid as a whole.

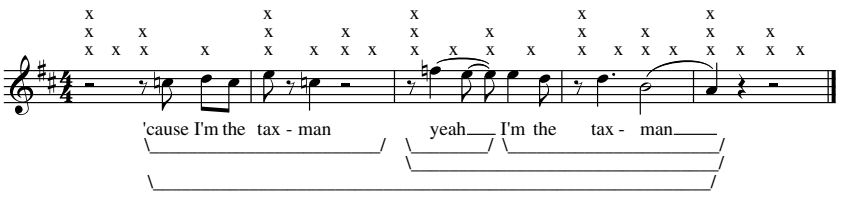

Fig. 4. Metrical and grouping structures of a phrase in *Taxman*.

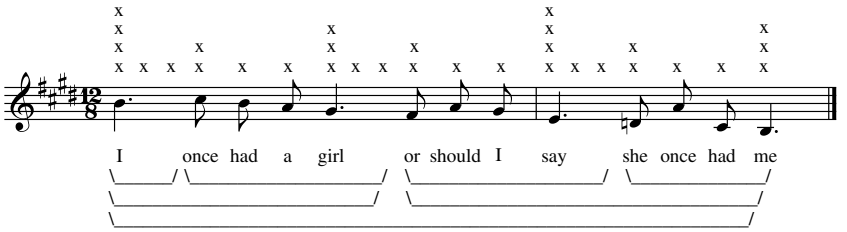

Fig. 5. Grouping and metrical structures for the first phrase of *Norwegian Wood*.

More subtle possibilities also exist. Fig. 5 shows the grouping and metrical structures of *Norwegian Wood*. Here, all the groups begin on beats that are strong at the lowest layer of the grid; but the second, third, and fourth groups begin on beats that are weak at the second layer of the grid (i.e., do not correspond to beats at the third level of the grid); their strongest beat is in fact on the final note of the group. This example shows that the notion of "upbeat" has to be construed relative to a particular layer in the metrical hierarchy.

In Western classical and popular music, metrical grids are typically regular, each level uniformly doubling or tripling the one below it. But there are occasional anomalies. Fig. 6a, from *Here Comes the Sun*, shows an irregularity at a small-scale metrical level, in which three triple-length beats are inserted into a predominantly duple meter; the listener feels this irregularity as a strong jolt. Fig. 6b, from *All You Need is Love*, shows a larger-scale irregularity, where the phrases have a periodicity of seven beats. Because of its greater time scale, not all listeners will notice the anomaly, but for trained musicians it pops out prominently.

Across the musical idioms of the world, regular metrical structures like those in Figs. 3 and 5 are very common. In addition, there are genres that characteristically make use of irregular periodicities of two and three at a small metrical level ($2+2+3$, $2+3+2+2+3$, etc.), for instance Balkan folk music (Nettl, 1973; Singer, 1974). The metrical complexity of polyrhythmic African music has been subject to mixed interpretations. On one interpretation, the polyrhythms project multiple metrical grids in counterpoint, arranged so that they align only at the smallest level and at some relatively large level, but proceeding with apparent independence

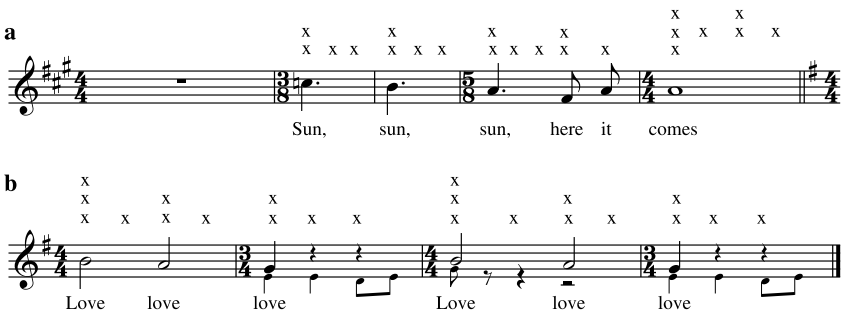

Fig. 6. Metrical irregularities in (a) *Here Comes the Sun* and (b) *All You Need is Love*.

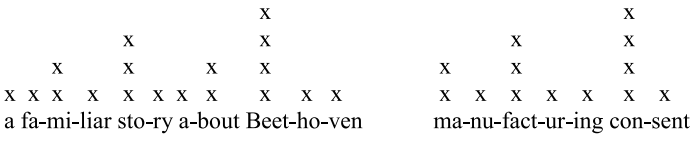

Fig. 7. Grids for linguistic stress.

at levels in between. Alternatively, the polyrhythms are understood as syncopations against a regular metrical framework (Locke, 1982; Temperley, 2001). Western listeners may experience music in these traditions as exciting, but their mental representations of its metrical structure are less highly developed than are those for whom this music is indigenous – to the extent that they may have difficulty reliably tapping their feet in time to it (Trehub and Hannon, this issue). Finally, there are numerous genres of chant and recitative throughout the world in which a temporally rigid metrical grid is avoided and the music more closely follows speech rhythms.

The principles that associate a metrical grid with a musical surface (GTTM, chapter 4; Temperley, 2001), like those for establishing grouping, are defeasible principles whose interaction in cooperation and competition has to be optimized. Prominent cues for metrical strength include (a) onsets of notes, especially of long notes, (b) intensity of attack, and (c) the presence of grouping boundaries. The first of these principles is overridden in syncopation, such as the second half of Fig. 4, when note onsets surround a relatively strong beat. The second principle is overridden, for instance, when a drummer gives a ''kick'' to the offbeats. The third principle is overridden when groups begin with an upbeat. Generally, beats are projected in such a way as to preserve a maximally stable and regular metrical grid. But even this presumption is overridden when the musical surface provides sufficient destabilizing cues, as in Figs. 6a and b. In other words, the construction of a metrical grid is the result of a best-fit interaction between stimulus cues and internalized regular patterns.

At this point the question of what makes music special begins to get interesting. Musical metrical grids are formally homologous to the grids used to encode relative stress in language, as in Fig. 7 (Liberman & Prince, 1977). Here a beat is aligned with the onset of the vowel in each syllable, and a larger number of x's above a syllable indicates a higher degree of stress. So we can ask if the formal homology indicates a cognitive homology as well.

Two immediate differences present themselves. In normal spoken language, stress grids[3] are not regular as are metrical grids in music (compare Fig. 7 to Figs. 3 and 5), and they are not performed with the degree of isochrony that musical grids are. Yet there are striking similarities. First, just as movements such as clapping or foot-tapping are typically timed so as to line up with musical beats, hand gestures accompanying speech are typically timed so as to line up with strong stresses (McNeill, 1992). Second, among the most important cross-linguistic cues for stress is the heaviness of

[3] We use the term ''stress grid'' here, recognizing that the linguistic literature often uses the term ''metrical grid'' for relative degrees of stress. We think it important to distinguish in both music and poetry between often irregular patterns of stress (the ''phenomenal accents'' of GTTM, chapter 2) and the regular metrical patterns against which stresses are heard.

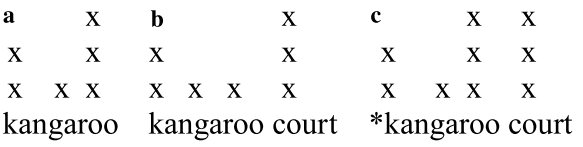

Fig. 8. Stress shift in *kangaroo*.

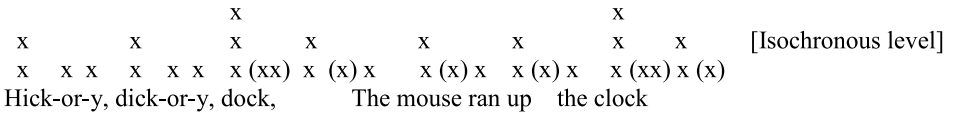

Fig. 9. An example of a metrical grid in poetry.

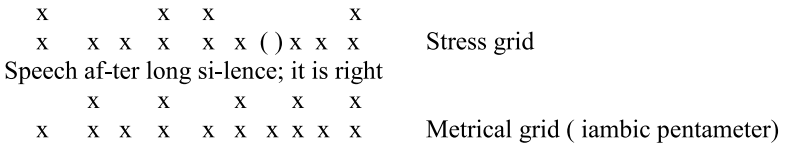

Fig. 10. A line from Yeats, *After Long Silence*, with accompanying stress and metrical grids (adapted from Halle and Keyser (1971)).

a syllable, where (depending on the language) a syllable counts as heavy if it has a long vowel and/or if it closes with a consonant (Spencer, 1996). This corresponds to the preference to hear stronger musical beats on longer and louder notes. Third, there is a cross-linguistic preference for alternating stress, so that in some contexts the normal stress of a word can be distorted to produce a more regular stress pattern, closer to a musical metrical grid. For instance, the normal main stress of *kangaroo* is on the final syllable, as in Fig. 8a; but in the context *kangaroo court* the main stress shifts to the first syllable to make the stress closer to an alternating pattern (Fig. 8b instead of Fig. 8c).

In poetry the parallels become more extensive. A poetic meter can be viewed as a metrical grid to which the stress grid in the text is optimally aligned (Halle & Keyser, 1971). Especially in vernacular genres of poetry such as nursery rhymes and limericks, the metrical grid is performed quasi-isochronously, as in music – even to the point of having rests (silent beats) in the grid (Burling, 1966; Oehrle, 1989).[4] Fig. 9 illustrates this point; note that its first phrase begins on a downbeat and the second on an upbeat. In sophisticated poetry, it is possible, within constraints, to misalign the stress grid with the metrical grid (the poetic meter); this is a counterpart of syncopation in music. Fig. 10 illustrates one instance.

[4] In this connection, Lerdahl (2003) applies the analytic procedures of GTTM to the sounds of a short poem by Robert Frost, "Nothing Gold Can Stay". Although the syllable count indicates that the poem is in iambic trimeter, the analysis treats the poem as in iambic tetrameter, with a silent beat at the end of each line, an interpretation motivated by the semicolon or period at the end of each line. It has recently come to our attention that Frost's own reading of the poem, recorded in Paschen and Mosby (2001), follows this silent-beat interpretation.

Given these extensive similarities, it is reasonable to suppose that the two systems draw on a common underlying cognitive capacity. But it is then necessary to account for the differences. Here is one possibility. The principles of metrical grids favor hierarchical regularity of timing. However, stress in language is constrained by the fact that it is attached to strings of words, where the choice of words is in turn primarily constrained by the fact that the speaker is trying to convey a thought. Therefore regularity of stress usually has to take a back seat. Stress-timed poetry adds the constraint that the metrical grid must be regular, and in that sense more ideal. By contrast, in much music, the ideal regular form of the metrical grid is a primary consideration. Because metrical regularity is taken for granted, stresses on the musical surface can be played off against it to a greater extent than is the case in poetry. In short, the same basic cognitive system is put to use in slightly different ways because of the independent constraints imposed by other linguistic or musical features with which it interacts.

