THE PSYCHOLOGY OF LEARNING AND MOTIVATION

Advances in Research and Theory

EDITED BY BRIAN H. ROSS

BECKMAN INSTITUTE AND
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
URBANA, ILLINOIS

Volume 46

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Academic Press is an imprint of Elsevier

WHAT IS MUSICAL PROSODY?

Caroline Palmer and Sean Hutchins

Prosody is the music of everyday speech.
(Wennerstrom, 2001).

I. Introduction

Music performance is prevalent everywhere. Concert attendance, recording sales, and critics' reviews suggest that people are sensitive to and have definite preferences for particular performances. Performers add variation to music; they manipulate the sound properties, including frequency (pitch), time, amplitude, and timbre (harmonic spectrum) above and beyond the pitch and duration categories that are determined by composers. These manipulations are called "musical expression," and cognitive psychologists have speculated on why performances contain expression. One hypothesis is that musical expression communicates emotions (Juslin & Sloboda, 2001). Another is that it clarifies structure (Kendall & Carterette, 1990). That structure may be indicated by a composer in a musical score (Lerdahl & Jackendoff, 1983), as well as by a performer in an interpretation—the performer's shaping of the music according to his or her own intentions (Apel, 1972). Each of these hypotheses relies on the fact that music expresses something, and the object of study is often what is expressed in music. But to understand *what* is expressed requires that we understand *how* it is expressed; the *how* often delineates or identifies the *what*. Examples might include

THE PSYCHOLOGY OF LEARNING
AND MOTIVATION VOL. 46
DOI: 10.1016/S0079-7421(06)46007-2

remembering a familiar tune by its tempo or recognizing the end of a phase by marked decrease in tempo and loudness.

The ways in which musicians manipulate acoustic signals to create expression bear considerable resemblance to the ways in which talkers manipulate speech. In speech, the acoustic changes in frequency, amplitude, and duration that form grouping, prominence, and intonation are called "prosody"; these features are systematic within each language and are thought to be rule governed and distinct from other structural levels of linguistic analysis (Beckman, 1996; Pierrehumbert, 1999). We refer here to musical expression as "musical prosody" because, as in speech, performers manipulate music for certain expressive and coordinating functions. Acoustic properties of speech and music can be manipulated to a certain extent without changing the categorical information (the words as they might be written or the musical pitches as they might be notated). Because linguistic and musical items are constrained by different acoustic properties (e.g., timing defines duration categories in music but not in speech), some acoustic dimensions may play a larger role in one domain than in another. Thus, prosody reflects acoustic variations on dimensions within each domain, while maintaining important categorical distinctions. We focus here on the forms and functions of musical prosody and whether it is rule-governed and distinct from other forms of musical structure.

Not much attention has been paid to the role of musical prosody. One reason for this is an assumption that most aspects of memory for musical pitch or duration are relative (not absolute) and that subcategorical expressive features (such as frequency differences smaller than musical pitch categories) cannot be retained in memory once the items are perceived (Raffman, 1993). This position is similar to normalization views of speech perception in which extralinguistic speaker variation was treated as noise to be filtered out (Pisoni, 1997). Counterevidence to this position toward performer variation in music, however, is gaining to suggest that listeners do retain absolute details about pitch height (Levitin, 1994; Schellenberg & Trehub, 2003) and tempo of familiar music (Levitin & Cook, 1996), as well as subcategorical note-to-note variations in musical expression (Palmer, Jungers, & Jusczyk, 2001). Infants can remember the absolute details of tone intensities and durations as well (Palmer et al., 2001; Trainor, Wu, & Tsang, 2004). Thus, music is a good example of an intricate acoustic nonverbal system in which expression is remembered.

The role that prosodic features of music play in acquisition is particularly of interest; prosody may provide information to help novices (infants as well as late acquisition learners) parse a continuous acoustic stream into meaningful units and bootstrap learning of complex hierarchical relationships among

those units. Listeners' knowledge of complex tonal and rhythmic structure in music is acquired through implicit knowledge gained from long-term exposure to music of a particular culture (Bharucha, 1987; Krumhansl, 1990; Trehub, 2000); it is unknown how much of that acquisition is influenced by prosodic features. Later in this chapter, we consider how prosodic features of music influence segmentation and implicit learning of musical structure.

A. Musical Prosody is Obligatory

Musicians perform in a variety of ways in Western cultures. In many performances, musicians express their interpretation of previously composed music; the composition is learned either by reading a notated musical score or by hearing it performed. In improvisatory performance, musicians create the composition as they perform or create a variation on something they have heard or read. A musical composition denotes categorical pitches and durations. However, the physical instantiation of intensity, tempo (rate), articulation (relations of tone offsets to successive tone onsets, such as in staccato/legato), and timbre are usually underdetermined by the composition; in most cases, they must be determined by the performer. Because some instantiation of these factors is necessary, we consider prosody to be obligatory (not that there is an obligatory way in which prosody must be instantiated). Although frequency and timbre are fixed on piano and other keyboard instruments, most musical instruments permit variation in frequency and timbre, as well as variation in intensity and tempo. Some forms of expression are notated in a musical score; for example, intensity, tempo, and even emotional connotations can be indicated by composers or by editors, although these markings tend to be in a small number of large categories. However, the majority of acoustic features that the musician must choose are obligatory, in the same way that they are obligatory in speech (Pierrehumbert, 1999).

There are several ramifications of an obligatory prosody—first, every musical performance represents some choice of the physical variables that give rise to perception of stress, accent, rhythm, and intensity contour (including crescendos and decrescendos). Even computer-generated music, in which all tones have fixed and equivalent durations, intensities, and tempo, requires some choice of tempo and absolute intensity level by the composer/programmer. Those choices have important perceptual outcomes; computer-generated performances are perceived as flat or dull (Kendall & Carterette, 1990; Palmer, 1996b), similar to how computer-generated speech can be perceived as mechanical. Second, a musician tends to reproduce performances of the same music with the same prosodic choices. An example of a pianist's rendition of a Mozart piano sonata (Palmer, 1996a) is shown in Fig. 1; the prosodic choices of timing and intensity for each tone were highly similar across the repeated section of the performance (measures 1–16). Finally, musicians' prosodic choices change systematically with the musical context in which the tones appear. Excerpts from musical compositions are produced with different expressive nuances when performed in or out of a particular melodic context; for example, phrase-final lengthening (the choice of making final tones in a sequence longer than others) will be greater for the last tones in a musical excerpt when it is performed on its own than when it is embedded in a longer sequence. Also, musical excerpts are produced with different nuances when they are placed in one melodic context than in another (Palmer et al., 2001). One ramification is that experiments that use musical events produced in isolation, such as chords or tones, will reflect different prosodic structure

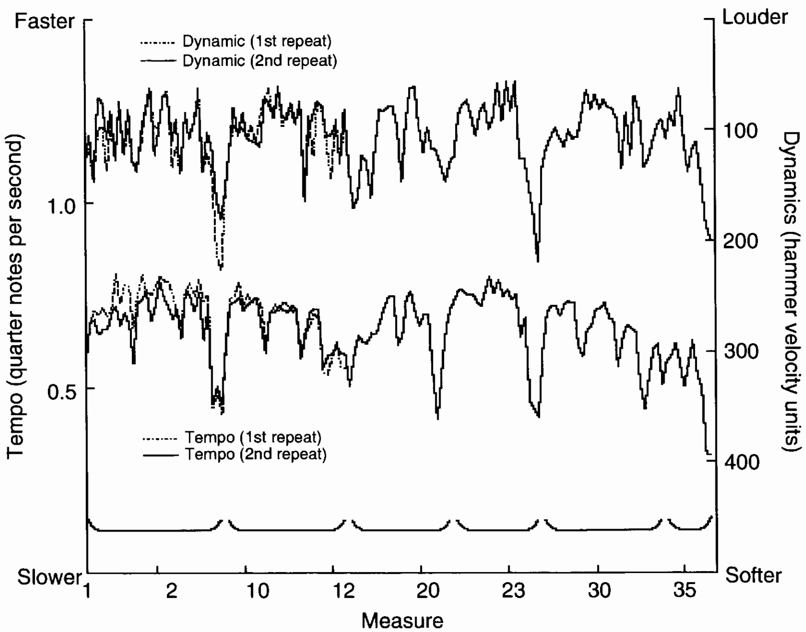

Fig. 1. Reprinted with permission from C. Palmer (1996a). Anatomy of a performance: Sources of musical expression. *Music Perception, 13*, 433–454, © by the Regents of the University of California. Performance of 54-measure section of a Mozart piano sonata; dynamics and tempo measures for the 16-measure section that the pianist performed twice. Both dynamics and tempo are consistent across repetitions.

from what is produced for the same musical events in context. Similar points have been made for speech (Pierrehumbert, 1999).

