Music Perception
Spring 2004, Vol. 21, No. 3, 289–311

© 2004 BY THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
ALL RIGHTS RESERVED.

Music and Language: A Developmental Comparison

ERIN McMULLEN & JENNY R. SAFFRAN
University of Wisconsin-Madison

The possible links between music and language continue to intrigue scientists interested in the nature of these two types of knowledge, their evolution, and their instantiation in the brain. Here we consider music and language from a developmental perspective, focusing on the degree to which similar mechanisms of learning and memory might subserve the acquisition of knowledge in these two domains. In particular, it seems possible that while adult musical and linguistic processes are modularized to some extent as separate entities, there may be similar developmental underpinnings in both domains, suggesting that modularity is emergent rather than present at the beginning of life. Directions for future research are considered.

On the surface, music and language are wildly different. No listener would ever confuse a Beethoven sonata with a political speech. By the time we reach adulthood, we possess vast and distinct arrays of knowledge about each of these domains. However, from the perspective of a naïve listener such as an infant, the two systems may not seem quite so different. The purpose of this article is to consider the learning processes facing infants as they are confronted with their native musical and linguistic systems, and to examine the extent to which the developmental trajectories of linguistic and musical skills may rest upon the same learning capacities, at least early in life. In some cases, a single mechanism might underlie learning in both domains. In other cases, distinct mechanisms might be required, given the nature of the learning tasks confronting the child in each domain.

It is common for scholars in cognitive science and cognitive neuroscience interested in music to draw comparisons with language. Why is language the domain most likely to be considered as a contrast to music? Unlike other universal domains of human expertise such as vision or social organization, both music and language (included signed languages) are orga-

Address correspondence to Jenny R. Saffran, Department of Psychology, University of Wisconsin-Madison, Madison, WI 53706. (e-mail: jsaffran@wisc.edu)

ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

nized temporally, with the relevant structures unfolding in time. Furthermore, spoken languages, like music, reach our perceptual system as frequency spectra, arrayed as pitches. In both cases, some of the pertinent systematicities are universal, whereas others are culturally specific. For instance, all languages consist of phonemes, and all musical systems consist of notes; however, the specific sets thereof will vary from culture to culture.

Such comparisons are necessarily limited when one considers the kinds of information communicated by music versus language; it is evident that the communicative uses of the two domains are vastly different. However, from the perspective of the youngest listeners, who must learn about each system before discovering its communicative intent, the similarities between music and language may be heightened. This is particularly true when one distinguishes between competence-knowledge about one's native linguistic or musical system-and performance-one's ability to use this knowledge communicatively, which is affected by many other factors. The difference between comprehension and production is particularly evident in the early stages of first language acquisition; parents are keenly aware that their infants understand far more than they can produce throughout infancy and early childhood. From this perspective, considerations of music focus not on skilled performance as a function of organized music lessons, but on the knowledge of music gained implicitly from the musical exposure that is ubiquitous in children's lives. As with language, this process involves inducing structure from environmental input. Learning, combined with inherent perceptual and cognitive predispositions, eventually renders adult knowledge in each domain.

Another reason for interest in comparisons between language and music is the broader theoretical issue of modularity of mind: to what extent are cognitive processes specifically tied to particular domains? In other words, is there a neural system underlying music that is either somewhat or entirely distinct from the neural system underlying language? Modularity is a central theoretical construct in cognitive science, and different conceptualizations of modularity suggest different possible relationships between language and music. Fodor's (1983) classic formulation of modularity suggests that distinct architectural brain regions might subserve language and music, and that these regions are informationally encapsulated such that there is no cross-talk between them. On this view, the hardware and software underlying processing in each area are distinct. Peretz and Coltheart (2003) present evidence from neurologically impaired adults to support such a division, suggesting that the neural faculties subserving music are quite distinct from other processes in the brain.

Granting for the moment that some anatomical distinction between music and language may well exist in adults, it remains unclear whether we begin life with this neural specialization or whether it emerges as a function of experience with the two domains as distinct from one another. Although discussions of modularity are often confounded with assumptions about innateness, Karmiloff-Smith and others have pointed out that modularity and innateness of specific mental capacities are not inextricably linked (Elman et al., 1996; Karmiloff-Smith, 1992). Furthermore, as Elman and colleagues (1996) argue, the sheer quantity of genetic material that would be required to directly encode specific modules for higher cognition is quite unwieldy. Finally, although a specific module for language perception and comprehension would have plausible sexual selection advantages, the direct benefits of a music module are less obvious (but see Huron, 2003; Trehub, 2003). The modularity question remains an open and vexing issue, and significant further research is required to clarify it; some potential avenues for such research may be suggested by the comparisons of the ontogeny of music and language drawn here.