On this view, metrical structure is part of the broad musical capacity. It remains to ask *how* broad. We see little evidence that metrical grids play a role in other human (or animal) activities besides music and language. To be sure, other activities such as walking and breathing involve temporal periodicities. But periodicity alone does not require metrical grids: metrical grids require a differentiation between strong and weak beats, projected hierarchically. For example, walking involves an alternation of legs, but there is no reason to call a step with one leg the strong beat and a step with the other the weak beat. And these activities certainly present no evidence for metrical grids extended beyond two levels, that is, with the complexity that is routine in language and especially music. A promising candidate for metrical parallels with music is dance, where movement is coordinated with musical meter. We know of no other activities by humans or other animals that display symptoms of metrical grids, though perhaps observation and analysis will yield one. We tentatively conclude that metrical structure, though part of the broad musical capacity, is not widely shared with other cognitive systems. It thus presents a sharp contrast with grouping structure, which is extremely broad in its application, extending even to static visual grouping and to conceptual groupings of various kinds.

There is little evidence of non-human behavior that requires a metrical grid. Bonobos may engage in synchronously pulsed chorusing, which requires periodicity (a "pulse"), but not a hierarchical metrical grid with alternating strong and weak beats (see also examples and discussion in Trehub & Hannon, this volume). By contrast, human children spontaneously display movements hierarchically timed with music, often by the age of two or three (and we personally have observed even younger children attempting to synchronize their movements with music). Phillips-Silver and Trainor (2005) demonstrate that even 7-month-old infants are sensitive to the difference between 2-beat and 3-beat metrical regularity (though the counterpart with non-humans remains to be investigated). Thus there seems to be something special going on in humans even at this seemingly elementary level of rhythm.

3. Pitch structure

We begin our discussion of pitch structure by observing that harmony in Western music is not representative of indigenous musical idioms of the world. In other idioms (at least before the widespread influence of Western music) it is inappropriate to characterize the music in terms of melody supported by accompanying chords. Our contemporary sense of Western tonal harmony started in the European Middle Ages and coalesced into approximately its modern form in the 18th century. Developments of tonal harmony in the 19th and 20th centuries are extensions of the existing system rather than emergence of a new system.

3.1. Tonality and pitch space

What is representative of the world's musical idioms is not harmony, but rather a broad sense of tonality that does not require or even imply harmonic progression and that need not be based on the familiar Western major and minor scales (Nettl, 1973). Western harmony is a particular cultural elaboration of this basic sense of tonality. In a tonal system in this sense (from now on we will just speak of a tonal system), every note of the music is heard in relation to a particular fixed pitch, the *tonic* or *tonal center*. The tonic may be sounded continuously throughout a piece, for instance by a bagpipe drone or the tamboura drone in Indian raga; or the tonic may be implicit. Whether explicit or implicit, the tonic is felt as the focus of pitch stability in the piece, and melodies typically end on it. Sometimes, as in modulation in Western music, the tonic may change in the course of a piece, and a local tonic may be heard in relation to an overall tonic. The presence of a tonal center eases processing (Deutsch, 1999) and is a musical manifestation of the general psychological principle of a cognitive reference point within a category (Rosch, 1975). (A prominent exception to tonicity is Western non-tonal music since the early 20th century, an art music that is designed in part to thwart the listener's sense of tonal center, and we set it aside here).

A second essential element of a tonal system is a *pitch space* arrayed in relation to the tonic. At its simplest, the pitch space associated with a tonic is merely a set of pitches, each in a specified interval (a specified frequency ratio) away from the tonic. Musicians conventionally present the elements of such a space in ascending or descending order as a *musical scale*, such as the familiar major and minor diatonic scales. Of course, actual melodies present the elements of a scale in indefinitely many different orders.[5]

[5] It is possible for a scale to include different pitches depending on whether the melody is ascending or descending. A well-known example is the Western melodic minor mode, which has raised sixth and seventh degrees ascending and lowered sixth and seventh degrees descending. It is also possible, if unusual, for the pitch collection to span more than an octave, and for the upper octave to contain different intervals than the lower. Examples appear in Nettl (1960, p. 10) and Binder (1959, p. 85); the latter is a scale commonly used in American synagogues for torah chant on Rosh Hashana and Yom Kippur.

A pitch space usually has more structure beyond the distinction between the tonic and everything else in the scale. The intuition behind this further organization is that distances among pitches are measured not only psychophysically but also cognitively. For example, in the key of C major, the pitch D♭ is closer to C in vibrations per second than D is, but D is cognitively closer (i.e., closer in terms of tonality) because it is part of the C major scale and D♭ is not. Similarly, in C major, G above the tonic C is cognitively closer to C than is F, which is psychophysically closer, because G forms a consonant fifth in relation to C while (in the conventions of classical harmony) F forms a relatively dissonant fourth in relation to C. On both empirical and theoretical grounds (Krumhansl, 1990, TPS, chapter 2), cognitive pitch-space distances are hierarchically organized. Further, pitch-space distances can be mapped spatially, through multi-dimensional scaling and theoretical modeling, into regular three- and four-dimensional geometrical structures. There is even provisional evidence that these structures have brain correlates (Janata et al., 2003). It is beyond our scope here to pursue the geometrical representations of pitch space.

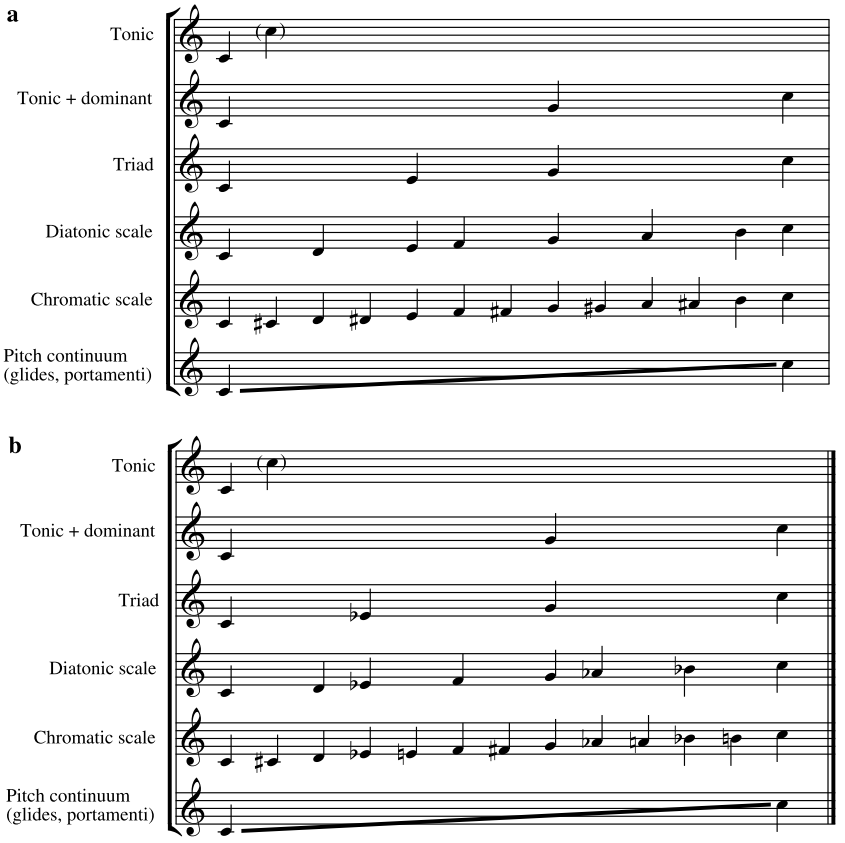

Fig. 11. Pitch space for (a) C major and (b) C minor.

Fig. 11 gives the standard space for the major and minor modes in common-practice Western music. As with strong beats in a metrical grid, a pitch that is relatively stable at a given level also appears at the next larger level. The topmost layer of the taxonomy is the tonic pitch. The next layer consists of the tonic plus the dominant, a fifth higher than the tonic. The dominant is the next most important pitch in the pitch collection, one on which intermediate phrases often end and on which, in Western harmony, the most important chord aside from the tonic chord is built. The third layer adds the third scale degree, forming a triad, the referential sonority of harmonic tonality. The fourth layer includes the remaining notes of the diatonic scale. The fifth layer consists of the chromatic scale, in which adjacent pitches are all a half step apart (the smallest interval in common-practice tonality). Tonal melodies often employ chromatic pitches as alterations within an essentially diatonic framework. The bottom layer consists of the entire pitch continuum out of which glissandi and microtonal inflections arise. Microtones are not usually notated in Western music, but singers and players of instruments that permit them (e.g., in jazz, everything but the piano and drum set) frequently use glides and "bent" notes before or between notes for expressive inflection.

The taxonomy of a pitch space provides a ramified sense of orientation in melodies: a pitch is heard not just in relation to the tonic but also in relation to the more stable pitches that it falls between in the space. For instance, in Figs. 11a–b, the pitch F is heard not just as a fourth above the tonic, but also as a step below the dominant, G, and a step above the third, E or Eb. In Fig. 11a, chromatic D# (or E♭), a non-scale tone, is heard in relation to D and E, the relatively stable pitches adjacent to it. A pitch "in the cracks" between D# and E will be heard as out of tune, but the same pitch may well be passed through by a singer or violinist who is gliding or "scooping" up to an E, with no sense of anomaly.[6]

How much of the organization of pitch space is special to music? This question can be pursued along three lines: in relation to psychoacoustics, to abstract cognitive features, and to the linguistic use of pitch in intonation and tone languages.

3.2. Tonality and psychoacoustics

People often sing in octaves without even noticing it; two simultaneous pitches separated by an octave (frequency ratio 2:1) are perceptually smooth. By contrast, two simultaneous pitches separated by a whole step (ratio 9:8 in "just" intonation), a half step (ratio 16:15), or a minor seventh (16:9) are hard to sing and are perceived as rough. Other vertical intervals such as fifths (3:2), fourths (4:3), major thirds (5:4), and major sixths (5:3) lie between octaves and seconds in sensory dissonance. In general, vertical intervals with small integer frequency ratios (allowing for small, within-

[6] It is an interesting question whether the "blue note" in jazz, somewhere between the major and minor third degree (i.e., between E♭ and E in C major), is to be regarded as an actual scale pitch, as Sargent (1964) analyzes it, or as a conventionalized out-of-scale pitch. More than other pitches of the scale, the blue note is unstable: performers characteristically "play with the pitch".

category deviations) are perceived as more consonant than those with large-integer frequency ratios.

Beginning with the Pythagoreans in ancient Greece, theorists have often explained consonant intervals on the basis of these small-integer ratios (originally in the form of string lengths). From Rameau (1737) onward, attempts have been made to ground consonance and dissonance not only in mathematical ratios but also in the physical world through the natural overtone series. Broadly speaking, modern psychoacoustics takes a two-component approach. First, the physiological basis of Helmholtz's (1885) beating theory of dissonance has been refined (Plomp & Levelt, 1965). If two spectral pitches (i.e., fundamentals and their overtones) fall within a proximate region (a critical band) on the basilar membrane, there is interference in transmission of the auditory signal to the auditory cortex, causing a sensation of roughness. Second, at a more cognitive level, the auditory system attempts to match spectral pitches to the template of the harmonic series, which infants inevitably learn through passive exposure to the human voice even before birth (Lecanuet, 1996; Terhardt, 1974). Vertical intervals that fit into the harmonic template are heard in relation to their "virtual" fundamentals, which are the psychoacoustic basis for the music-theoretic notion of harmonic root. A chord is dissonant to the extent that it does not match a harmonic template, yielding multiple or ambiguous virtual fundamentals.