B. Musical Prosody Alters with Intent: Interpretations

One of the earliest empirical findings in music cognition was that musicians never perform events evenly (with equal duration, intensity, and articulation—note offset to successive note onset) (Seashore, 1938a). Performers often reproduce a musical performance with the same expression (Seashore, 1937), which indicates that the variations in how events are produced are not random. Furthermore, attempts to play mechanically (without expression) reduce the expressive nuances, and attempts to play exaggeratedly enhance the expressive nuances (Gabrielsson, 1987; Palmer, 1989a; Seashore, 1938b). An example of a pianist's intent to perform with a particular phrasing interpretation is shown in Fig. 2 (Palmer, 1989b). The difference between the expressive performance (solid line) and the intent to perform without expression (the dashed line) indicates the degree to which prosodic features of slowing down at phrase

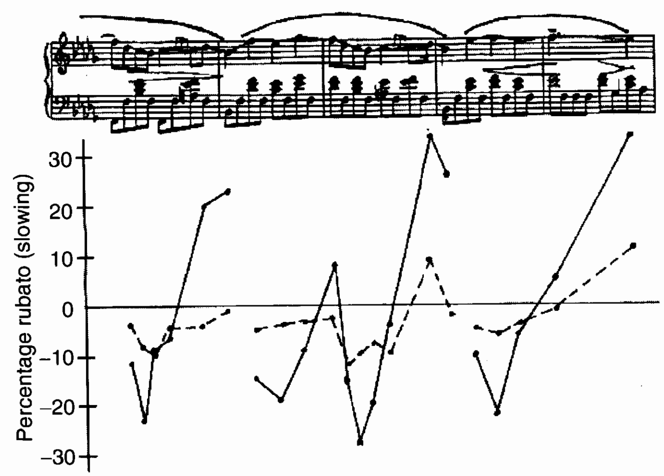

Fig. 2. Reprinted with permission from C. Palmer (1989b). Structural representations of music performance. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society* (pp. 349–356). © Erlbaum. Tempo changes that alter with performer's interpretation from five measures of performance of a Prelude by Chopin. Solid lines above music notation indicate performer's intended phrasing. Tempo changes in expressive performance (solid line) show phrase-final lengthening near intended phrase boundaries. Inexpressive performance (dashed line) shows reduced phrase-final lengthening.

endings alter with intent. Thus, expressive nuances are systematic and at least some nuances correspond to musical intentions.

Most music contains some structural ambiguity with regard to where the phrase boundaries or metrical units will be, how important one musical voice is relative to another, etc. Musical interpretations allow the performer a certain degree of flexibility in how they apply prosodic features. Interpretations reflect a conductor's or performer's modeling of a piece according to their own musical thoughts or ideas (Apel, 1972). Music teachers often discuss the contents of interpretations with students, based on musicological and historical elements, aesthetic elements, and elements of expressive performance that are not necessarily conveyed by a printed score. Although interpretations permit a certain degree of freedom across musicians and performances, many theorists believe that some interpretations are more preferred or appropriate than others (Cyr, 1992; Neumann, 1982). For example, polyphonic music of J. S. Bach's era does not permit the same freedom of tempo change as romantic music of Chopin's era; measurements of rubato are greatest in performances of romantic music across a range of performers and compositions (Palmer & van de Sande, 1995; Repp, 1992a; Todd, 1985). Some expressive nuances, such as phrase-final lengthening or progressive slowing of tempo toward the end of major phrasal units, correspond to musical grouping structure as determined by the composition (Todd, 1985, 1992). Other expressive nuances seem tied to performers' individualistic decisions about the prominence given to some musical events over others, such as which tones form major phrase boundaries (Palmer, 1989a; Shaffer, 1992) or which tones in chords (simultaneities), are most important (i.e., the melody) (Palmer, 1989a, 1996b).

Given that most music contains structural ambiguity, one can ask whether musical prosody is the result of individualistic interpretations. Do performers have the freedom to express the musical features in acoustic signals in a primarily individualistic manner, or is this mapping constrained by cognitive principles shared across performers and listeners? If the former is true, then we should not expect to see perceptual or cognitive generalities across musical styles or cultures; other factors, such as historical period, geographical region, and instrument-specific degrees of freedom, may shape musical expression. Although there are many studies of musical expression within Western tonal cultures, there is a paucity of evidence to address cultural differences in musical expression. If the latter is true, we should see general principles of musical prosody both in the forms that it takes and the functions that it serves. This chapter focuses on whether the term "musical prosody" captures general principles of musical expression above and beyond individualistic features of musical interpretation.

II. Definitions of Prosody

To define musical prosody, we first consider definitions of linguistic prosody. Classical study of linguistic prosody, interestingly, returns us to music. In traditional study of classical grammar, which focused on poetry prosody was defined as the study of accent, or phonetic properties of syllables and words relevant to the measure of rhythm and meter, especially in verse (La Driere, 1993). Meter, which derives from the Greek *metron*, refers in both verse and in music to metrical units defined by accents, measured in syllables or feet in poetry and in beats in music (*American Heritage Dictionary of the English Language*, 2004). Grammatical and rhetorical elements were often combined with musical elements in the late Middle Ages, when poetry was again closely associated with music (La Driere, 1993). Thus, it is not surprising that aspects of metrical phonology have developed similarly in linguistic and music analyses.

Today, prosody is considered a distinct component of phonological theory, which posits a separate prosodic tier for metrical structure (Levelt, 1989; Liberman & Prince, 1977; Selkirk, 1984). Prosody can refer both to an abstract level of phonological structure and to its acoustic realization in speech (Cutler, Dahan, & Donselaar, 1997; Shattuck-Hufnagel & Turk, 1996). Prosodic cues in speech include acoustic variations in fundamental frequency, spectral information, amplitude, and relative durations of speech. One approach from phonological theory characterizes prosody in terms of the cognitive structures that are implicit in the minds of speakers in order to explain their use of these variations in speech production and speech perception (Pierrehumbert, 1999). Prosodic structure within the word is important in lexical access, influencing how segmentation works in each language and the set of active competitors for any given word at any given time (Cutler, 1995). Prosodic structure above the word level (phrasing and phrasal prominence) often reflects other dimensions of linguistic analysis, including syntax, semantics, and discourse structure (Ferreira, 2002).

Most definitions of speech prosody distinguish between pitch and time dimensions. For example, prosodic structure of rhythm, grouping, and prominence is often separated from intonational (melodic) structure (Cutler, 1999; Ferreira, 2002; Pierrehumbert, 1999). Intonational structure refers to pitch at the phonological phrase level, often described as the "melody" of the speech. Intonation is carried in the fundamental frequency of the voice, perceived as pitch. The primary functions attributed to intonation in non-tone languages (stress-timed or syllable-timed languages) are to mark the pragmatic information or the emotional information in an utterance (Pell, 2001). In lexical tone languages, such as Mandarin or Thai, lexical features interact with intonational features to determine pitch contours assigned to phrases. Intonation and rhythmic prominence (including meter) place simultaneous constraints on each other in prosody (Beckman, 1996).

In contrast to speech, musical pitch is relatively fixed by categorical constraints that are specified in a composition. Although most musical instruments allow some manipulation of fundamental frequency, performance conventions constrain the amount of pitch variability that can be introduced (vibrato, or a small fluctuation in pitch on a single tone, is an exception). While soloists may slightly alter their intonation at times, this is generally not feasible for ensemble musicians who must perform simultaneously with others, and is not possible for performance on instruments, such as keyboards, which only allow discrete pitch changes. Thus, most of our discussion of musical prosody focuses on rhythm, grouping, and prominence, and not intonation.

The temporal aspects of music and speech prosody bear important similarities, in particular, between musical meter and linguistic stress (Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983; Selkirk, 1984). The rhythm of speech involves patterning of strong beats that coincide with some syllables. Syllable-timed languages like Spanish are described as having a regular beat pattern in which each syllable coincides approximately with a beat, whereas stress-timed languages like English tend to have beats on stressed syllables and equal time intervals between stressed syllables (Pike, 1945). Vowels in stressed syllables are longer than in unstressed syllables in many (but not all) languages (Lehiste, 1972). Similar to how words differ in the amount of stress that is assigned to their constituent syllables, musical tones differ in the amount of prominence or accent assigned to them. Musical accent, like stress, is hierarchical—usually there are two to four levels of accent in a musical phrase (Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983). Metrical grids are a common formalism for representing both the different levels of stress that syllables receive and the levels of accent that musical tones receive (Lerdahl & Jackendoff, 1983; Liberman & Prince, 1977). Syllables that are stressed tend to be longer or bear a pitch accent (Ferreira, 2002), and musical tones that bear metrical accent tend to be longer or produced with greater intensity (Drake & Palmer, 1993; Sloboda, 1983).

A major difference in the temporal aspects of music and speech prosody is the degree of isochrony, or temporal regularity. Music of many cultures displays a strong beat regularity as evidenced in written notation and in aural traditions; furthermore, musically untrained listeners can clap along systematically to unfamiliar music that is performed with expressive tempo variations (Drake, Penel, & Bigand, 2000). Languages of the world do not display the same degree of regularity in production or perception. Interstress intervals tend to vary as a function of the material within the interval, although evidence of a tendency toward isochrony is seen in some studies when principle determinants of duration (sentential context, lexical content, etc.) are controlled for (Cooper & Eady, 1986; Kelly & Bock, 1988). Differences between speech and music in their degree of temporal regularity raise the possibility that prosodic variation may be constrained differently. Prosodic variation in music, whose temporal regularity is high, may permit more deviation and still yield perceptual constancy, whereas prosodic variation in speech, whose temporal regularity is lower, may be pressured to align with other structural (syntactic, semantic) dimensions.

In sum, although music and language differ on many important structural dimensions, both domains have realizations of prosody. Music may be more constrained in terms of pitch and duration categories, but both domains utilize the flexibility available within the domain. Temporal and pitch manipulations are used prosodically in music, although not to the extent permitted in speech. Linguistic stress and musical accent are both realized in duration and intensity, and are manipulated in similar ways for prosodic purposes. Linguistic meter and musical meter are modeled with similar hierarchical rules, and producers appear to manipulate those dimensions in similar ways to serve prosodic functions. The precise functions are discussed in later section.