What Do We Learn

THE SOUND STRUCTURE OF LANGUAGE AND MUSIC

From the perspective of learning about linguistic and musical systems during development, where might we expect to see overlap and divergence? Here we focus on the acquisition of tacit competence; that is, musical and linguistic knowledge, rather than the skills required to produce these systems. Both (spoken) language and music are generated from a finite set of sounds (notes or phonemes), carved out of a larger possible set of sounds. These sounds are organized into discrete categories, facilitating representation and memory. For example, auditory events in both domains are subject to the process of categorical perception: it is extremely difficult to discriminate between sounds that are part of the same category (e.g., different versions of the syllable /ba/), whereas sounds that are equally physically different that span category boundaries are readily perceived (e.g., the switch from /ba/ to /pa/). Such auditory discontinuities are evident for speech sounds from early infancy (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971), and appear not to be linked to speech perception alone. Certain nonspeech stimuli that share similar auditory properties also elicit categorical perception in both adults and infants, such as the difference between a plucked and a bowed string (e.g., Cutting & Rosner, 1974; Jusczyk, Rosner, Cutting, Foard, & Smith, 1977). Musical materials can also be perceived categorically, even by nonmusicians; adults taught labels for musical intervals (e.g., "Here Comes the Bride" for the perfect fourth) perceive those intervals categorically (Smith, Kemler Nelson, Grohskopf, & Appleton, 1994). Although infants have not been tested for categorical perception of musical materials, the fact that they do show categorical perception for nonspeech analogs of particular consonant contrasts created from tones suggests that it is likely that they, too, would show categorical perception for at least some musically relevant auditory input (Jusczyk, Rosner, Reed, & Kennedy, 1989).

The auditory system helps to determine which categories are to play a role in each domain. Nonhuman species that presumably did not evolve to perceive speech appear to detect many of the same phonetic categories as humans (e.g., Kuhl & Miller, 1975; Kuhl & Padden, 1982). Moreover, these speech categories are perceived by infants whose native languages do not use them (for review, see Aslin, Jusczyk, & Pisoni, 1998). Thus, in the absence of experience, infants chunk auditory space into the speech sound repertoire from which human languages sample; for example, young Japanese infants treat /r/ and /l/ as members of two distinct categories, unlike Japanese adults. Similarly, infants show preferences for particular types of musical sounds. In particular, infants prefer consonant intervals over dissonant intervals from quite early in life (Trainor & Heinmiller, 1999; Trainor, Tsang, & Cheung, 2002; Zentner & Kagan, 1998). Some of these musical predispositions can also be demonstrated in nonhuman animals. For example, rhesus monkeys demonstrate octave equivalence only given tonal materials; atonal materials do not induce this perceptual capacity, suggesting auditory predispositions for tonality processing in nonhuman primates (Wright, Rivera, Hulse, Shyan, & Neiworth, 2000). Similarly, neurophysiological recordings in rhesus monkeys and humans suggest similar neural signatures for consonant versus dissonant materials (Fishman, 2001).

If linguistic and musical systems never varied across cultures, then these predispositions might be sufficient to explain how such systems are processed. However, the languages and musical systems of the world exhibit substantial differences. Thus, infants and young children must learn the specific features of the systems in their environments. By 6 months of age, infants' speech perception abilities are attuned to the vowels in their native language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), suggesting that just a few months of passive exposure to ambient speech is sufficient to shape infants' vowel processing; extensive experience with speech production is not required. Similarly, infants' consonant perception is attuned to the native language by 10 to 12 months of age (Werker & Lalonde, 1988). In both cases, infants have shifted from categorizing all speech sounds, regardless of their status in the native language, to discriminating contrasts between native-language categories differently than nonnative language categories. This nonnative to native shift implicates powerful learning abilities that operate implicitly during the first year of life. Somehow, infants are able to learn which sounds mark meaningful differences in their language before they have extensive access to word meanings.

Although the precise learning mechanisms underlying the nonnative to native shift in speech processing remain unknown, it appears that the statistical distributions of speech sounds in the input may play a critical role. Speech sounds that are part of the same category in a given language cluster together in different ways than speech sounds that are part of different categories, and this appears to be information that is useable by learners. For example, the shift in vowel processing observed in 6-month-old infants can be induced in nonhuman animals exposed to vowel distributions akin to those of particular languages, such as English versus Swedish (Kluender, Lotto, Holt, & Bloedel, 1998). Experiments explicitly manipulating the statistical distributions of consonants have demonstrated that this information affects infants' speech perception (Maye, Werker, & Gerken, 2002). It thus seems likely that the learning process that underlies the Japanese infant's discovery that /r/ and /l/ are part of the same category is mediated by the statistical structure of the sounds of Japanese.