The two pitch spaces in Fig. 11 reflect psychoacoustic (or sensory) consonance and dissonance in their overall structures. The most consonant intervals appear in the rows of the top layers, and increasingly dissonant intervals appear in successive layers. Thus the octave is in the top layer, the fifth and fourth in the second layer, thirds in the third layer, seconds (whole steps and two half steps) in the fourth layer, and entirely half steps in the fifth layer.

The pitch space for a particular musical idiom, however, may reflect not only sensory dissonance, which is unchanging except on an evolutionary scale, but also musical dissonance, which is a cultural product dependent only in part on sensory input. To take two cases that have caused difficulties for theorists (such as Hindemith, 1952 & Bernstein, 1976) who attempt to derive all of tonal structure from the overtone series: (1) in the second layer of Figs. 11a–b, the fourth (G to upper C) appears as equal to the fifth (lower C–G), whereas in standard tonal practice the fourth is treated as the more dissonant; (2) in the third layer, the major triad in Fig. 11a (C–E–G) and the minor triad in Fig. 11b (C–E♭–G) are syntactically equivalent structures, even though the minor triad is not easily derivable from the overtone series and is more dissonant than the major triad. But these are small adjustments on the part of culture. It would be rare, to take the opposite extreme, for a culture to build stable harmonies out of three pitches a half step apart. The conflict between intended stability and sensory dissonance would be too great to be viable. Cultures generally take advantage of at least broad distinctions in sensory consonance and dissonance.

Traditional Western tonality has sought a greater convergence between sensory and musical factors than have many cultures. Balinese gamelan music, for instance, is played largely on metallic instruments that produce inharmonic spectra (i.e., overtones that are not integer multiples of the fundamental). Consequently, Balinese cul-

ture does not pursue a high degree of consonance but tolerates comparatively wide deviations in intervallic tuning. Instead, a value is placed on a shimmering timbre between simultaneous sounds, created by an optimal amount of beating (Rahn, 1996). The contrasting examples of Western harmonic tonality and Balinese gamelan illustrate how the underlying psychoacoustics influences but does not dictate a particular musical syntax.

Psychoacoustic factors affect not only vertical but also horizontal features of music. Huron (2001) demonstrates this for the conventional rules of Western counterpoint. For example, parallel octaves and fifths are avoided because parallel motion between such consonant intervals tends to fuse two voices into one, contradicting the ideal in Western counterpoint of independent polyphonic voices. (Melodies sung in parallel fifths have a "medieval" sound to modern Western ears.) Parallel thirds and sixths, common in harmonization of modern Western melodies, are acceptable because these intervals are sufficiently dissonant to discourage fusion yet not so dissonant as to cause roughness. Cultures that do not seek a polyphonic ideal, however, have no need to incorporate such syntactic features into their musical idioms.

Intervallic roughness/dissonance pertains only to *simultaneous* presentation of pitches and says nothing about sequential presentation in a melody. Given the rarity of harmonic systems in non-Western and pre-modern tonal traditions, *sequential* presentation is at least as pertinent to the issue of the psychological "naturalness" of tonality. Small intervals such as whole and half steps are harmonically rough. Yet in the context of a melody, such intervals are most common, most stable, least distinctive, and least effortful. By contrast, although the interval of an octave is maximally smooth harmonically, octaves are relatively rare and highly distinctive as part of a melody (for instance, in the striking opening leap of *Somewhere Over the Rainbow*).

The naturalness of small melodic intervals follows in part from two general principles, both of which favor relatively small frequency *differences* rather than small-integer frequency *ratios*. First, in singing or other vocalization, a small change in pitch is physically easier to accomplish than a large one. Second, melodic perception is subject to the gestalt principle of good continuation. A melody moving discretely from one pitch to another is perceptually parallel to visual apparent motion; a larger interval corresponds to a greater distance of apparent motion (Gjerdingen, 1994). A pitch that is a large interval away from the melody's surrounding context is perceptually segregated, especially if it can be connected to other isolated pitches in the same range (Bregman, 1990). For instance, in Fig. 12a the three extreme low notes pop out of the melody and are perceived as forming a second independent line, shown in Fig. 12b.

The factors behind a preference for small melodic intervals are not unique to music. Stream segregation occurs with non-musical auditory stimuli as well as musical stimuli. As in the visual field, auditory perception focuses on, or attends to, psychophysically proximate pitches (Scharf, Quigley, Aoki, Peachery, & Reeves, 1987). Likewise, in spoken language, large frequency differences function more distinctively than small ones.

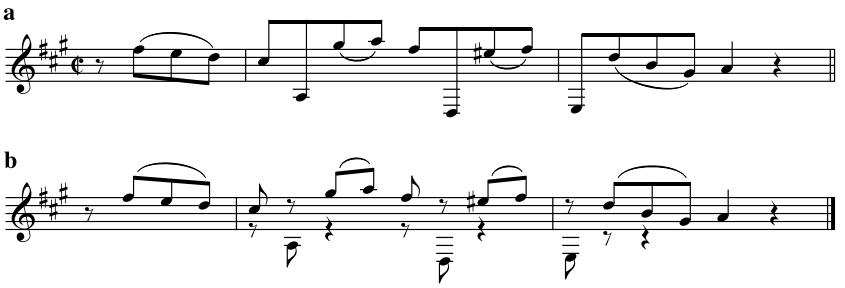

Fig. 12. Except of a line from Mozart's Clarinet Quintet, fourth movement, (a) as written and (b) as heard.

3.3. Cognitive features of tonality

The structure of pitch spaces has further cognitive significance. First, the elements of a pitch space are typically spaced asymmetrically yet almost evenly (Balzano, 1980; Clough, Engebretsen, & Kochavi, 1999). For instance, in Fig. 11 the dominant pitch (G) divides the octave not in half but almost in half (in terms of ratios, which are the relevant criterion for measuring intervals): it is a fifth above the lower tonic and a fourth below the upper one. The diatonic major mode in Fig. 13a distributes half steps unevenly between two and three whole steps. Similarly, the pentatonic scale in Fig. 13b has an asymmetrical combination of whole steps and minor thirds. A common mode in Jewish liturgical music and klezmer music, called "Ahava raba" or "Fregish", has the configuration in Fig. 13c, using half steps, whole steps, and an augmented second. By contrast, scale systems built out of equal divisions of the octave, such as the six-pitch whole-tone scale, are rare in "natural" musical idioms. Asymmetrically distributed intervals help listeners orient themselves in pitch space (Browne, 1981), just as they would in physical space. (Imagine trying to orient yourself inside an equal-sided hexagonal room with no other distinguishing features; the view would be the same from every corner. But if the room had unequal sides dis-

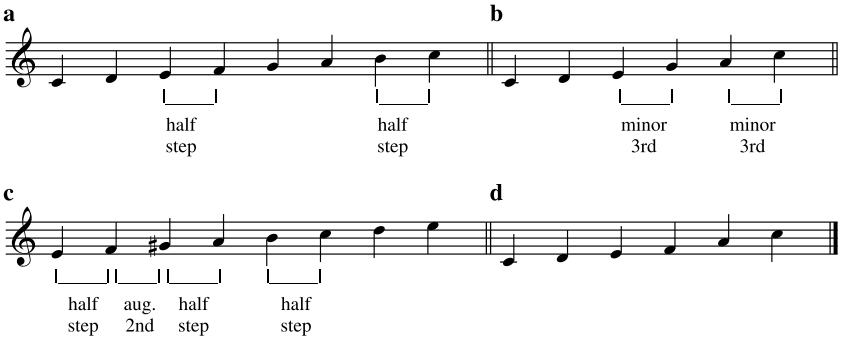

Fig. 13. Examples of scale spaces: (a) diatonic scale with C as tonic; (b) pentatonic scale with C as tonic; (c) Fregish scale with E as tonic; and (d) scale that is non-preferred because of its unevenness.

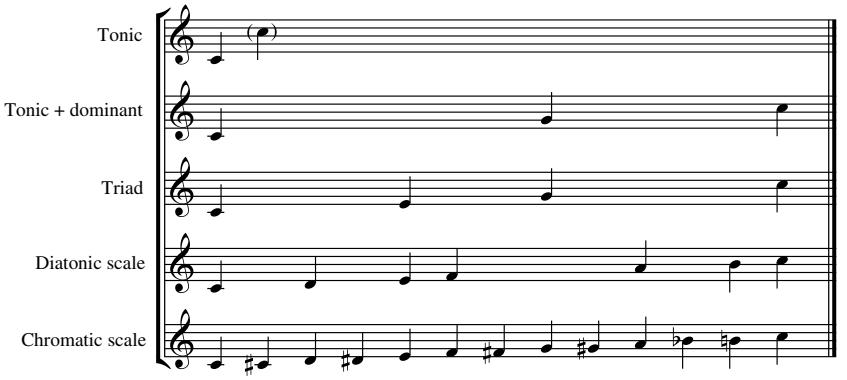

Fig. 14. An example of an ill-formed space.

tributed unevenly, each corner would have a distinctive vista.) However, asymmetry without approximate evenness is undesirable: Fig. 13d is a non-preferred space because its scale is quite uneven, leaving steps that feel like skips between F–A and A–C. A highly preferred tonal space, such as those in Figs. 11 and 13a–b, distributes its pitches at each layer asymmetrically but as evenly as possible given the asymmetry.

A second cognitive feature of pitch spaces lies in the strict taxonomy of Fig. 11: each layer consists of pitches selected from the layer below it. Thus scales are built from the repertory of pitches, chords are built from scale members, and tonics come from either scales or chords, depending on whether the idiom in question uses chords. Thus a pitch space like Fig. 14 is ill-formed because there is a G in the tonic triad that is not also a member of the diatonic scale.[7]

Two more cognitive features of tonal pitch spaces play a role in the organization of melody and will be taken up in somewhat more detail in Section 3.5. The first is that the pitch space facilitates intuitions of tonal tension and relaxation. The tonic pitch is home base, the point of maximal relaxation. Motion away from the tonic – whether melodically, harmonically, or by modulation to another key – raises tension, and motion toward the tonic induces relaxation (TPS, chapter 4). Because music is processed hierarchically, degrees of tension and relaxation take place at multiple levels of musical structure, engendering finely calibrated patterns. Second, pitch space fosters intuitions of tonal attraction (TPS, chapter 4; Larson, 2004). An unstable pitch tends to anchor on a proximate, more stable, and immediately subsequent pitch (Bharucha, 1984, 1996). Tonal attractions in turn generate expec-

[7] Tonal spaces resemble metrical grids in their abstract structures (compare the grid in Fig. 3), except that in Western music the time intervals between beats in metrical grids are typically equal, unlike the case with pitch intervals. In West African drumming music, however, there are standard rhythmic patterns that correspond to the asymmetrical structure of the diatonic and pentatonic scales (Pressing, 1983; Rahn, 1983). Thus in this idiom, which lacks tonal structure, the rhythmic domain possesses a counterpart of some of the structural richness of other idioms' tonal systems.

tations. The listener expects a pitch or chord to move to its greatest attractor. If it does so move, the expectation is fulfilled; if it does not, the expectation is denied (also see Meyer, 1956; Narmour, 1990). In general, tension and attraction are inversely related: motion toward a stable pitch reduces tension while it increases the expectation that the stable pitch will arrive.