III. Functions of Musical Prosody

What functions would musical prosody serve? Linguistic prosody serves several functions, some of which do not have direct equivalents in music performance. One such function is to signal the illocutionary intent of the speaker (such as making a statement or a request in "She has left the room (?)"). Another is to disambiguate the meaning of words with similar segmental structure (phonemic stress), such as "greenhouse" versus "green house." But other prosodic functions that relate less to semantic content may be similar to the functions of music prosody, such as segmenting a continuous acoustic stream into its component units, highlighting items of relative importance (focus and prominence), coordination among producers (including turn-taking), and attributing emotional states to producers. We consider here these functions that musical prosody may play.

A. Segmentation

A major focus of auditory communication is how people segment a continuous signal into important events, their sources, and the relations among them (Bregman, 1990). Music and speech are organized in units of varying temporal extent. A short segment can serve as a unit at one level and then join with other segments to form longer units at higher levels of organization, in a hierarchical structure (Deutsch & Feroe, 1981; Lerdahl & Jackendoff, 1983). Musical tones combine to form melodic and rhythmic figures, phrases, and larger sections. Listeners, with and without musical training, are adept at segmenting the musical stream into these units. How do prosodic cues influence this segmentation?

One musical unit that is often marked for segmentation by prosodic cues is the phrase. A musical phrase varies in length and is described as a unit of meaning, often defined by elements at its boundaries (Cogan & Escot, 1976). Performers often mark phrase boundaries with changes in intensity, tone duration, and articulation (Henderson, 1936). Performers on a wide variety of musical instruments tend to use phrase-final lengthening at boundaries by increasingly lengthening successive tones as they approach a structural boundary, even when those tones are notated as equivalent duration in the musical composition (Gabrielsson, 1987; Kendall & Carterette, 1990; Palmer, 1989a,b; Todd, 1985). The patterns of tempo modulations often indicate a hierarchy of phrases, with the amount of slowing at a boundary corresponding to the depth of the phrase embedding (Shaffer & Todd, 1987; Todd, 1985, 1989). Todd (1985, 1989) proposed a computational model of phrase-final lengthening in music performance; the more important the musical segment in a given phrase, based on a hierarchical analysis of meter and grouping principles (Lerdahl & Jackendoff, 1983), the greater the phrase-final lengthening. The correspondence between phrase-final lengthening and intensity changes in music performance is also strong; unit boundaries are marked by both slowing in tempo and decreased amplitude (Drake & Palmer, 1993; Palmer, 1996b).

Are musical phrases always marked prosodically? Usually, yes. All published research on performance of Western tonal music shows phrase-final lengthening at boundaries. Compositional structure marks phrase boundaries as well with longer duration categories, and listeners attribute phrase structure in the absence of prosodic cues (Palmer & Krumhansl, 1987). Listeners are sensitive to phrase markings as early as infancy. When pauses were placed at appropriate phrase boundaries, infants displayed longer orientation times in a head-turn preference procedure task than to music with phrase boundaries placed at inappropriate phrase boundaries (based on the structural information in the compositional score) (Krumhansl & Jusczyk, 1990). Jusczyk and Krumhansl (1993) further demonstrated that the decreasing pitch height and increasing tone duration that typically marked phrase boundaries were critical for infants' sensitivity to music that is segmented at phrase boundaries. These two variables are the same ones suggested to underlie infants' segmentation of speech (Jusczyk et al., 1992)—change in fundamental frequency at the ends of clauses and phrases, and lengthening of syllables before important syntactic boundaries. These similarities raise the possibility that listeners respond to general acoustic properties of acoustic signals, which serve to mark important events in auditory perception, rather than duplicating domain-specific segmentation mechanisms.

Other prosodic features of music help to distinguish simultaneities, such as musical voices from each other, presumably to aid listeners in focusing limited attentional resources. Performers often manipulate the relationship between the timing of individual parts that are notated as simultaneous, with intensity differences or onset timing differences; onset asynchronies are a primary aid to stream segregation (Bregman, 1990). Musical ensembles display inter-instrument onset asynchronies among tones notated as simultaneous (chords) that can help listeners distinguish between voices in multi-voiced music (Rasch, 1978). Ensemble players who produced the melody (voice of primary importance) tended to precede other voices in string and wind trios by 30–50 ms (Rasch, 1979, 1988). Palmer (1989a, 1996b; see also Thompson and Cuddy, 1997) showed further evidence of melodic leading in solo piano performance, giving evidence that timing asynchronies are the result of structural emphasis rather than solely a result of coordination between performers. Goebl (2001) showed that the amount of the melody leads correlated with hammer velocity differences (how fast the keys were struck to create louder sounds) on acoustic piano.

Do prosodic features influence music segmentation? Palmer et al. (2001) reported performances of the same musical excerpt (a musical measure, or metrical unit) placed in different melodic contexts. Figure 3 shows the original prosodic cues produced by a pianist for the excerpt (marked by the box) and the melody in which the excerpt was presented in music notation. Prosodic features of intensity and articulation were used to mark the metrical beats as implied by the meter in which the excerpt was presented. The prosodically marked excerpts were then spliced and placed in computer-generated melodic contexts whose musical structure either matched or mismatched the implicit structure of the prosodic cues. Musically trained and untrained listeners were then familiarized with a performance of the excerpts out of context, and later had to recognize which excerpts they had heard when the excerpts were placed in the computer-generated melodic contexts. All listeners were less able to recognize the excerpt when its prosodic cues matched the context; when the excerpt was heard in the incorrect context, it became more salient, suggesting that the prosodic cues influenced listeners' segmentation processes. In sum, prosodic cues may not always be necessary for segmentation in music, but when present, prosodic features have a significant influence on how segmentation occurs.

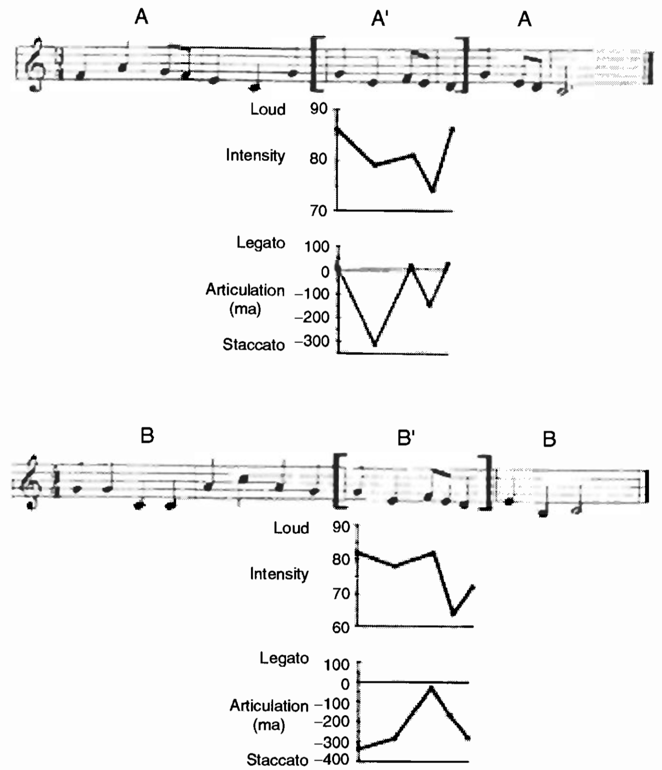

Fig. 3. Reprinted with permission from Palmer et al. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45*, 526–545, © Elsevier. Musical expression changes with context and aids segmentation. Pianist's performance of musical excerpt placed in different metrical contexts (top: ternary meter, bottom: binary meter). Intensity and articulation patterns (legato/staccato) increase on metrical accents.

Does musical prosody carry information in the absence of tonal or rhythmic content? For example, listeners can still analyze major prosodic phrasing when segmental content is removed in low-pass-filtered speech or in hummed speech (Kreiman, 1982, with low-pass-filtered speech; Collier & Hart, 1975, with hummed sentences). Only a few studies have compared listeners' responses to musical prosody in the absence of pitch (monotonic tone) or rhythmic content (isochronous rhythm). When pitch variation was removed and only durational performance variation remained, listeners were able to identify the phrase structure on the basis of duration-lengthening (Palmer & Krumhansl, 1987). Likewise, when durational variation was removed and

pitch variation remained, they were able to indicate phrase structure on basis of pitch variation. Thus, these results suggest that musical prosody is sufficient to signal phrase structure, similar to the speech findings.

B. Prominence

Linguistic prominence, or acoustic highlighting with stress, accent, and other prosodic forms, is used to signal the focus, or the more significant elements of a sentence (Dogil, 2003). Musical prosody may also function to signal the relative prominence of events. Prominence is especially important in ambiguous forms of music for which the meter, phrasing, or the melody (most important voice) might be underdetermined by the composition. Performers use prosodic cues to emphasize the metrical and phrase structure of the music (above and beyond the segmental marking of boundaries), especially when the structure is ambiguous. Performers lengthen tones, perform them louder, or change the articulation (more legato or more staccato) to mark metrically important events (Sloboda, 1983, 1985). Gabrielsson (1974) documented in a wide variety of rhythms that musicians performed metrically important beats louder, longer, and more legato relative to other beats. Many studies document a decreased intensity and/or slowing tempo that performers introduce at phrase boundaries (Palmer, 1989b; Repp, 1992a; Shaffer, 1981; Todd, 1985), similar to phrase-final lengthening in speech. Vibrato, or small, rapid variations in pitch, is another method for increasing prominence. Although less research has addressed the use of vibrato, some studies indicate that singers and string players use vibrato to emphasize musical events (Seashore, 1938; Small, 1937).