To sum up so far, psychoacoustics provides a defeasible and culturally non-binding foundation for aspects of tonality and pitch structure. In particular, small-integer frequency ratios produce relative sensory consonance; but smaller frequency differences (which are harmonically rough) provide a better basis for melodic continuity. At the same time, some abstract cognitive features of pitch space relate to features that exist elsewhere in cognition. The typical asymmetry of pitch spaces makes it possible for the listener more easily to orient with respect to the tonic; and the hierarchical organization of pitch space creates the possibility of intuitions of tonal attraction, tension and relaxation.

Yet the pitch organization of almost any musical idiom achieves a specificity and complexity far beyond these general influences. In particular, psychoacoustic considerations alone do not explain why music is invariably organized in terms of a set of fixed pitches organized into a tonal pitch space. Moreover, although general gestalt principles of proximity and good continuation lie behind a preference for small melodic intervals, they do not explain why the particular intervals of the whole step and half step are so prevalent in melodic organization across the musical idioms of the world. We conclude that the mind/brain must contain something more specialized than psychoacoustic principles that accounts for the existence and organization of tonality.

3.4. Pitch structure and language

Could this additional bit of specialization be a consequence of something independently necessary for language (as we found in the case of metrical structure)? Two linguistic features are reminiscent of musical pitch. First, prosodic contours (sentences and breath-groups within sentences) typically move downward in pitch toward the end, with exceptions such as the upward intonation of yes–no questions in English. Such contours parallel the typical shape of melodies, which also tend to move downward toward the ends of phrases, as seen for instance in Figs. 3–6. (Huron, 1996 provides statistical support for this tendency in a large database of European folksongs.) In fact, non-linguistic cries also exhibit such a downward intonation – and not in humans alone (e.g., Hauser & Fowler, 1991). From this we might conclude that some aspects of melodic shape follow from extra-musical considerations.

However, prosodic contours, even when they pass through a large pitch interval, are not composed of a sequence of discrete pitches the way melodies are. Rather, the pitch of the voice typically passes continuously between high and low points. Current accounts of intonation (Beckman & Pierrehumbert, 1986; Ladd, 1996) analyze prosodic contours in terms of transitions between distinctive high and low tones, so it might be possible to treat intonation as governed by a pitch space whose layers are (a) the high and low tones (with the low tone perhaps as tonic) and (b) the pitch

continuum between them. But even so, the high and low tones are not fixed in frequency throughout a sequence of sentences in the way that the dominant and tonic are fixed in pitch space.

Another linguistic feature possibly analogous to tonal space is the use of pitch in tone languages such as Chinese and many West African languages (Yip, 2002). Tone languages have a repertoire of tones (high, low, sometimes mid-tone, and sometimes rising and falling tones of various sorts) that form an essential part of the pronunciation of words. The tones form a fixed set that can be seen as playing a role parallel to a pitch space or scale in tonal music. But the analogy is not exact. The tones are typically overlaid with an intonation contour, such that the entire range from high to low tone drifts downward in the course of a phrase. Moreover, in the course of down-drift the frequency ratio between high and low tones also becomes smaller (Ladd, 1996; Robert Ladd, personal communication). In music, the improbable parallel would be a melody in which not only the pitches sagged gradually in the course of a phrase, as if an analog recording were slowing down, but the intervals also got smaller, octaves gradually degrading to fifths, fifths to thirds, and so on. Thus neither the pitches of tone languages nor the intervals between the pitches are fixed, as they are in musical spaces.

These comparisons to language amplify the conclusion reached at the end of the previous subsection. Although some features of musical pitch are consequences of more general cognitive capacities, a crucial aspect is *sui generis* to music: the existence of a fixed pitch set for each musical mode, where each pitch is heard in relation to the tonic and in relation to adjacent pitches at multiple layers of pitch space.

Some of these characteristics are provisionally confirmed by neuropsychological evidence. There appear to be two distinct brain systems concerned with pitch, the one involving recognition of pitch contours and the other involving recognition of fixed pitches and intervals. Impairment in the former results in intonational deficits in both music and language; impairment in the latter affects music but spares language (Peretz, 2001). This evidence suggests that there is something special about detecting fixed pitches and intervals. Brain correlates of the more complex aspects of tonality are yet to be discovered (but see Janata et al., 2003).

3.5. Hierarchical structure in melody

So far we have spoken only of the collection of pitches and intervals out of which melodies are constructed. We now turn to some of the structural principles governing the sequential ordering of pitches into melodies. We will avoid issues of harmonic progression and modulation as too complicated for present purposes; they are in any event not germane to most musical idioms of the world.

The first phrase of *Norwegian Wood*, with its unchanging tonic harmony, again serves as a useful example. The understanding of this melody goes beyond just hearing the sequence of notes. In particular, the melody is anchored by the long notes (“I. . .girl. . .say. . .me”), which spell out notes of the E major triad, B–G#–E–B. These anchors are relatively stable points, as they belong to the tonic–chord layer of the pitch space for E major, which is shown in Fig. 15a (with a flatted seventh,

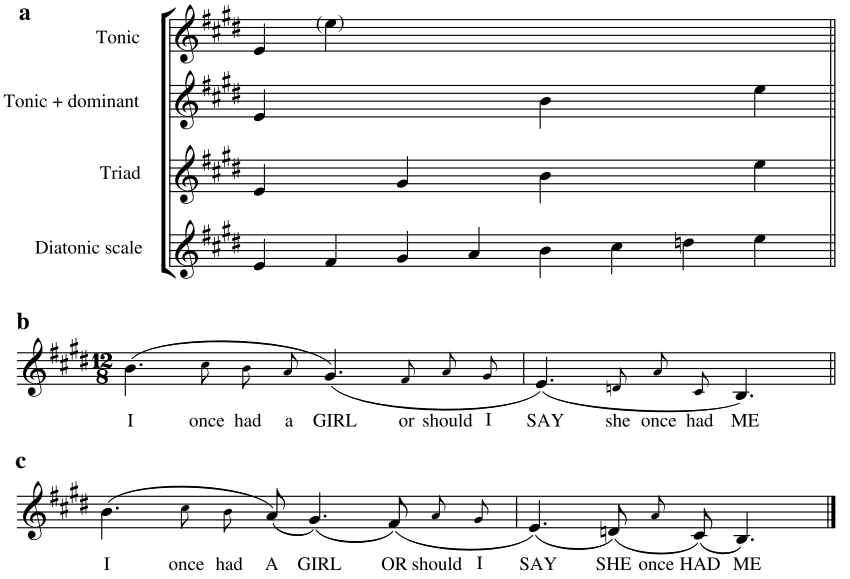

Fig. 15. (a) pitch space of E major with flatted 7th; (b) hierarchical analysis of the first phrase of *Norwegian Wood* so that its main notes form a descending triad; (c) elaboration of (b) so that the main notes form a descending diatonic scale.

D instead of D#, because of the modal coloring of this song). The shorter notes in the phrase are understood as transitions from one anchor to the next; for example, the notes C#–B–A (“once had a”) take the melody from B (“I”) to G# (“girl”). This analysis is given in Fig. 15b: the slurs connect the anchoring arpeggio, and the transitional notes appear in smaller note heads.

Within the transitions there is the further organization shown in Fig. 15c. The most direct way to move from the first anchor, B, to the second, G#, is via A, the note between them in the next layer down in pitch space. And indeed this note is present in the transition C#–B–A (on the word “a”). As a consequence, we understand the A as essential to the transition, and the C# and B as ornamental. Another way to think of this distinction is to say that if only the A were present and the other two notes were deleted, we would hear a smooth stepwise movement from B to G#. The forces within pitch space are such that the A is attracted to the G#. Similarly, between the second and third anchors (G# and E), the transition contains the direct transition F# (“or”) followed by two ornamental notes. Between the third and fourth anchors, the tune takes two stepwise movements to get from E (“say”) down to B (“me”); hence both the D (“she”) and C# (“had”) are essential to the transition, and only the A (“once”) is ornamental. The anchors plus the essential transitions form a descending scale. The melody is structured in terms of different levels of abstraction: at a relatively abstract level it spells out a tonic chord; at another, closer to the surface, it spells out a descending scale. A note of the melody that belongs to more abstract levels is understood as relatively stable, a point around which surrounding notes are heard.

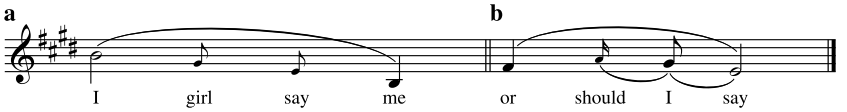

Fig. 16. (a) Hierarchical differentiation of (a) the descending triad in Fig. 15b; (b) the third sub-group in the phrase.

We can detail this organization still further. At the very largest scale, this phrase is heard as moving from the first note, a high B (“I”), to the last, the B an octave lower (“me”). The intermediate anchors G# and E are transitional in this octave descent, as shown in Fig. 16a. (In the notation here, note values represent not durations but structural importance: the larger the note value, the more stable the pitch.) At the smallest scale, consider the little group F#–A–G#–E (“or should I say”). We have already said that the middle two notes, A–G#, are ornaments to the stepwise descent from F# to E. But they are moreover differentiated from each other, in that a transition F#–G#–E is more stable than F#–A–E in this context. Thus, the G# is the more essential of the two in this context, and the A is understood as an embellishment on the way from F# to G#. This is shown in Fig. 16b.

This analysis demonstrates that the stability of a note cannot be determined by its pitch alone. The same pitch may play a different role at different points in the melody, and its relation to adjacent notes – at each level of abstraction from small to large – is crucial. For instance, the pitch G# occurs first in *Norwegian Wood* as an anchor (“girl”) and then again as an ornament (“I”). In the first case, its role is as a transition in the large-scale spanning of the octave: it relates most directly to the adjacent anchors B and E, as shown in Fig. 16a. In the second case, its role is as a transition from the “essential” transition F# to the anchor E; it is reached from the F# by way of the still more ornamental A between them, as shown in Fig. 16b.

The result of this analysis is a hierarchical, recursive structure in which each note of the melody is related to more stable notes in the structure. The related notes need not be adjacent at the musical surface, but they must be adjacent at some level of abstraction. This kind of organization, which in music–theoretic circles is often called a *pitch reduction* (in the tradition of Schenker, 1935), is notated in GTTM as a tree structure; see Fig. 17. The higher a note’s branch is attached in the tree, the more essential the note is to the skeletal structure of the melody; the lower a note’s branch is attached, the more ornamental the note is.[8]

Reductional structure plays a role in determining degrees of tension and relaxation in a melody. A relatively stable note in a reduction – one that is attached relatively high up in the tree – marks a relatively relaxed point in this contour. A relatively unstable note – one that is attached relatively low – marks a relatively tense

[8] Fig. 17 shows each note attached not just between the two surrounding notes, but as either a right branch to what precedes or a left branch to what follows. Space precludes justifying this aspect of the notation here; see GTTM, chapters 8–9, and TPS, chapter 1, for discussion. In particular, we are glossing over the important distinction in GTTM between Time-Span Reduction and Prolongational Reduction.

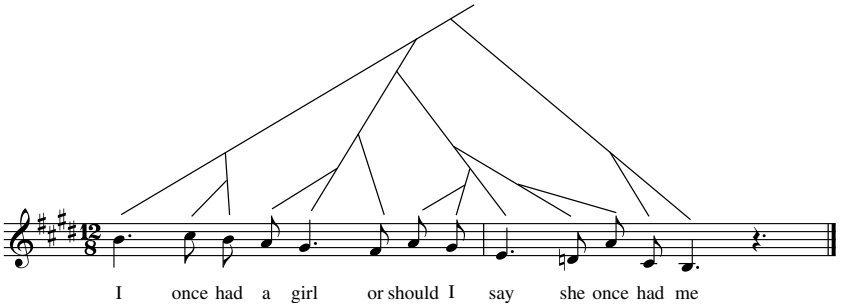

Fig. 17. Tree representation of the hierarchy of pitches in the first phrase of *Norwegian Wood*.

point. To the degree that a point in a melody is tense, it calls for relaxation, that is, continuation to a point of greater stability. One can see this particularly with the left branches in the tree in Fig. 17: C# (''once'') is attracted to and resolves on B (''had''); similarly for A (''a'') and G# (''girl''), and so forth.

TPS works out quantitative metrics of degrees of tension and attraction within a melody and/or chord progression at any point. The metrics are supported by experimental investigation (Lerdahl & Krumhansl, 2004; Smith & Cuddy, 2003). Their details are beyond the scope of the present article, but we can mention some factors beyond depth of embedding that contribute to the experience of melodic tension and attraction. Compare the three occurrences of the note A in *Norwegian Wood* (''a'', ''should'', and ''once''). The first is approached stepwise from above and goes on stepwise to the next note. This transition is melodically simple and smooth. Hence, the small tension peak engendered by this particular note is attributable to its being a transition between B and G# rather than a principal or anchor note of the melody.

Next look at the second A (''should''). This is approached by an upward leap of a third from F#, a more demanding transition; however, like the previous A, it is attracted to and resolves naturally into the following G#. Hence, it is a somewhat greater point of tension in the phrase. Finally, the third A (''once'') is approached by a quite large upward leap of a fifth from the preceding D; it does not resolve in a comfortable way to the proximate G# but jumps back down to C#. Furthermore, none of the three notes D–A–C# are high in the tonal hierarchy of Fig. 15a. Thus, the A is an unusual interposition in the contour and engenders the most tension of any note in the phrase. The overall tension contour of the phrase, then, is a gradual decrease in tension, as the main pitches in the reduction go downward. But in the interstices, each successive transition from one stable point to the next is tenser than the preceding one.

We should emphasize that the tension associated with the third A does not arise from a violation of the listener's conscious expectations. Most of us have been familiar with *Norwegian Wood* for many years, so this note is by all means consciously expected. Rather, the tension is a consequence of the unconscious attractive forces (or grammatical expectations) on the melody at this point. The unconscious expectation of the A to resolve to G# in its own register is overridden by the competing

unconscious expectation of the lower line, of what is now a polyphonic melody, to continue its scalar descent from D to C# and B. This contrast of conscious and unconscious expectations relies on a modular view of music processing: the encapsulated music module, constructing the structure of the music in real time, unconsciously computes its moment-to-moment tensions and attractions regardless of the listener's conscious memory (Jackendoff, 1992; Meyer, 1967).

The tree structure in Fig. 17 gives an idea of the cognitive structure associated with melody (Q1 from the outset of our discussion). We must next ask what cognitive principles allow the listener to infer (or construct or derive) such a structure on the basis of the musical surface (Q2). As in the case of grouping and meter, the principles are along the lines of defeasible constraints. Here are some of them, very informally. They are already illustrated to some degree in the discussion above. The overall "best" or "most stable" structure for a piece as a whole results from the optimal interaction of all the principles, applied at each point in the piece.

Local good form:

- (Interaction with contour) Pitches that are approached by small intervals from preceding notes (at any level of reduction) should be relaxed relative to their context (at the requisite level of reduction).
- (Interaction with meter) Relatively relaxed melodic points should be aligned with beats of relatively strong metrical importance.
- (Interaction with pitch space) Pitches that are relatively stable in the tonal pitch hierarchy should be relatively relaxed melodically. In an idiom with harmony, this principle is supplemented by:
  - Pitches that align with (or are consonant with) the current harmony should be relatively relaxed.
  - Harmonies close to the current tonic are relatively relaxed.

Global good form:

- (Interaction with grouping) The most relaxed points in a group should be at or near the beginning and end.

Pitch considerations at the musical surface:

- Lower in the pitch range is relatively relaxed.
- Less extreme in the pitch range is relatively relaxed.

Notice the effect of the last two principles. The first determines that rising melodic lines are generally tensing, and falling melodic lines are generally relaxing. However, it also interacts with the second principle. Both constraints mark very high notes as tense. Going down reduces tension until we begin to reach extremely low notes, at which point tension increases again (think of a tuba showing off how low it can go).

These principles collectively require an integration of all the musical factors reviewed above: the two components of rhythm (grouping and meter), the pitch-

space hierarchy, and the formation of melodic tension and attraction contours from sequences of pitches. In any piece of music, these factors constantly go in and out of phase, and their interplay is a further source of tension (we might call it "meta-tension"): situations where principles are in conflict with each other are more tense than those where the principles produce concordant (and redundant) outcomes (Temperley, 2001, chapter 11). A general consequence is that the end of a melody (which is necessarily the end of a large group) will tend to be a point of minimal tension – on a strong metrical beat, on the most stable pitch in the tonal hierarchy, and in a relatively low register – so that all the principles are maximally satisfied. This does not mean that all melodies will end that way (indeed, the melody of *Norwegian Wood* ends not on the tonic but on the dominant), nor do these principles dictate in general that melodies must have any particular shape.

Different musical idioms, as we have seen, specify a range of possible metrical structures and pitch spaces. In addition, most idioms have a stock of conventionalized melodic and rhythmic formulas that can be incorporated as building blocks of melodies, interspersed with freely composed segments. This stock of formulas might be thought of as rather like the lexicon in a language, but it differs in two important ways. First, although sentences have to be built entirely out of stored words and morphemes, melodies need not consist entirely of melodic and rhythmic formulas. Indeed, melodies are individual to the extent that they are not so composed. Second, musical formulas need not be just fragments of a musical surface but can be quite abstract frameworks in terms of which melodies are constructed, such as the 12-bar blues, the 32-bar pop song form, or classical sonata form. These abstract patterns can be freely modified at the composer's whim. So if the stock of formulas resembles anything in language, it is not words and morphemes, but rather a continuum running from words and morphemes through idiosyncratic constructions to general grammatical rules, as has been posited by recent "constructionist" approaches to linguistic theory (e.g. Culicover & Jackendoff, 2005; Culicover, 1999; Goldberg, 1995; Jackendoff, 2002; Tomasello, 2003).

How does a listener acquire principles of melodic organization (Q3 above), and what innate resources assist this acquisition (Q4)? To understand a melody in a given idiom, one must have sufficient exposure to the idiom's grouping and metrical possibilities, its pitch space, and at least some of the melodic and rhythmic formulas that are essential to finding similarities and differences among pieces of music within the idiom. Many of these factors are addressed in Trehub and Hannon (this volume), and we defer to their exposition. We conjecture, however, that one does not have to learn the basic principles of "good form" listed above, which assign contours of tension and attraction to the musical surface. Rather, these are part of the human capacity for music. But because these principles refer to an idiom's metrical and tonal particulars and its stock of conventionalized formulas, all of which must be learned, different idioms will yield different tension and attraction contours, calibrated to these particulars. Thus each idiom will have its own characteristic structures, created out of the interaction of idiom-specific tonal and metrical principles with universal principles of tension and relaxation.

As with other aspects of music, we can then ask which principles of melodic organization are specific for music (the narrow capacity), and which follow from more general properties of cognition (the broad capacity) (Q5). Melodic organization does share some principles with other domains. The issues of pitch range and the typical downward direction of melody are general and apply to all sorts of mammalian call systems. That small melodic intervals produce less tension than do large intervals is also general, following from gestalt principles of proximity and good continuation as well as muscular constraints on vocal production. That a melody can break up into separate streams or voices is a musical instantiation of general principles of auditory scene analysis (Bregman, 1990). And that principles of melodic organization are defeasible and interactive is characteristic of many cognitive systems.

But tonal space – the system of fixed pitches and intervals, and its hierarchy of pitches, chords, and keys and distances among them – is entirely specific to music and therefore to melodic organization. So are the principles for the treatment of dissonance that arise in conjunction with a particular form of tonal space. And so are the principles for understanding melodies in terms of pitch reductions. Moreover, although recursion of some sort or another is widespread among human cognitive systems (Pinker & Jackendoff, 2005), the kind of recursion appearing in pitch reductions seems to be special to music. In particular, there is no structure like it in linguistic syntax. Musical trees invoke no analogues of parts of speech, and syntactic trees do not encode patterns of tension and relaxation.[9] Insofar as musical tree structures are specializations for music, so must be the principles of good form connecting them to the musical surface.

If the principles of tonal systems and melodic structure are specializations for music, it is of interest to ask what their evolutionary precursors might be, based on results from non-human subjects. D'Amato (1988) sets cebus monkeys and rats the task of distinguishing between two different pitch contours. He finds that they make their choices not on the basis of melodic contour, but rather on the basis of one or two distinctive points within the contour. By contrast, Wright, Rivera, Hulse, Shyan, and Neiworth (2000) find that rhesus monkeys *can* make reliable same-different judgments on 6-note melodies. Moreover, they find that judgments are significantly more reliable when the melodies conform to a diatonic pitch space than when they do not. In both experiments, octave transpositions of contours were judged as similar to the original; in the latter case, transpositions of a tritone were typically judged different.

Our take on these experiments is that they show monkeys to have some command of pitch, interval, and contour. They do not, however, show that monkeys are capable of music cognition in the human sense. In particular, the monkeys' ability to recognize octave transposition may well be a matter of psychoacoustics rather than melodic perception, along lines suggested in Section 3.2. Humans can perceive a mel-

[9] Musical trees do, however, share properties with phonological stress trees (GTTM, chapter 11) and hierarchical syllabic patterns in poetry (Lerdahl, 2003). As in the case of metrical structure, poetry apparently "borrows" musical structure – invokes the musical capacity – in a way ordinary language does not.

ody as the same melody at any degree of transposition. More telling is that these experiments do not test for a sense of tonal stability or for a capacity to understand melodies in terms of reductions.

We conclude that the properties of music do not all follow from other more general cognitive principles. Thus, there is a genuine need to posit a narrow musical capacity. Perhaps some homologues to pitch spaces and reductional structures will emerge as we come to understand a wider range of human cognitive structures in formal detail. But for the moment these look like musical specializations.