Prominence can be used to exaggerate some aspects of musical structure at the expense of others, similar to the goals of contrastive focus in speech. For example, duration patterns that are intended to have certain ratios, such as 3:1, are often performed with higher ratios (>3:1), perhaps to distinguish them from similar (2:1) ratios commonly found in music (Gabrielsson, 1987). Shackford (1961, 1962) presented similar evidence from the production of simultaneous musical intervals; when two violinists performed a single tone a tritone apart from each other (a pitch interval of six semitones between the two tones), the interval they produced was exaggerated as a larger pitch interval when it was notated as an augmented fourth (an interval created by expanding a perfect fourth—an interval of five semitones) than when it was notated as a diminished fifth (an interval created by reducing a perfect fifth—an interval of seven semitones), even though the augmented fourth and the diminished fifth denote equivalent pitch intervals of six semitones. These exaggerations give prominence to a desired interpretation over an alternative interpretation, similar to how accented vowels tend to be articulated in a more exaggerated fashion and made more distinct from each other than unaccented vowels (Beckman, 1996; de Jong, 1995).

Performers and listeners develop expectations for the relationship between musical structure and prosodic prominence. Pianists are better at imitating the note-to-note prosodic variations in performances whose expressive timing deviations match the implicit phrase structure (Clarke, 1993; Clarke & Baker-Short, 1987); they are not as accurate at imitating expressive timing that does not match the structural expectations they form (Repp, 2000). Listeners develop structural expectations that influence their perception of prosodic cues as well; for example, a tone with increased duration is harder for listeners to detect when it occurs at a position in which lengthening is expected (toward the end of a phrase) relative to an unexpected position (in the middle of a phrase) (Repp, 1992c). Listeners' ability to detect lengthened tones placed in a computer-generated performance (one in which all tone durations, intensities, and tempo are constant) was inversely correlated with the points in time at which performers typically lengthen tones. Thus, structural expectations by performers and listeners can influence production and perception of prosody in systematic ways.

Interpretive (performer-specific) prosody also signals the relative prominence of musical events. Kendall and Carterette (1990) and Palmer (1989a) present evidence that different performers playing the same musical composition apply different prosodic cues to emphasize particular structural intentions. Palmer (1989a) documented correspondences between performers' notated interpretations of phrase boundaries and melody (primary voice) with melodic asynchrony, rubato, and articulation (staccato/legato) in pianists' performances of the same piece. Other studies document aspects of performance that are common between different performances of the same music (Gabrielsson, 1999; Repp, 1992b). In sum, musical prosody marks prominent musical events that are determined both by the composition and the performer.

C. Coordination

An important function of prosody in speech is to regulate turn-taking and interspeaker coordination in discourse, signaled primarily through intonation patterns and pauses (Cutler et al., 1997; Swerts & Geluykens, 1994). Most forms of music performance are collaborative; while they are not strictly turn-based, they require coordination between performers to achieve simultaneities. Sometimes, this is externally regulated by a conductor; however, in many small ensembles, performers must maintain coordination among themselves. Most research on synchronization among performers has focused on segmentation and prominence issues (described in an earlier section) and less on coordination. Performers must keep their onsets within

30–50 ms (Rasch, 1979) of each other for simultaneities to maintain cohesion. A conductor usually accomplishes this purpose in large ensembles by visual cues to the beat; more research is needed to investigate how performers maintain cohesion in ensembles without a conductor. It may be that one member is designated the "conductor" (perhaps the member who tends to lead in asynchronous performance), and other members follow their visual or auditory cues (Rasch, 1979). Synchronization may also be accomplished from memory or from schematic knowledge of stylistic norms.

Some researches address turn-taking among musicians. Perhaps the most important factor for regulating the coordination required for turn-taking is rate or tempo; a performance is considered a failure if performers do not maintain the same tempo. Jungers, Palmer, and Speer (2002) demonstrated that performers show strong tempo persistence from music they heard previously. Musicians and speakers heard sequences produced at different rates, and then produced sequences presented in written form. Musicians mimicked the rate of what they had just heard in their subsequent performances. This finding of rate priming can ensure successfully coordinated performances. Speakers' rates were also primed by the speech rates they had just heard, although to a lesser degree; their speech rate was better predicted by their preferred rate. Speech may show less rate priming because of the less strict demand for tempo coordination in spoken discourse than in music performance, which often requires that musicians perform simultaneously. However, other factors may ensure turn-taking in speech than in music; Jungers et al. (2002) found that speakers' intonational structure mimicked that of the speech they just heard; this parallel was not possible in the music because the primes were produced with equivalent loudness contours, and the notated music constrained the pitch values that were to be produced.

In some musical styles, such as improvisatory jazz, performers alternate creating solos in a stylistic turn-taking that is more like spoken discourse. Like conversation in which speakers elaborate upon their discourse partners' linguistic phrases (Fais, 1994), improvisatory jazz soloists elaborate upon previously heard performers' musical ideas (Johnson-Laird, 1991; Pressing, 1988). Although there is not much research on musical discourse, musical improvisation (in which the content is not predetermined) is an area, which would be an appropriate focus of questions such as: "Which prosodic cues distinguish the end of a musical solo from its middle?" and how much of turn-taking is prearranged or relies on learned musical patterns?

D. Emotional Response

Music has been characterized as the language of emotions (Peretz, Gagnon, & Bouchard, 1998), and cognitive approaches to music understanding have focused on the emotions that can be conveyed through music (Meyer, 1956). Until recently, little research focused on music's emotional content, both because of difficulties in defining musical emotion and possible lack of consensus in listeners' emotional response to music (Sloboda, 1992). There are distinctions between musical communication of emotional states and speech communication of emotional states. Whereas listeners often attribute emotional states to speakers based on their prosody (at least in everyday speech), listeners do not normally attribute the emotional state to the performer or assume that the sad music itself is suffering (Davies, 2001). Developments suggest there are some generalities in listeners' response to music that extends to infants and across cultures (Balkwill & Thompson, 1999), although the field of cross-cultural musical response is young. We do not know yet whether universals exist in emotional response to musical styles across cultures.

Study of emotional functions in music has paralleled speech research in separating emotional prosody from linguistic (structural) prosody. The term emotional prosody in speech refers to melodic and rhythmic components that listeners use to gain insight into a speaker's emotive disposition (Scherer, Ladd, & Silverman, 1984). The emotive significance of speech prosody tends to be considered as independent from linguistic-propositional content of prosody (Gandour, Wong, & Hutchins, 1998), in part due to empirical findings that speakers' emotional states were signaled in the fundamental frequency independent of the verbal content (Protopapas & Lieberman, 1997; Scherer et al., 1984). Neurological findings also suggest some dissociation between emotional and linguistic prosody (Heilman, Bowers, Speedie, & Coslett, 1984; Ross, Thompson, & Yenkosky, 1997). However, comparisons indicate that either right- or left-hemisphere damage can produce emotional prosody deficits of similar magnitude (Pell & Baum, 1997). A few studies suggest neurological dissociation between emotional response and structural response to music. Peretz, Belleville, & Fontaine (1997) tested a patient with brain damage whose musical abilities were impaired while her speech and intellectual functions were spared, a case of amusia without aphasia. The patient could classify happy and sad tunes correctly (emotional content), but she could not recognize familiar tunes or discriminate gross changes in pitch contour (Peretz, 2001; Peretz & Gagnon, 1999; Peretz et al., 1998).

Emotional expression in music performance is typically measured in experiments in which performers are instructed to express different emotions for the same piece of music in different performances. The performances are evaluated in listening experiments to see whether listeners recognize the intended expression (usually in a forced choice procedure), and the prosodic

parameters of the performances are analyzed with regard to the intended emotion. The assumption is that because the structural contents of the music remain the same, any changes in performances or listeners' reactions must be due to the intended emotional expression. The focus of this research has been on five primary emotions (Ekman, 1973)—happy, sad, angry, fearful, and surprised.

The most successful emotions communicated in the prosodic features of music performance are happy and sad (Gabrielsson & Juslin, 1996; Krumhansl, 1997). Juslin (2001) conducted a meta-analysis of which expressive cues performers used to communicate the five basic emotions across studies. The primary means of expression included tempo, sound level, timing, articulation, vibrato, tone attacks, tone decays, and pauses. Sad emotions were associated with slow tempo, low-sound level, legato articulation, whereas happy expressions were associated with fast tempo, high-sound level, and staccato articulation. Children used the same prosodic cues to express emotions in song—by 4-year olds, they used a fast tempo and high intensity in happy expression, and a low tempo and low intensity in sad expression (Adachi & Trehub, 1998).

Although some studies document listeners' success in choosing the emotion intended in professional performances, they also document individual differences (Gabrielsson & Juslin, 1996; Juslin & Madison, 1999; Kotlyar & Morozov, 1976). Music's emotional connotations are dependent on many factors, including the musical instrument, the particular musical structure, and the particular performance (Gabrielsson & Lindstrom, 2001; Juslin, 2001). Furthermore, different acoustic parameters of music performance interact in emotional judgments. For example, anger judgments, which are influenced by the sharpness of attacks or tone onsets, were expressed well on an electric guitar but not on a flute (Gabrielsson & Juslin, 1996).

There are several models of emotional expression in music, based on Hevner's (1935a,b, 1936, 1937) groundbreaking studies of emotional response to music. This work first documented that music played at a fast tempo tends to be perceived as exciting and happy, whereas music played at a slow tempo is serene and dreamy. Some of these musical features, such as pitch height, are specified by the composition and some, such as tempo, by the performance prosody. Juslin and Laukka (2003) model the expression of emotional states following a functional lens model (Brunswik, 1956) that begins with the encoding of emotional states by certain voice and speech characteristics in the signal. The emotional arousal of the speaker is accompanied by physiological changes that affect respiration, phonation, and articulation in such a way as to produce emotion-specific patterns of acoustic parameters that are understood by listeners (Scherer, 1986; for a detailed description in speech). According to this view, vocal music is physiologically related to vocal expressions of emotion; emotion influences the production of speech and vocal music in the same way (instrumental music is included by association) (Juslin & Laukka, 2003).

Is musical prosody necessary to convey emotion? Peretz et al. (1998) played listeners commercial recordings of music that contained normal expressive features and a computer-generated (prosody-less) version of the same music. Musical pieces that were originally interpreted by the experimenters as happy or sad were classified correctly, even in the absence of prosodic cues. Thus, some information as to emotional content is conveyed through structural features not considered part of performance expression. A second study manipulated the mode (minor or major key signature) and tempo (fast or slow) of the music in versions of the same musical composition. Listeners' emotion judgments were influenced by both tempo (faster tempi evoked happiness judgments) and mode (minor keys evoked sadness judgments). Thus, prosodic features (tempo) conveyed emotion in addition to structural features (mode). Interestingly, listeners were able to discriminate happy/sad musical segments, based primarily on tempo cues, for excerpts as short as 500 ms. This finding is consistent with a view of linguistic prosody as immediate and occurring early in processing (Dogil, 2003), compared with syntactic analysis (such as mode), in which structural cues may occur over larger time-spans (Peretz et al., 1998).

Listener-specific factors, such as musical training or familiarity with the particular music, also influence emotional response (Peretz, Gaudreau, & Bonnel, 1998). Scherer and Zentner (2001) propose that listeners' emotional response to music arises from an interaction of prosodic features (which they label performance features), structural features (such as phrase structure), listener-specific features (such as familiarity), and contextual features (such as location and event of hearing). Thus, this work suggests that prosody may be sufficient, as evidenced in the communication of different emotions for the same musical structure (Juslin, 2001) but not necessary to convey musical emotions.

In sum, musical prosody can serve various functions—segmenting a continuous acoustic stream into its component units, highlighting items of relative importance (prominence), coordination among producers (including turn-taking), and attributing emotional states to performances. Of those, lending prominence may be the single greatest function that prosody serves; without prosody, the ambiguity in musical structure could not easily be resolved. Although there is less research devoted to the role of prosody in coordination or in emotional states, research suggests that prosody is one of several variables that serve those functions.

IV. Rule-Based Models of Musical Prosody

Is the relationship between musical structure and prosody rule-governed? Can an "appropriate" musical prosody be derived from compositional information without appealing to performer variability or individual interpretation? Several rule-based models have been proposed for mapping music's compositional structure to performed expression (Battel, Bresin, De Poli, & Vidolin, 1994; Clynes, 1995; Sundberg, Askenfelt, & Frydén, 1983; Sundberg & Frydén, 1985; Sundberg, Friberg, & Frydén, 1989; Todd, 1985, 1995), each of which applies to particular types of musical structure. With a few exceptions (Todd, 1992), these systems are composed of rules that take the musical structure notated in a score as input, and generate prosodic manipulations of pitch, duration, and intensity as output. Because performers have flexibility about which prosodic manipulations they use and to what degree, these rules are not meant to be deterministic ones that yield the only correct prosodic structure; instead, they are preference rules that lead to a common or preferred prosodic structure. The distinction between preference rules and well-formed rules was first raised to explain different interpretations of ambiguous musical structure (Lerdahl & Jackendoff, 1983).

Sundberg et al. (1983, 1989; Sundberg & Frydén, 1985) were instrumental in first developing rule-based models with a technique of "analysis-by-synthesis," in which the rules were based on the musical intuitions of a trained performer. The rules can be grouped generally into two classes—those that enhance segmentation by articulating group boundaries or harmonically important events, and those that enhance the prominence of tones by exaggerating pitch or categorical (notated) duration differences. An example of a rule that enhances prominence is to further shorten note durations that are notated as short (such as an eighth-note), and to lengthen note durations that are notated as long (such as a half-note). A rule that marks segmentation is to insert micropauses (of 80 ms) between subphrases and to lengthen note durations at the end of phrases (Sundberg et al., 1983, 1989). These rules can be weighted, and so can be applied moderately or over-expressively. When multiple rules affect the same tone, they are applied additively.

Perceptual tests of the rules were conducted to compare listeners' preferences for the rule-based grammars. Listeners preferred the rule-based performances to the computer-generated performances in which the rules were not applied, and musically trained listeners showed greater sensitivity to the application of the rules than nonmusicians (Sundberg, Friberg, & Frydén, 1991). Breslin, De Poli, and Vidolin (1992) showed that artificial neural networks could learn some of the performance rules specified in Sundberg's

(1983, 1989) grammar. These networks produced performances comparable to those produced by the rule-based systems (Sundberg et al., 1983, 1989), and listeners preferred the network-performances over computer-generated performances that contained no prosodic cues.

Rule-based models of composer-appropriate musical expression have also been proposed as well. Clynes (1983) advocated the incorporation of individual composers' pulses into the musical microstructure—composer-specific differences in the duration and the amplitude of each group of tones within a musical pulse. These pulse manipulations, derived from the author's intuitions and musical experiences, were argued to yield more acceptable interpretations for that composer's music than interpretations with no such manipulations, random manipulations, or the pulses of other composers. Clynes (1995) showed evidence that these pulses were preferred by musically trained listeners, though there is some conflicting evidence from perceptual tests (Repp, 1990), and the composer-specific pulses do not seem to be supported by judgments from untrained listeners, as do Sundberg et al.'s (1983, 1989) rules.

Rule-based models of phrase-final lengthening in music performance have been proposed as well. Based on Lerdahl and Jackendoff's (1983) theory of grouping and meter in Western tonal music, Sundberg and Verillo (1980) proposed a simple model of the final ritard—the deceleration of performance tempo seen at the end of a musical performance. They proposed a linear decrease in tempo to the end of the piece. Todd (1985) predicted the amount of slowing at phrase boundaries from a structural analysis of the musical composition. This model assigned greater lengthening to tones at major structural breaks (higher hierarchical levels in Lerdahl and Jackendoff's (1983) analysis). This approach is similar to Grosjean, Grosjean, and Lane's (1979) model of interword pauses in spoken sentences; Grosjean et al. (1979) found that a large degree of variation in speech pauses was predictable from the hierarchical representation of the utterance's syntactic phrase structure.

Other models have tried to model musical prosody with kinematic laws of physical motion. Feldman, Epstein, and Richards (1992) modeled the timing characteristics of ensemble performances in terms of a cubic polynomial model that smoothly connected sections of constant tempi with those of changing tempi, minimizing abrupt changes in acceleration, a desirable property in physical acceleration. Todd (1992, 1995) proposed a general-purpose algorithm for the timing of phrase-final lengthening, based on kinematic principles of physical motion. This model proposed a linear deceleration in tempo across tones, and treated musical space as continuous (rather than as discrete tone onsets). Based on findings that tempo changes are often coupled with changes in loudness (Gabrielsson, 1987; Palmer, 1996a), so that tempo and intensity increase together, Todd (1992) proposed

a model that linked the two, so that intensity is proportional to the square of the number of musical events per unit time, and proposed that musical expression creates the perception of motion in listeners. Although kinematic models of music performance are not universally accepted (Desain & Honing, 1992), they formalize the relationship between music performance and principles of motion.

In sum, rule-governed models focus primarily on notated compositional scores as input and yield prosodic manipulations as output. They reflect general prosodic features related to structural features of the music that can aid listeners in terms of segmentation and prominence. However, it is not clear how well these models generalize; although they are not instrument- or timbre-specific, these potential factors have not been thoroughly investigated. Also, the models do not account for interpretive (performer-specific) variation. Kendall and Carterette (1990) and Palmer (1989a) argued against a rule-governed performance grammar generated solely from compositional structure, based on evidence that performers apply different prosodic cues systematically to the same composition that support different structural intentions. Individual interpretation appears to play a role, as well as compositional structure, and some of these sources of variance can be accounted for by rules that are conceived as preferences, rather than as deterministic outcomes (Lerdahl & Jackendoff, 1983). A deterministic application of these structure-based rules—without any interpretive or individualistic variation—would lead to only one of many appropriate realizations of musical prosody. Empirical measurements of individual performances are necessary to distinguish between structure-based prosodic cues that transcend performers and prosodic features that are specific to performers. A combination of rule-based and measurement-based approaches may be necessary to explicate both the form and function of musical prosody.

V. Acquisition of Musical Prosody

Musical prosody is always present in human performance and appears to have many functions for listeners. What role does prosody play in learning? Prosody may aid perceptual learning of primitive units. Prosodic features in speech can provide low-level cues to aid segmentation and learning of hierarchical relationships (Gleitman & Wanner, 1982; Hirsh-Pasek et al., 1987; Jusczyk & Kemler Nelson, 1996), including delineating word boundaries (Christophe, Doupoux, Bertoncini, & Mehler, 1994; Gout, Christophe, & Morgan, 2004) and marking syntactic relationships (Fisher & Tokura, 1996). Listeners' abilities to identify smaller musical units may likewise bootstrap their ability to perceive higher-order relationship among those units.