4. Remarks on affect in music[10]

4.1. Defining the problem

We turn now to the question that practically everyone other than music theorists considers the primary point of interest in the psychology of music: the relation of music to affect. We must necessarily be speculative here, but it is worth bringing up a number of possibilities in the context of the present discussion of musical structure. Our overall view is that there are several distinct converging routes from musical surface to musical affect, which range from fairly general psychological responses to effects that are quite specific to music.

The issue that we address here in terms of "affect" is usually phrased as the relation of music to *emotion*, as in the titles of two prominent books, Leonard Meyer's *Emotion and Meaning in Music* (1956), and Juslin and Sloboda's collection *Music and Emotion* (2001a). We prefer the term *affect* because it allows a broader inquiry than *emotion*. For example, several papers in Juslin and Sloboda's volume seek to identify passages of music with replicable basic emotions such as *happy, sad,* and *angry*. We do not deny that, say, *Yellow Submarine* is happy and *Michelle* is rather sad, and that these judgments are correlated with their respective modes (major and minor), rhythms, and pitch ranges. But there is a vastly wider range of descriptors that deserve characterization. A passage of music can be gentle, forceful, awkward, abrupt, static, earnest, opening up, shutting down, mysterious, sinister, forthright, noble, reverent, transcendent, tender, ecstatic, sentimental, longing, striving, resolute, depressive, playful, witty, ironic, tense, unsettled, heroic, or wild. Few of these can be characterized as emotions per se. And while a passage of music can be disgusting, it is hard to imagine attributing to a piece of music the basic human emotion of disgust, i.e., to say the music is "disgusted".

Some philosophical/conceptual issues have to be addressed to approach the problem (Davies, 1994, 2001 offers a detailed survey of positions on these issues; we concur with most of his assessments). First, we do not want to say that affect or emotion is the "meaning" of music, in the sense that language is meaningful. Unlike lan-

[10] The discussion in this section is indebted to chapters in Juslin and Sloboda (2001a), especially Juslin and Sloboda (2001b), Sloboda and Juslin (2001), Davies (2001), Gabrielsson and Lindström (2001), and Scherer and Zentner (2001).

guage, music does not communicate propositions that can be true or false. Music cannot be used to arrange a meeting, teach someone how to use a toaster, pass on gossip, or congratulate someone on his or her birthday (except by use of the conventional tune). Moreover, it trivializes music to say, for instance, that one piece "means" happiness and another "means" sadness. Under this interpretation, all happy pieces would mean the same thing – so why should anyone bother to write another happy one? Insofar as music can be characterized as meaningful (and insofar as it is produced with an intention to make it meaningful), it is so in the generalized sense that we say any sort of experience is meaningful, namely that it makes an affective impression on us.

Raffman (1993), like us, rejects emotion pure and simple as the meaning of music. She speaks of the meaning of music as a "quasi-semantics", consisting of the feelings that one experiences upon hearing the structure of music in detail. Although we agree with the sentiment, we prefer not to invoke the term "semantics" in this context, inasmuch as nothing like propositional inference ensues from the perception of music. We concur with Davies (1994) in using the term "the listener's understanding of a piece of music" to denote the cognitive structures (grouping, metrical, and tonal/reductional) that the listener unconsciously constructs in response to the music. We would then characterize Raffman's sense of "musical meaning" as the affects that the listener associates with the piece by virtue of understanding it.

A further conceptual difficulty: one might think that affects ought to be ascribed only to sentient agents such as people and perhaps animals. So what does it mean to say a string of notes is playful or sentimental? This question actually has a scope wider than music. How can we characterize a novel, a poem, or a painting (especially an abstract painting) as cheerful, static, or playful? It does not necessarily mean the characters or objects in it are cheerful, static, or playful. Nor need we be talking about the emotions of an author or performer, since we can describe a natural landscape as gloomy or wild. (Kivy, 2001 argues that these "emotive properties" in music are "perceptual properties pure and simple". We disagree, as will be seen below.)

This problem is not confined to aesthetic experience. To call something *boring* or *valuable* ascribes to it a putatively objective characteristic akin to its size or temperature. Yet something can not be boring if no one is bored by it; something cannot be valuable if no one values it. That is, such evaluative predicates covertly involve the reactions of an observer. Jackendoff (2006, chapter 7) proposes that this covert observer is understood as a non-specific generic individual of the sort invoked by German *man* and French *on* – and in English by the generic *one* and some uses of unstressed *you* ("You donot hear the Grieg piano concerto played much any more"). We propose that the affective predicates applying to music are of this kind: a listener deems a passage of music mysterious if it is judged to evoke a sense of mystery in the generic observer – usually, with the listener him/herself taken to stand in for the generic observer.

We have to be careful about what is intended by "evoke" here. It does not necessarily mean "cause to experience". Aside from masochists, people do not normally want to deliberately make themselves sad; yet people flock to hear all sorts of sad

music. Again, this is a more general issue. People have flocked to *Hamlet* and *Oedipus Rex* for centuries, too. And people like to eat very spicy foods and to indulge in hazardous activities such as caving (Bharucha, Curtis, & Paroo, this issue; Davies (1994)). A solution to this puzzle is that the perception of music and drama is *framed*, in the sense of Goffman (1974): it is approached with a mindset distinct from ordinary life, like a picture in a frame. Experiencing art is not the only possible frame. Such frames as practicing a task and playing games also detach an activity from normal goals. In listening to music, perceivers can choose how much to invest themselves in the material within the frame and how much to remain detached; the emotional effect is greater, the more one invests in the framed material – while still recognizing it as framed. Composers and performers have similar choices: one need not feel sinister to compose or perform sinister music.[11]

Musical activity can itself be embedded in frames. For instance, when practicing a piece of music for future performance, one often holds affect entirely in abeyance. In addition (in the spirit of Goffman), consider the different frames or mindsets involved in listening to a chorus perform, in performing in a chorus, in participating in congregational singing as part of a religious service, in singing the national anthem at a sporting event, in singing quietly to oneself while walking, in singing a lullaby to a child, and in experiencing (or *not* experiencing) muzak or background music in a film. Each of these circumstances changes the overall stance in terms of which musical affect is experienced.

Part of the normal framing of music is an association with aesthetic appreciation, which can occur in any modality (including food!). Raffman (1993, p. 60) speaks of a "peculiar aesthetic emotion"; Kivy (2001, p. 105) speaks of the affect that comes with this aesthetic engagement as "an enthusiasm, an intense musical excitement *about* what I am hearing".[12] In earlier times, this might have been called "appreciation of beauty". But in 20th-century Western culture, it became possible to detach the framing of an object or activity as consciously produced art from its perceivable properties, permitting the production of such famous examples as Andy Warhol's Campbell's soup can and pieces by John Cage in which the performers tune several radios simultaneously at random. These cases rely on the perceiver experiencing an affect associated with an aesthetic frame that transcends the content of the object or activity under contemplation.

Beyond the general frame of aesthetic experience, music partakes in other wide-ranging sources of affect. One is the affect that goes with admiring virtuosity of any sort, be it by a violinist, an acrobat, a star quarterback, or an ingenious criminal. Another is the affect provoked by nostalgic familiarity ("Darling, they're playing our song"), which is shared by familiar foods, customs, and geographic locales, among

[11] RJ recalls being admonished by a chamber-music coach, "*You*'re not supposed to dance, you're supposed to make the *audience* want to dance". Indeed, effective performance (at least in some genres) requires a considerable degree of detachment.

[12] However, Kivy sees this engagement as the *only* source of affect in music, rejecting the other sources of affect discussed here. Like many others, he seems to presume that there can be only a single source of affect; our position is that there are multiple interacting sources.

other things. The situation in music becomes more complex when composers deliberately build such affects into their music. With respect to virtuosity, well-known display pieces, from coloratura arias to concertos to hot jazz improvisation, tap into this vein of appreciation. As for nostalgic (or perhaps ironic) evocation, consider Beatles songs such as *Your Mother Should Know* (''Let's all get up and dance to a song/That was a hit before your mother was born'') and *Honey Pie* (''Honey Pie, you are making me crazy/I'm in love but I'm lazy ...'') that are written in a style of an earlier era. Also in this category belong the use of folk elements in works by such classical composers as Haydn, Mahler, and Bartók. To achieve the appropriate affect, of course, one must be familiar with the style alluded to and its extra-musical connotations. For example, one's appreciation of Stravinsky's *Rite of Spring* is amplified by acquaintance with some Russian folk music.

There are also circumstances in hearing music where the frame is dropped altogether, described perhaps as ''losing oneself in the music'' or ''getting swept up in the music''. This sensation, too, is not peculiar to music. It appears, for instance, in states of religious ecstasy, sexual abandon, and mob behavior.

4.2. General-purpose components of musical affect

Some aspects of affect in music are easily attributable to general characteristics of audition. A clear case is the startle (and fear?) reaction to sudden loud noises, which carries over to sudden loud outbursts in music. Some sounds are inherently pleasant (songbirds) or unpleasant (buzz saws), and music with similar acoustic character evokes similar affect.

Equally clear are musical phenomena that simulate affective characteristics of vocal production. Not only human but much mammalian communication modulates vocal pitch, volume and timbre to convey threat, reconciliation, fear, excitement, and so on (Hauser, 2001). These modulations can be carried over into musical performance, sometimes in the character of melodic contour, but often also in the performer's manipulation of vocal or instrumental tone production (Bharucha et al., this issue; Juslin & Laukka, 2003; see also Section 4.3). Listeners respond affectively to such manipulations in the same way as they respond to the corresponding vocal communications – that is why we can speak of ''sighing violins''.

At a larger scale, overall affective tone can be influenced by the pitch range of a melody: as in speech, small range and overall low pitch correspond to subdued affective tone; wide range corresponds to more outgoing and expressive affective tone (Juslin & Laukka, 2003). This disposition can be used dynamically. Much of the melody in *Michelle*, shown in Fig. 18, moves in a relatively small range in the mid-to-low vocal range, with a generally descending contour (bars 1–6); the overall affect is subdued. But this is interrupted by a repetitive and affectively passionate outburst in a higher pitch range (bars 7 and 8), paralleling the text, which gradually subsides into the original range (bars 9 and following).

At a larger scale of organization, we find a source of affect that is shared with language: what might be called ''rhetorical effects''. A simple example is the use of rep-

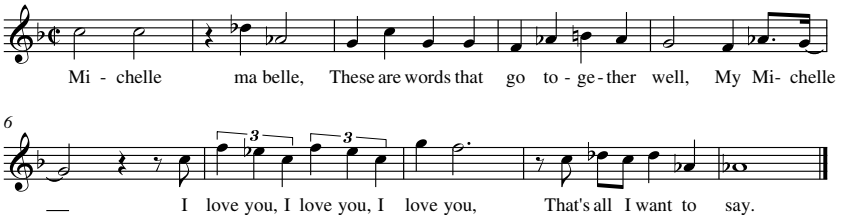

Fig. 18. An except from *Michelle*.

etition as a source of intensification, such as the music for "I love you, I love you, I love you" in Fig. 18. Related to this is use of a musical refrain to which the melody returns, perhaps parallel to the use of refrains in the rhetorical mode of evangelical preaching.