Prosodic cues may be especially important in a domain like music that contains structural and emotional ambiguity. In a seminal theory, Meyer (1956) proposed that listeners' expectations, based on the hierarchical structure of the music, are critical to their emotional response. Music creates expectations, which cause listeners to experience tension when unresolved and release upon their resolution; these expectations are based on how the music will continue, as opposed to extramusical ideas. The better the listener can grasp the hierarchical structure of a piece, the more precise the listener's expectations and the more emotion is conveyed. Prosodic cues may help a listener grasp the emotional content of the piece by clarifying the musical structure.

Infants' responses to speech and music offer explanations of why prosody is important. Infant-directed utterances across many cultures contain musical features that are often described as "melodies" (Fernald, 1989; Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990). According to Fernald (1992), infants are predisposed to selectively attend to the distinctive pitch contours of infant-directed speech, whose primitive emotional meanings can be decoded in the absence of language. Cross-cultural similarities in emotional response to adult speech support this argument (Frick, 1985; Krauss, Curran, & Ferleger, 1983). Intention and affect can be communicated in intonation; caretakers use salient intonational patterns that draw preverbal infants' attention (Fernald, 1985, 1993). Parents use melodic contours in consistent ways to engage babies—infant-directed speech is typically higher in pitch with more exaggerated intonation contours than adult-directed speech (Fernald et al., 1989). Different melodic contours (level, rise, fall, bell-shaped, U-shaped, etc.) can consistently map to different behaviors that parents attempt to elicit, including discouraging unfavorable behavior, encouraging imitation, encouraging play, and contingent rewarding (Papoušek, Papoušek, & Symmes, 1991). Furthermore, adults can successfully discriminate emotional states in both infant-directed and adult-directed speech, but the acoustic correlates of emotional expression are more widely found in infant-directed speech (Trainor, Austin, & Desjardins, 2000). This work suggests that intonation and prosody may aid parent–infant bonding and emotional communication (Trainor et al., 2000), although attempts to relate particular melodic contours to particular emotions have generally failed (Scherer, 1985).

Adults use prosodic features in music as well, to draw the attention of the infant. Caregivers' songs to infants tend to have higher pitch, slower tempo, and more jitter in fundamental frequency and intensity than songs directed at adults (Trainor, Clark, Huntley, & Adams, 1997; Trehub & Trainor, 1998; Trehub et al., 1997), which have been associated with increased emotional expression. These prosodic features appear in songs sung by young children to their infant siblings (Trehub, Unyk, & Henderson, 1994). Parents can adapt their song to the mood and abilities of their listeners. The same song may be sung in a playful way or a soothing way (Trainor & Rock, 1997; Trehub et al., 1997). Mothers can alter their performances of the same song for infants (at a higher pitch level) or for preschool children (in enunciated lyrics).

Prosodic cues in music or speech can draw infants' attention to statistical regularities in the transitional probabilities among sequence items. In a statistical learning paradigm, infants were familiarized with recurring sequences of syllables in brief exposure to an artificial language (Saffran, Aslin, & Newport, 1996). After familiarization, the infants were able to discriminate novel syllable combinations from familiar ones, indicating that they were sensitive to the transitional probabilities between phonemes in strings of speech sounds. Infant-directed speech that contained more prosodic differentiation in pitch height, pitch range, and pitch peaks elicited better statistical learning among infants than did adult-directed speech (Thiessen, Hill, & Saffran, 2005). In a similar experiment, Thiessen, Hill, Emerson, and Saffran (2005) showed that infants learned number sequences that were sung with a unique rhythmic/intonational contour more easily than spoken number sequences. These findings suggest that prosody can benefit statistical learning in music and speech, although several studies demonstrate that statistical learning occurs in the absence of additional prosodic cues (Saffran, Johnson, Aslin, & Newport, 1999; Saffran et al., 1996).

How musical prosody bootstraps learning is likely to be related to the alternatives to be learned. For example, different musical styles and genres permit different degrees of variation in how compositions can be manipulated by performers to make some elements more prominent. In addition, musical instruments differ in how many prosodic features they allow performers to alter; for example, violin permits vibrato but piano does not; most instruments permit change in pitch but drums do not. The musical environment of the infant will shape the space of possible learning alternatives. Because infants' musical environments contain infant-directed music, mostly from their parents' voices, it would not be surprising if the prosodic features of the singing voice are learned first and are then generalized to other nonvocal forms of music, as suggested earlier (Juslin & Laukka, 2003). In this view, it would not be surprising if musical prosody and speech prosody were related. Evolutionary theories of music's origins suggest that musical behavior evolved in conjunction with—or as an adaptation of—vocal communication (Brown, 2000; Dissanayake, 2000). An area for future research is whether prosodic marking of infant-directed song varies across musical styles and across languages. Also of interest is whether musical prosody can aid infants' learning of musical statistical regularities, as in language learning.

VI. How are Musical and Linguistic Prosody Related?

Researchers have sought more direct links between musical and speech prosody. Patel and Daniele (2003) found differences in the amount of durational variation used by French- and English-speaking composers. English, a stress-timed language that alternates in strong and weak stresses, typically shows larger variation in the production of interconsonantal intervals (Ramus, Nespor, & Mehler, 1999) and vowel duration (Grabe & Low, 2002) than French, a syllable-timed language. The amount of rhythmic variation in a corpus of 19th century music compositions, measured by note-to-note changes in notated durations, was smaller in a collection of French musical themes than in the English themes (later confirmed by Huron & Ollen, 2003, in a larger corpus). The rhythmic variability of the music was attributed to the influence of the composers' knowledge of their language on the compositions, even though the instrumental music contained no words. Although these finding are based on analysis of notated musical compositions (which contain no prosodic performance variation), the suggestion of different sensitivity to musical variation related to linguistic knowledge is intriguing.

Some neurological evidence suggests a direct connection between musical and linguistic prosody. Stroke victims who have impairments in musical discrimination and perception often have related impairments in perception of linguistic prosody (Nicholson et al., 2003; Patel, Peretz, Tramo, & Labreque, 1998). Patel et al. (1998) documented two stroke-related cases of amusia—one patient (CN) showing deficits in musical memory, the other (IR) showing perceptual deficits. Perceptual discrimination tasks were created based on sentence pairs, which differed only in linguistic prosody, and musical analogs of the prosodic stimuli were created (only fundamental frequency and duration information were retained from the linguistic stimuli). IR had difficulty on both the linguistic and musical prosody tasks, whereas CN (whose deficits involved long-term memory) performed similarly to control subjects, suggesting shared neural mechanisms between the two domains. Congenitally amusic individuals (with similar music perception deficits but no history of brain damage) did not show deficits in linguistic prosody (Ayotte, Peretz, & Hyde, 2002). The congenitally amusic individuals may have developed different neural mechanisms than the stroke-damage patients to process linguistic prosody.

Music training is associated with increased sensitivity to pitch processing in language tasks. Musicians detect fundamental frequency changes better than nonmusicians; they also show similar ERP responses to small frequency manipulations in music and speech, whereas nonmusicians show similar neural responses in music and speech only for large frequency changes (Schön, Magne, & Besson, 2004). Thompson, Schellenberg, and Husain

(2004) tested whether musical training influenced listeners' ability to detect the emotional connotations in speech. Musically trained and untrained listeners heard semantically neutral utterances spoken with a particular emotional prosody, or musical tone sequences that mimicked the utterances prosody, created in the same way as Patel et al.'s (1998) musical analogs to spoken sentences. Musically trained adults outperformed untrained adults at identifying sadness, fear, and emotionally neutral terms. The influence of music lessons on identifying emotions in speech prosody was extended to children as well; a training study with 6-year olds randomly assigned to 1 year of keyboard, vocal, drama, or no lessons indicated that the keyboard group performed equivalently to the drama group and better than the no-lessons group at identifying anger or fear (Thompson et al., 2004). These studies suggest that musical training facilitates the ability to decode emotional meaning in speech prosody (although motivational differences among groups may play a role as well).

Direct comparisons of musical prosody and speech prosody, both in terms of their functions and their neurological substrates, are beginning to yield areas of overlap. Musical and speech prosody manipulate the same acoustic variables, yield emotional expression, and serve some of the same segmentation and prominence functions. It seems likely that future studies will yield additional similarities in the functions of coordination and how infants acquire prosodic knowledge. Whether these similarities can distinguish a single evolutionary basis or homologous structures for musical and speech prosody is not obvious; perhaps this distinction is not as important as understanding whether the functions served by prosody within each domain imply abilities that transcend domains.

VII. Conclusions and Caveats

Musical prosody is a complex, rule-governed form of auditory stimulation, and it can move listeners emotionally in systematic ways. The acoustic cues in which musical prosody is instantiated are shared across performers, and performers share with listeners ideas of what constitutes appropriate prosody. Although musical prosody contains individual components of interpretation and prominence, it also reflects constraints based on shared cognitive principles of perceptual organization, emotional response, and even motor production. Thus, it is partially influenced by the structural ambiguity of music that individual interpretation addresses, but it is also determined by cognitive structures that arise from human perceptual and motor biases. Sensitivity to prosodic cues aids listeners in identifying commonalities and differences in sources of musical sound. In sum, this evidence suggests that musical "prosody" captures general principles of musical expression above and beyond individualistic features of musical interpretation.

One important caveat is that not all prosodic variability in music performance is intentional or expressive in nature. Movement demands on performers contribute to the acoustic characteristics of performance; for example, temporal variance in performed event durations has been modeled in terms of an internal timekeeper and motor response delays (Wing & Kristofferson, 1973). Shaffer (1981) applied similar modeling in analyses of timing in piano performance, concluding that the timing variability arose in part from timekeepers that controlled the timing of individual hands. Expressive timing, internal timekeeper variance, and motor response delays are all part of the same measurement of interonset durations, similar to how phrase-final lengthening and the time it takes to produce certain phonetic features are expressed in speech.