Looking at still larger scales of organization, consider the treatment of extended musical forms. A piece may consist of a sequence of repeated verses (as in a typical folk song), or a sequence of unrelated episodes (as in a chain of dances or the second half of the Beatles' *Abbey Road*). Alternatively, there may be large-scale cohesion that involves more than concatenation. Most simply, after one or more unrelated episodes, a repetition of the initial section may return (e.g., da capo aria, minuet and trio, or the Beatles' *A Day in the Life*). A piece may gradually build in intensity to a climax, which is resolved triumphantly or tragically, or by the restoration of repose. A piece may begin with an introductory passage that sets a mood from which the rest of the music departs; examples are the "vamp" at the beginning of a pop song and the introductory quiet passage at the beginning of a raga before the tabla drums enter. A piece may incorporate highly embedded dependency structures, as in classical sonata form. There may be stretches of music where nothing of consequence happens, and tension is built only by the passing of time and the sense that something has to happen soon. These structural elements can be combined and embedded in various ways, creating a wide range of large-scale forms, all of which have some sort of affect associated with them.

These larger structures have a great deal in common with structure in narrative and drama. The "vamp" plays much the same role as scene-setting in narrative; one is creating a mood and waiting for the action to begin. The plot of a novel or play often involves a slow building of tension to a climax, followed by rapid denouement. Often the resolution is postponed by long stretches of inaction, or alternatively by deflection to a subplot. Of course, the literary devices used to build these dramatic structures are entirely different from those used in music – but nevertheless the overall large-scale rhythm of tension and relaxation are strikingly similar. We conjecture that both music and language make use of idealized event structures in terms of which humans understand long-range connections of tension and resolution among events.

In short, many affective qualities of music and their integration into larger frames are shared with other aspects of human activity and experience. Setting these aside, we now turn to what might be to some degree more specific to music.

4.3. Affective characteristics more specific to music

There remains a core of affective expression in music that we will now address provisionally. We believe that, in addition to the factors mentioned above, musical affect arises in large part from its relation to physical patterns of posture and gesture. That is, the use of the term "musical gesture" is directly motivated by its relation to physical gesture. (This position is shared by Bierwisch, 1979 & Davies, 1994; we differ from Juslin & Laukka, 2003, who claim that musical affect comes only from its similarity to vocal affect.) Posture and gesture are strong cues for affect in others; they are usually produced unconsciously and are often detected unconsciously. We are immediately sensitive, for example, to a person in the room having a slumping, depressed posture or making a joyful gesture.

Some of the cues for recognition of affect in others do not depend on our first characterizing the individual as human and then judging affect. Rather, the character of motion alone can convey affect and in turn lead to ascription of animacy. This point is strikingly demonstrated by the well-known experimental cartoon by Heider and Simmel (1944), in which triangles move about in such a fashion that observers cannot help seeing them as characters that act aggressively and sneakily, and that experience anger, frustration, and joy. Damasio (1999) and Bloom and Veres (1991) report related experiments. Temporal patterns in music can similarly invoke perceptions of affect.

Evidence for this view is the deep relationship between music and dance. Dancing does not just involve timing one's movements to the beat of the music. One could waltz in time with a march, but the juxtaposition would be incongruous, because the gracefulness of waltz movements is sharply at odds with the rigidity and heaviness of march music (see Mitchell & Gallaher, 2001 and references therein for experimental support of these judgments). In our personal experience, even young children appreciate these differences in the character of music and improvise dances accordingly (see Trehub, 2003 on the sensitivity of very young children to musical affect). Similarly, orchestra conductors do not simply beat time: rather, their posture and the shape of their gestures convey the affective sense of the music. A conductor moves entirely differently when conducting a waltz and a march, and players respond accordingly. The ability of dancers to convert musical into gestural shape and that of performers following a conductor to do the reverse is instinctive (though it can be refined by training). Equally instinctive is the ability of audiences to interpret these relationships spontaneously. We think that such abilities are intrinsic to musical affect.

For another bit of evidence for the correlation, recall the list of descriptive terms we used for musical affects at the beginning of Section 4: gentle, forceful, awkward, abrupt, static, earnest, opening up, shutting down, mysterious, sinister, forthright, noble, reverent, transcendent, tender, ecstatic, sentimental, longing, resolute, depressive, playful, witty, ironic, tense, unsettled, heroic, or wild. All of these are also used to describe gestures, postures, facial expressions, or some combination thereof.

There is a further distinction to make. So far we have spoken of the perception of affect in others. Sometimes there is musical affect associated with such perception; for instance, witty, mysterious, or sinister music does not make one feel witty, mysteri-

ous or sinister, but rather as if in the presence of someone or something witty, mysterious, or sinister. We might call this *reactive* affect. However, the more important variety of musical affect is experienced as though in empathy or attunement with the producer of the gesture: one may well feel joyous, reverent, or unsettled upon hearing joyous, reverent, or unsettled music. We might call this *empathic* affect (the term "emotional contagion" is sometimes used in this connection). Dancing, then, can be taken as externalizing empathic affect – converting it into posture and gesture.[13]

If our line of reasoning is on the right track, it is somewhat of a misdirection to look for a direct connection of music to emotion. The affects themselves, and their connection with posture and gesture, belong to general psychology. Rather, there are three problems for psychology of music per se: first, how features within music itself are correlated with affective posture and gesture; second, how these features come to be treated empathically in addition to reactively; and third, what brain mechanisms are responsible for these effects.

Our analysis of musical structure allows us to approach at least the first of these questions. The features in music that connect to posture and gesture can be found at two time scales. The larger time scale, the *macro* level, concerns tempo, rhythm, broad dynamics, melodic contour, and melodic and harmonic relationships. This is the level to which the structures discussed in Sections 2 and 3 pertain. To the extent that music is notated or individuated into remembered pieces (*Norwegian Wood*, *Happy Birthday*, the Grieg piano concerto) it is the macro level that differentiates them. At a smaller time scale, the *micro* (or *nuance*) level, musical affect can be manipulated through micro timing of the amplitude, onset, and offset of individual notes. These effects are mostly not notated (or, in classical music, they are notated by expression terms such as *dolce* and *risoluto*), and they are more open to the performer's discretion. In the Western classical tradition, the macro level is mostly determined by the composer and the micro level by the performer. But in genres such as jazz, the performer's discretion extends into the actual choice of notes, and many other genres hardly distinguish composer from performer at all.

Our discussion of rhythm and melody above lays out many of the parameters of the macro level. Basic tempo is fairly straightforward: fast, slow, or moderate tempo evokes fast, slow, or moderate movement, hence corresponding degrees of arousal and corresponding affective possibilities. In addition, the rhythmic character of a melody plays a role in affect – for example, a steady flow versus dotted rhythms (alternating long and short) versus a variety of note lengths. Steady flow may correspond more closely to motor activity such as walking or hammering. A varied flow

[13] Included in empathic affect would be the sensation of resisting or giving in to outside forces on the body, as in pushing through an unwilling medium or being stopped in one's tracks; these too can be evoked by music. This account might also explain the earlier observation that music does not express disgust: there is no canonical reaction, neither reactive nor empathic, to someone else's evincing disgust. Kivy, 2001, again insisting that there is only one kind of affect, recognizes the validity of only reactive affect, not of empathic affect.

It is hard to resist appealing to the possible contribution of mirror neurons (Rizolatti, Fadiga, Gallese, & Fogassi, 1996) to the ability to experience and externalize empathic affect, though our impression is that currently not enough is known about them to cash out the appeal in the necessary fine detail.

may correspond more closely to expressive speech, with its clusters of syllables of varying length and interspersed pauses, or in other cases, to a sense of hesitating movement. These patterns can convey an overall affective tone.

A more detailed correspondence between musical structure and affective response is in the tension-relaxation contour of melody discussed in Section 3.5. The terms "tension" and "relaxation" give away the correspondence: these are descriptors primarily of bodily states and only derivatively of strings of sounds. Melodic attraction, the complementary aspect of melodic tension, is equally embodied. For example, the leading tone (the note immediately below the tonic) seems impelled toward the stable tonic a half step above it; it metaphorically "wants" to resolve on the tonic. Pitch contours that move against the attractive forces seem to have a will of their own, as if they were animate agents. The notes of a melody progressing through pitch space act like the triangles in Heider and Simmel's cartoon.

It is the tension and attraction contours that above all give music its dynamic quality. Music does not just express static emotions or affects such as nobility or gloom. It moves from one state to another in kaleidoscopic patterns of tension and attraction that words cannot begin to describe adequately.

Unlike many aspects of musical affect discussed above, this one is heavily dependent on the listener's ability to use the grammar of the musical idiom to construct tension and attraction contours. Thus, as pointed out by Davies (1994), this aspect of musical affect will be available in detail only to experienced listeners, and will not be available to amusic subjects (Bharucha et al., this issue; Peretz, this issue).

At the micro time-scale of musical expression, the performer can manipulate details of individual notes and transitions between notes to amplify the expressivity of the macro level (these cues for affect are not discussed by Davies, 1994; and little by Raffman, 1993). Sloboda and Juslin (2001) use the term "vitality affects", introduced by Stern (1985), to describe the affective aspect of these micro-level manipulations. Such affective details are not confined to music: they can appear in the global modulation of any motor activity, from dancing to conducting an orchestra to bowing a violin. They are evident in perception of another's movement, and are taken as cues for affect.

Master performers have excellent command of this modulation of the musical surface and are able to creatively vary the micro-level character from note to note in coordination with the macro-level tension and attraction contours. Performers have at their disposal micro-level variations in tuning, vibrato, amplitude envelope, and tone color, possibilities for gliding up or down to a pitch (portamento), moment to moment alterations in overall tempo, and delays or anticipations in the onsets of individual notes (Clarke, 1999; Gabrielsson, 1999; Palmer, 1996; Repp, 1998, 2000; Sloboda & Lehmann, 2001; Windsor & Clarke, 1997). Genres differ in what sorts of micro-level variations are considered acceptable or stylish. For example, classical music of the Romantic period permits far wider micro variation in tempo (the underlying beat) than do Baroque music or jazz, but jazz calls for far more extensive pitch modification, vibrato, and variation in early and late onset of notes in relation to the beat than does any classical genre. These possibilities for micro variation, and the differences in them among styles, cannot be conveyed by explicit

verbal description: there are no words for them. They are learned by imitation and "intuition" and are passed down by tradition. More than anything else, it is mastering the appropriate parameters for manipulating micro structure that constitutes "getting the feel" of a musical style and that distinguishes a truly artistic performer from one who is just "playing the notes" (not enough manipulation) or who is "tasteless" (exaggerated manipulation).

The upshot is that one cannot talk of "the" affect of a piece of music. Rather, what makes musical expression special is its manifold possibilities for complex and ever-changing contours of affect, and for reinforcement and conflict among the various sources of affect in framing, general audition, interpretation of mammalian vocalization, and coding of patterns of gesture.

5. Concluding thoughts

We have attempted to provide a synoptic view of the full complexity of the musical capacity. Particularly in the last section on affect, it has proven virtually impossible to disentangle the parts that belong to the narrow musical capacity, those that are shared with other art forms, those that are shared with general auditory perception, those that are shared with vocal communication, and those that partake of more general cognition. However, at the very least, certain parts of the musical capacity emerge as special: isochronic metrical grids, tonal pitch spaces, and hierarchical tension and attraction contours based on the structure of melody.

These specifically musical features are richly interwoven with many other cognitive and affective mechanisms in such a way that it is impossible to think of music as a module in the sense of Fodor (1983). The looser sense of modularity in Jackendoff (2002), with many smaller interacting modules, may be applicable, perhaps along lines proposed by Peretz and Coltheart (2003). In particular, we would expect the existence of overlaps with language as well as dissociations from language, as have been observed (Patel, 2003; Peretz & Coltheart, 2003; Peretz & Hyde, 2003).

We have proposed that the aspects of musical affect that distinguish it from other sources of affect should be pursued not directly, but rather in terms of the interaction of musical structure with motor patterns that evoke affect. In these terms, a leading question ought to be how temporal patterns in audition can be linked with temporal patterns in posture and gesture, and how these are in turn linked with affect. These are issues for venues larger than music cognition alone, but music can provide a superb source of evidence.

Acknowledgments

We are grateful to Peter Bloom, Robert Ladd, Louise Litterick, Jacques Mehler, Caroline Palmer, Isabelle Peretz, David Temperley, and two anonymous readers for important suggestions on the formulation of this article. This research was supported in part by NIH Grant DC 03660 (R.J.).

References

Ashley, R. (2002). Do[n't] change a hair for me: The art of jazz rubato. *Music Perception, 19*, 311–332.

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

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 15–70.

Bernstein, L. (1976). *The unanswered question*. Cambridge, MA: Harvard University Press.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16*, 485–518.

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

Bharucha, J. J., Curtis, M., & Paroo, K. (this issue) Varieties of musical experience, doi:10.1016/j.cognition.2005.11.008.

Bierwisch, M. (1979). *Musik und Sprache: Überlegungen zu ihrer Struktur und Funktionsweise. Jahrbuch Peters 1978, 9–102*. Leipzig: Edition Peters.

Binder, A. W. (1959). *Biblical chant*. New York: Philosophical Library.

Bloom, P., & Veres, C. (1991). The perceived intentionality of groups. *Cognition, 71*, B1–B9.

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

Browne, R. (1981). Tonal implications of the diatonic set. *In Theory Only, 5*, 3–21.

Burling, R. (1966). The metrics of children's verse: A cross-linguistic study. *American Anthropologist, 68*, 1418–1441.

Clarke, E. (1999). Rhythm and timing in music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 473–500). New York: Academic Press.

Clough, J., Engebretsen, N., & Kochavi, J. (1999). Scales, sets, and interval cycles. *Music Theory Spectrum, 10*, 19–42.

Cross, I. (2003). Music, cognition, culture, and evolution. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music*. New York: Oxford University Press.

Culicover, P. W. (1999). *Syntactic nuts: Hard cases in syntax*. Oxford: Oxford University Press.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford: Oxford University Press.

D'Amato, M. R. (1988). A search for tonal pattern perception in cebus monkeys: Why monkeys can't hum a tune. *Music Perception, 5*, 453–480.

Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.

Davies, S. (1994). *Musical meaning and expression*. Ithaca and London: Cornell University Press.

Davies, S. (2001). Philosophical perspectives on music's expressiveness. In Juslin and Sloboda (2001a), 23–44.

Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl and Jackendoff's Grouping Preference Rules. *Music Perception, 4*, 325–360.

Deutsch, D. (1999). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic Press.

Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic Press.

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In Juslin and Sloboda (2001a), 223–248.

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

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* New York: Harper and Row. Boston: Northeastern University Press.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Gould, S. J., & Lewontin, R. (1979). The spandrels of San Marco and the Panglossian Paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society, B205*, 581–598.

Halle, M., & Keyser, S. J. (1971). *English stress: Its form, its growth, and its role in verse*. New York: Harper & Row.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.
Hauser, M. D. (2001). The sound and the fury: Primate vocalizations as reflections of emotion and thought. In Wallin, Merker, & Brown (2001), 77–102.
Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve?. *Science 298*, 1569–1579.
Hauser, M. D., & Fowler, C. (1991). Declination in fundamental frequency is not unique to human speech: Evidence from nonhuman primates. *Journal of the Acoustic Society of America, 91*, 363–369.
Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience, 6*, 663–668.
Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243–259.
Helmholtz, H. (1885). *On the sensations of tone*. New York: Dover.
Hindemith, P. (1952). *A composer's world*. Garden City, NY: Doubleday.
Huron, D. (1996). The melodic arch in Western folksongs'. *Computing in Musicology, 10*, 3–23.
Huron, D. (2001). Tone and voice: A derivation of the rules of voice-leading from perceptual principles. *Music Perception, 19*, 1–64.
Huron, D. (2003). Is music an evolutionary adaptation?. In I. Peretz & R. Zatorre (Eds.) *The cognitive neuroscience of music*. New York: Oxford University Press.
Jackendoff, R. (1992). Musical parsing and musical affect. *Music Perception, 9*, 199–230. Also in Jackendoff, *Languages of the mind*, MIT Press, 1992.
Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.
Jackendoff, R. (2006). *Language, culture, consciousness: Essays on mental structure*. Cambridge, MA: MIT Press.
Janata, P., Birk, J., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2003). The cortical topography of tonal structures underlying Western music. *Science, 298*, 2167–2170.
Juslin, P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code?. *Psychological Bulletin 129*, 770–814.
Juslin, P. N., & Sloboda, J. A. (Eds.). (2001a). *Music and emotion: Theory and research*. Oxford: Oxford University Press.
Juslin, P. N., & Sloboda, J. A. (2001b). Music and emotion: Introduction. In Juslin and Sloboda (2001a), 3–20.
Kivy, P. (2001). *New essays on musical understanding*. Oxford: Oxford University Press.
Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
Larson, S. (2004). Musical forces and melodic expectations: Comparing computer models and experimental results. *Music Perception, 21*, 457–498.
Lecanuet, J. P. (1996). Prenatal auditory experience. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings*. Oxford: Oxford University Press.
Lerdahl, F. (2001). *Tonal pitch space*. New York: Oxford University Press.
Lerdahl, F. (2003). The sounds of poetry viewed as music. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music*. New York: Oxford University Press.
Lerdahl, F., & Jackendoff, R. (1983). *A generative grammar of tonal music*. Cambridge, MA: MIT Press.
Lerdahl, F., & Krumhansl, C. (2004). La teoría de la tensión tonal y sus consecuencias para la investigación musical. In J. Martín Galán & C. Villar-Taboada (Eds.), *Los últimos diez años en la investigación musical*. Valladolid: Servicio de Publicaciones de la Universidad de Valladolid.
Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*, 249–336.
Locke, D. (1982). Principles of offbeat timing and cross-rhythm in southern Ewe dance drumming. *Ethnomusicology, 26*, 217–246.
McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
Meyer, L. (1967). On rehearing music. In Meyer (Ed.), *Music, the arts, and ideas*. Chicago: University of Chicago Press.

Mitchell, R. W., & Gallaher, M. C. (2001). Embodying music: Matching music and dance in memory. *Music Perception, 19*, 65–85.
Narmour, E. (1990). *The analysis and cognition of basic melodic structures*. Chicago: University of Chicago Press.
Nettl, B. (1960). *Cheremis musical styles*. Bloomington: Indiana University Press.
Nettl, B. (1973). *Folk and traditional music of the western continents* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
Oehrle, R. (1989). Temporal structures in verse design. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology: Rhythm and meter*. New York: Academic Press.
Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception, 13*, 433–453.
Paschen, E., & Mosby, R. P. (Eds.). (2001). *Poetry speaks*. Naperville, IL: Sourcebooks.
Patel, A. D. (2003). Language, music, syntax, and the brain. *Nature Neuroscience, 6*, 674–681.
Peretz, I. (2001). Music perception and recognition. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology* (pp. 519–540). Philadelphia: Psychology Press.
Peretz, I. (2003). Brain specialization for music: New evidence from congenital amusia. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music*. New York: Oxford University Press.
Peretz, I. (this issue) The nature of music: A biological perspective.
Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688–691.
Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences, 7*, 362–367.
Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition 95*, 201–236.
Phillips-Silver, J., & Trainor, L. J. (2005). Feeling the beat: Movement influences infant rhythm perception. *Science, 308*, 1430.
Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America, 38*, 548–560.
Pollard, C., & Sag, I. (1987). *Information-based syntax and semantics*. Stanford: CSLI.
Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
Pressing, J. (1983). Cognitive isomorphisms between pitch and rhythm in world musics: West Africa, the Balkans and western tonality. *Studies in Music, 17*, 38–61.
Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. Piscataway, NJ: Rutgers University Center for Cognitive Science.
Raffman, D. (1993). *Language, music, and mind*. Cambridge, MA: MIT Press.
Rahn, J. (1983). *A theory for all music: Problems and solutions in the analysis of non-Western forms*. Toronto: University of Toronto Press.
Rahn, J. (1996). Perceptual aspects of tuning in a Balinese gamelan angklung for North American students. *Canadian University Music Review, 16.2*, 1–43.
Rameau, J.-P. (1737). *Génération harmonique*. Paris: Prault fils.
Repp, B. H. (1998). The detectability of local deviations from a typical expressive timing pattern. *Music Perception, 15*, 265–289.
Repp, B. H. (2000). Pattern typicality and dimensional interactions in pianists' imitation of expressive timing and dynamics. *Music Perception, 18*, 173–211.
Rizolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.
Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7*, 532–547.
Sargent, W. (1964). *Jazz: A history*. New York: McGraw-Hill.
Scharf, B., Quigley, S., Aoki, C., Peachery, N., & Reeves, A. (1987). Focused auditory attention and frequency selectivity. *Perception & Psychophysics, 42*, 215–223.
Schenker, H. (1935). *Free composition*. Tr. E. Oster. New York, Longman, 1979.
Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In Juslin & Sloboda (2001), 361–392.
Singer, A. (1974). The metrical structure of Macedonian dance. *Ethnomusicology, 18*, 379–404.

Sloboda, J. A., & Juslin, P. N. (2001). Psychological perspectives on music and emotion. In Juslin & Sloboda, 2001, 71–104.

Sloboda, J. A., & Lehmann, A. C. (2001). Tracking performance correlates of changes in perceived intensity of emotion during different interpretations of a Chopin piano prelude. *Music Perception, 19*, 87–120.

Smith, N., & Cuddy, L. (2003). Perceptions of musical dimensions in Beethoven's Waldstein sonata: An application of tonal pitch space theory. *Musicae Scientiae, 7.1*, 7–34.

Spencer, A. (1996). *Phonology*. Oxford: Blackwell.

Stern, D. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. New York: Basic Books.

Temperley, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

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

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience, 6*, 669–673.

Trehub, S. E., & Hannon, E. E. (this volume). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition*.

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

Wertheimer, M. (1923). Laws of organization in perceptual forms. Reprinted in W. D. Ellis (Ed.), *A source book of Gestalt Psychology* (pp. 71–88). London: Routledge and Kegan Paul, 1938.

Windsor, W. L., & Clarke, E. (1997). Expressive timing and dynamics in real and artificial musical performances: Using an algorithm as an analytical tool. *Music Perception, 15*, 127–152.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and cctave generalization in rhesus monkeys. *Journal of Experimental Psychology. General, 129*, 291–307.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.