Musical prosody appears to be sufficient for signaling segmentation, prominence, and emotional states; there is less known about the role of prosody in how musicians coordinate their performances. Musical prosody may be *necessary* only for signaling prominence, and, it seems, in promoting emotional communication during development. The inherent ambiguity in musical structure may *require* performers to make use of acoustic variability to encode which musical features are more important than others. Prosodic cues, such as tempo, intensity, timbre, and pitch changes, are redundant in music performance; multiple cues reinforce the same segmentation, prominence, etc. Whether musical prosody has a grammar of its own is less obvious. Some rules express well the relation between prominence and acoustic realization, whereas other aspects of performance expression are not well-expressed in rules. Nevertheless, the high degree of consistency in how performers use prosodic features suggests that study of musical prosody, separate from study of musical structure, will inform us about auditory cognition in general.

Two lines of developing research address further the significance of musical prosody: the first is from experiments that manipulate the acoustic parameters of musical prosody, especially in comparison with parameters of linguistic prosody through a continuum from music to speech. The second is neurological evidence from brain-damaged, as well as normal individuals who acquire musical training either early or later in life, mapping the functions of musical prosody to their structural mechanisms. These experimental and neurological techniques allow us to address interesting questions, such as: Is prosody the underpinning of language and music development? Are musical genres, like languages, points on a prosodic continuum that is bounded by general perceptual principles of rhythm, grouping, and prominence? Do individual differences in music or language abilities arise from

differences in sensitivity to prosody? A fully-fledged theory of musical prosody, we hope, can address these questions.

Acknowledgments

Caroline Palmer and Sean Hutchins, Department of Psychology, McGill University, 1205 Dr Penfield Ave, Montreal QC H3A 1B1, Canada. This paper was supported by NSERC 298173 to Caroline Palmer and by the Canada Research Chairs program, and a McGill University Tomlinson Fellowship to Sean Hutchins. The authors thank Shari Baum, Julie Boland, Melissa Jungers, Isabelle Peretz, Peter Pfordresher, Brian Ross, and Jenny Saffran for comments on an earlier draft.

References

Adachi, M., & Trehub, S. E. (1998). Children's expression of emotion in song. *Psychology of Music, 26*, 133–153.

American Heritage Dictionary of the English Language® (2004). *Meter* (4th ed.). Houghton Mifflin Company. http://www.bartleby.com/61/17/M0251700.html

Apel, W. (1972). *Harvard dictionary of music.* Cambridge, MA: Harvard University Press.

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

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17*, 43–64.

Battel, G. U., Bresin, R., De Poli, G., & Vidolin, A. (1994). Neural networks vs. rules system: Evaluation test of automatic performance of musical scores. In *Proceedings of the 1994 International Computer Music Conference* (pp. 109–113). San Franciso: ICMA.

Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes, 11*, 17–67.

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

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

Breslin, R., De Poli, G., & Vidolin, A. (1992). Symbolic and sub-symbolic rules system for real time score performance. In *Proceedings of the 1992 International Computer Music Conference* (pp. 325–327). San Francisco: ICMA.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, and S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: MIT Press.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments.* Berkeley, CA: University of California Press.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America, 95*, 1570–1580.

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception, 10*, 317–341.

Clarke, E. F., & Baker-Short, C. (1987). The imitation of perceived rubato: A preliminary study. *Psychology of Music, 15*, 58–75.

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (Vol. 39, pp. 76–181). Stockholm: Royal Swedish Academy of Music.
Clynes, M. (1995). Microstructural musical linguistics: Composers' pulses are liked most by the best musicians. *Cognition, 55*, 269–310.
Cogan, R. D., & Escot, P. (1976). *Sonic design: The nature of sound and music*. Englewood Cliffs, NJ: Prentice-Hall.
Collier, R., & Hart, J. (1975). The role of intonation in speech perception. In A. Cohen and S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 107–121). Heidelberg: Springer-Verlag.
Cooper, W. E., & Eady, S. J. (1986). Metrical phonology in speech production. *Journal of Memory and Language, 25*, 369–384.
Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.
Cutler, A. (1995). Spoken word recognition and production. In J. Miller and P. Eimas (Eds.), *Speech, language, and communication* (pp. 97–136). New York: Academic Press.
Cutler, A. (1999). Prosody and intonation, processing issues. In R. A. Wilson and F. C. Keil (Eds.), *The MIT encyclopedia of the cognitive sciences* (pp. 682–683). Cambridge: MIT Press.
Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*, 141–201.
Cyr, M. (1992). *Performing baroque music.*. Aldershot: Scholar Press.
Davies, S. (2001). Philosophical perspectives on music's expressiveness. In P. Juslin and J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 23–44). Oxford, England: Oxford University Press.
Desain, P., & Honing, H. (1992). *Music, mind, and machine: Studies in computer music, music cognition, and artificial intelligence*. Amsterdam: Thesis.
De Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America, 91*, 491–504.
Deutsch, D., & Feroe, F. (1981). Internal representation of pitch sequence in tonal music. *Psychological Review, 88*, 503–522.
Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, and S. Brown (Eds.), *The origins of music* (pp. 389–410). Cambridge, MA: MIT Press.
Dogil, G. (2003). Understanding prosody. In G. Rickheit, J. Herrmann, and W. Deutsch (Eds.), *Psycholinguistics: An international handbook* (pp. 544–566). Berlin: Mouton de Gruyter.
Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10*, 343–378.
Drake, C., Penel, A., & Bigand, E. (2000). Tapping in time with mechanically and expressively performed music. *Music Perception, 18*, 1–23.
Ekman, P. (1973). Darwin and cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression* (pp. 1–83). New York: Academic Press.
Fais, L. (1994). Conversation as collaboration: Some syntactic evidence. *Speech Communication, 15*, 231–242.
Feldman, J., Epstein, D., & Richards, W. (1992). Force dynamics of tempo change in music. *Music Perception, 10*, 185–204.
Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*, 181–195.
Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60*, 1497–1510.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow, L. Cosmides, and J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford: Oxford University Press.
Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*, 657–674.
Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477–501.
Ferreira, F. (2002). Prosody. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 258–265). New York: Nature Publishing.
Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development, 67*, 3192–3218.
Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin, 97*, 412–429.
Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandanavian Journal of Psychology, 15*, 63–72.
Gabrielsson, A. (1987). Once again: The theme from Mozart's piano Sonata in A Major (k.331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81–104). Stockholm: Royal Swedish Academy of Music.
Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 501–602). San Diego, CA: Academic Press.
Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music, 24*, 68–91.
Gabrielsson, A., & Lindstrom, E. (2001). The influence of musical structure on emotional expression. In P. Juslin and J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 223–248). Oxford: Oxford University Press.
Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport, 9*, 2115–2119.
Gleitman, L. R., & Wanner, E. (1982). Language acquisition: The state of the state of the art. In E. Wanner and L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 3–48). New York: Cambridge University Press.
Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? *The Journal of the Acoustical Society of America, 110*, 563–572.
Gout, A., Christophe, A., & Morgan, J. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language, 51*, 547–567.
Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven and N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 515–546). Berlin: Mouton de Gruyter.
Grosjean, F. H., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology, 11*, 58–81.
Heilman, K. M., Bowers, D., Speedie, L., & Coslett, H. B. (1984). Comprehension of affective and nonaffective prosody. *Neurology, 34*, 917–921.
Henderson, M. T. (1936). Rhythmic organization in artistic piano performance. In C. E. Seashore (Ed.), *Objective analysis of musical performance* (Vol. 4, pp. 281–305). Iowa City: University of Iowa Press.
Hevner, K. (1935a). The affective character of the major and minor modes in music. *American Journal of Psychology, 47*, 103–118.
Hevner, K. (1935b). Expression in music: A discussion of experimental studies and theories. *Psychological Review, 47*, 186–204.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48*, 248–268.
Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology, 49*, 621–630.
Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright-Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition, 26*, 269–286.
Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception, 21*, 267–272.
Johnson-Laird, P. N. (1991). Jazz improvisation: A theory at the computational level. In P. Howell, R. West, and I. Cross (Eds.), *Representing musical structure*. San Diego, CA: Academic Press.
Jungers, M. K., Palmer, C., & Speer, S. R. (2002). Time after time: The coordinating influence of tempo in music and speech. *Cognitive Processing, 1*, 21–35.
Jusczyk, P. W., & Kemler Nelson, D. G. (1996). Syntactic units, prosody, and psychological reality during infancy. In J. L. Morgan and K. Demuth (Eds.), *Signal to syntax* (pp. 389–408). Mahwah, NJ: Erlbaum.
Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 1–14.
Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*, 252–293.
Juslin, P. N. (2001). Communicating emotion in music performance: A review and theoretical framework. In P. Juslin and J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 2309–2337). Oxford: Oxford University Press.
Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770–814.
Juslin, P. N., & Madison, G. (1999). The role of timing patterns in recognition of emotional expression from musical performance. *Music Perception, 17*, 197–221.
Juslin, P. N., & Sloboda, J. A. (2001). *Music and emotion: Theory and research*. New York: Oxford University Press.
Kelly, M. H., & Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception & Performance, 14*, 389–403.
Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*, 129–164.
Kotlyar, G. M., & Morozov, V. P. (1976). Acoustic correlates of the emotional content of vocalized speech. *Soviet Physics Acoustics, 22*, 370–376.
Krauss, R. M., Curran, N. M., & Ferleger, N. (1983). Expressive conventions and the cross-cultural expression of emotion. *Basic and Applied Social Psychology, 4*, 295–305.
Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics, 10*, 163–175.
Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51*, 336–353.
Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1*, 70–73.
La Driere, J. C. (Driere 1993). Prosody. In A. Preminger (Ed.), *Encyclopedia of poetry and poetics*. Princeton: Princeton University Press.

Lehiste, I. (1972). Timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America, 51*, 2018–2024.
Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
Levelt, W. P. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics, 56*, 414–423.
Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception & Psychophysics, 58*, 927–935.
Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*, 249–336.
Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: Chicago University Press.
Neumann, F. (1982). *Essays in performance practice*. Ann Arbor: UMI Research Press.
Nicholson, K. G., Baum, S., Kilgour, A., Koh, C. K., Munhall, K. G., & Cuddy, L. L. (2003). Impaired processing of prosodic and musical patters after right hemisphere damage. *Brain and Cognition, 52*, 382–389.
Palmer, C. (1989a). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception & Performance, 15*, 331–346.
Palmer, C. (1989b). Structural representations of music performance. In *Proceedings of the 11th Annual Conference of the Cognitive Science Society* (pp. 349–356). Hillsdale: Erlbaum.
Palmer, C. (1996a). Anatomy of a performance: Sources of musical expression. *Music Perception, 13*, 433–454.
Palmer, C. (1996b). On the assignment of structure in music performance. *Music Perception, 14*, 21–54.
Palmer, C., Jungers, M., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45*, 526–545.
Palmer, C., & Krumhansl, C. L. (1987). Pitch and temporal contributions to musical phrase perception: Effects of harmony, performance timing, and familiarity. *Perception & Psychophysics, 41*, 505–518.
Palmer, C., & van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception & Performance, 21*, 947–962.
Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development, 13*, 539–545.
Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress language. *Infant Behavior and Development, 14*, 415–440.
Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87*, B35–B45.
Patel, A. D., Peretz, I., Tramo, M., & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language, 61*, 123–144.
Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America, 109*, 1668–1680.
Pell, M. D., & Baum, S. R. (1997). Unilateral brain damage, prosodic comprehension deficits, and the acoustic cues to prosody. *Brain & Language, 57*, 195–214.
Peretz, I. (2001). Listen to the brain: A biological perspective on musical emotions. In P. Juslin and J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 105–134). Oxford, England: Oxford University Press.
Peretz, I., & Gagnon, L. (1999). Dissociation between recognition and emotional judgment for melodies. *Neurocase, 5*, 21–30.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition, 68*, 111–141.
Peretz, I., Gaudreau, D., & Bonnel, A.-M. (1998). Exposure effects on music preference and recognition. *Memory and Cognition, 26*, 884–902.
Peretz, I., Belleville, S., & Fontaine, S. (1997). Dissociations entre musique et langage après atteinte cérébrale: Un nouveau cas d'amusie sans aphasie. *Canadian Journal of Experimental Psychology, 51*, 354–368.
Pierrehumbert, J. (1999). Prosody and intonation. In R. A. Wilson and F. C. Keil (Eds.), *MIT encyclopedia of the cognitive sciences* (pp. 679–682). Cambridge, MA: MIT Press.
Pike, K. (1945). *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press.
Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson and J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.
Pressing, J. (1988). Improvisation: Methods and models. In J. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition*. New York: Oxford University Press.
Protopapas, A., & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *Journal of the Acoustical Society of America, 101*, 2267–2277.
Raffman, D. (1993). *Language, music, and mind*. Cambridge, MA: MIT Press.
Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265–292.
Rasch, R. A. (1978). The perception of simultaneous notes as such in polyphonic music. *Acustica, 40*, 21–33.
Rasch, R. A. (1979). Synchronization in performed ensemble music. *Acustica, 43*, 121–131.
Rasch, R. A. (1988). Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 70–90). Oxford: Clarendon Press.
Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America, 88*, 622–641.
Repp, B. H. (1992a). A constraint on the expressive timing of a melodic gesture: Evidence from performance and aesthetic judgment. *Music Perception, 10*, 221–241.
Repp, B. H. (1992b). Diversity and commonality in music performance. An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America 92*, 2546–2568.
Repp, B. H. (1992c). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition, 44*, 241–281.
Repp, B. H. (2000). Pattern typicality and dimensional interactions in pianists' imitation of expressive timing and dynamics. *Music Perception, 18*, 173–211.
Ross, E. D., Thompson, R. D., & Yenkosky, J. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain and Language, 56*, 27–54.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.
Schellenberg, E. G., & Trehub, S. E. (2003). Accurate pitch memory is widespread. *Psychological Science, 14*, 262–266.
Scherer, K. R. (1985). Vocal affect signaling: A comparative approach. In J. Rosenblatt, C. Beer, M. Busnel, and P. J. B. Slater (Eds.), *Advances in the study of behavior* (pp. 189–244). New York: Academic Press.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*, 143–165.
Scherer, K. R., Ladd, D. R., & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *Journal of the Acoustical Society of America, 76*, 1346–1356.
Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. Juslin and J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361–392). Oxford, England: Oxford University Press.
Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*, 341–349.
Seashore, H. G. (1938a). An objective analysis of artistic signing. In C. E. Seashore (Ed.), *University of Iowa studies in the psychology of music: Objective analysis of musical performance* (Vol. 4, pp. 12–157). Iowa City: University of Iowa.
Seashore, C. E. (1938b). *Psychology of music*. New York: McGraw-Hill.
Selkirk, S. (1984). *Phonology and syntax*. Cambridge, MA: MIT Press.
Shackford, C. (1961). Some aspects of perception, Part I. *Journal of Music Theory, 5*, 162–202.
Shackford, C. (1962). Some aspects of perception, Part II. *Journal of Music Theory, 6*, 66–90.
Shaffer, L. H. (1981). Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology, 13*, 326–376.
Shaffer, L. H., & Todd, N. P. (1987). The interpretive component in musical performance. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 139–152). Stockholm: Royal Swedish Academy of Music.
Shaffer, L. H. (1992). How to interpret music. In M. R. Jones and S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 263–278). Washington: APA.
Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 193–247.
Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology A, 35*, 377–396.
Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Clarendon Press.
Sloboda, J. A. (1992). Empirical studies of emotional response to music. In M. R. Jones and S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 33–46). Washington: APA.
Small, A. M. (1937). An objective analysis of artistic violin performance. In C. E. Seashore (Ed.), *University of Iowa studies in the psychology of music: Objective analysis of musical performance* (Vol. 4, pp. 172–231). Iowa City: University of Iowa.
Sundberg, J., Askenfelt, A., & Frydén, L. (1983). Musical performance: A synthesis-by-rule approach. *Computer Music Journal, 7*, 37–43.
Sundberg, J., & Frydén, L. (1985). Teaching a computer to play melodies musically. In L. Anders and E. Kjellberg (Eds.), *Analytica: Studies in the description and analysis of music* (Vol. 1, pp. 67–76). Uppsala: Aeta University.
Sundberg, J., Friberg, A., & Frydén, L. (1989). Rules for automated performance of ensemble music. *Contemporary Music Review, 3*, 89–109.
Sundberg, J., Friberg, A., & Frydén, L. (1991). Threshold and preference quantities of rules for music performance. *Music Perception, 9*, 71–92.
Sundberg, J., & Verrillo, V. (1980). On the anatomy of the ritard: A study of timing in music. *Journal of the Acoustical Society of America, 68*, 772–779.
Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech, 37*, 21–43.
Thiessen, E. D., Hill, E. A., Emerson, D. R., & Saffran, J. R. (2005). Effects of prosodic structure on infants' statistical learning. Paper presented at the Cognitive Development Society, October 2005.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*, 53–71.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*, 46–64.

Thompson, W. F., & Cuddy, L. L. (1997). Music performance and the perception of key. *Journal of Experimental Psychology: Human Perception & Performance, 23*, 116–135.

Todd, N. P. M. (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33–58.

Todd, N. P. M. (1989). A computational model of rubato. *Contemporary Music Review, 3*, 69–88.

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

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

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science, 11*, 188–195.

Trainor, L. J., Clark, E. D., Huntley, A., & Adams, B. (1997). The acoustic basis for infant-directed singing. *Infant Behavior and Development, 20*, 383–396.

Trainor, L. J., & Rock, A. M. L. (1997). Distinctive messages in infant-directed lullabies and play songs. Presented at the Biennial Meeting of the Society for Research in Child Development, Washington DC.

Trainor, L. J., Wu, L., & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science, 7*, 289–296.

Trehub, S. E. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker, and S. Brown (Eds.), *The origins of music* (pp. 427–448). Cambridge, MA: MIT Press.

Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. In C. Rovee-Collier and L. Lipsitt (Eds.), *Advances in infancy research* (pp. 43–77). Norwood, NJ: Ablex.

Trehub, S. E., Unyk, A. M., & Henderson, J. L. (1994). Children's songs to infant siblings: Parallels with speech. *Journal of Child Language, 21*, 735–744.

Trehub, S. E., Unyk, A. M., Kamenetsky, S. B., Hill, D. S., Trainor, L. J., Henderson, J. L., & Rasaza, M. (1997). Mothers' and fathers' singing to infants. *Developmental Psychology, 33*, 500–507.

Wennerstrom, A. K. (2001). *The music of everyday speech: Prosody and discourse analysis.* Oxford: Oxford University Press.

Wing, A., & Kristofferson, A. B. (1973). The timing of interresponse intervals. *Perception & Psychophysics, 13*, 455–460.