Some similar shifts have been observed in the domain of music, on the level of scale structures. In many instances, young infants show little to no effect of Western tonal conventions on their perception of and memory for musical stimuli. Before 1 year of age, infants detect changes to a melody irrespective of whether the new note conforms to the implied harmony of the original melody (8-month-olds: Trainor & Trehub, 1992) or to the diatonic context in general (9- to 11-month-olds: Trehub, Cohen, Thorpe, & Morrongiello, 1986). Similarly, 9-month-olds are able to detect deviations from a standard melody regardless of whether the standard conforms to Western tonal conventions (Schellenberg & Trehub, 1999). By the time they start school, however, Western listeners' responses to stimuli begin to show influences of typical Western musical structure. Four- to six-year-old children detect changes in a diatonic melody more easily than in a nondiatonic melody (Trehub et al., 1986). Similar performance effects of changed notes differing by key (at 5 years) and harmony (at 7 years) were shown by Trainor and Trehub (1994). With 7- and 9-year-olds, pairwise discrimination is likewise enhanced by the nontonality of at least one of the members (Wilson, Wales, & Pattison, 1997). Scale structure has its greatest effect on memory for melody when the stimuli contain repeated notes (Schellenberg & Trehub, 1999). These influences continue to be seen throughout adulthood (Trainor & Trehub, 1992), though as with children, the effects are most pronounced with redundant melodies (Schellenberg & Trehub, 1999). Available correlational data on distributional properties of Western music (Budge, 1943; Krumhansl, 1990) indicates that statistical induction is a plausible mechanism for this kind of learning, as in the case of phonemes. Notes and chords that are considered the strongest fit within many Western musical contexts-at the top of Krumhansl's tonal hierarchy-tend to occur in music more often than other, less structurally important notes and chords. However, no experiments using artificial musical grammar induction techniques have tested this hypothesis specifically, so conclusions must be somewhat more tentative than in the linguistic case.

THE PROSODIC STRUCTURE OF LANGUAGE AND MUSIC

Moving beyond the segmental level (phonemes and tones), it is clear that the suprasegmental cues in both language and music are highly salient to infants. These patterns of rhythm, stress, intonation, phrasing, and contour most likely drive much of the early processing in both domains. Such prosodic information is the first human-produced external sound source available in utero; the filtering properties of the fluid-filled reproductive system leave rhythmic cues intact relative to high-frequency information. Fetuses avail themselves of the incoming rhythmic patterns; again, this is a process of implicit, nonreinforced learning. For example, newborn infants prefer their mother's voice on the basis of prenatal learning (DeCasper & Fifer, 1980). Fetal learning also encompasses the rhythmic patterns of the mother's native language, allowing newborn infants to use this experience to differentiate between languages (Mehler et al., 1988). It is likely that infants acquire specific information about musical rhythmic information in their prenatal environments as well, assuming sufficient exposure and quality of auditory input (maternal singing is presumably the best source of such input).

After birth, infants continue to be attuned to prosodic information in both domains. This may be in part due to learning in the womb. It is also likely a function of the richness of the prosodic structure in the infants' environments. Both linguistic and musical input are modified by caregivers in ways that appear to be maximally attractive to infants. Infant-directed speech, in comparison to adult-directed speech, is characterized cross-linguistically by a slower rate of speech, higher fundamental frequency, greater range of pitch variation, longer pauses, and characteristic repetitive intonation contours (Fernald, 1992). Other modifications also might enhance early learning; for example, vowels in infant-directed speech are produced in a more extreme manner, resulting in heightened distinctiveness between vowel categories (Kuhl, Andruski, Chistovich, & Chistovich, 1997). Infant-directed speech captures infants' attention more readily than adult-directed speech (Cooper & Aslin, 1990; Fernald, 1985). Moreover, learning appears to be facilitated by the exaggerated prosodic contours of infant-directed speech; infants detect word boundaries in fluent speech more readily when the same items are spoken with infant-directed prosody as opposed to adult-directed prosody (Thiessen, Hill, & Saffran, 2004).

Caregivers also engage musically with their infants in ways that differ from adult-directed music. The play-songs and lullabies that dominate infant-directed music share features cross-culturally, including simple, repeated pitch contours (e.g., Trehub & Trainor, 1998). This is sensible, as contour is one of the first aspects of music to be discriminated by infants (Trehub, 2003). Interestingly, renditions of caregivers' songs tend to be sung at the same pitch; when a given mother is recorded singing a particular song on different days, she will tend to use the same key and absolute pitches (Bergeson & Trehub, 2002). As with infant-directed speech, these songs are preferred by infants from early in life (e.g., Masataka, 1999; Trainor, 1996). In both domains, the affective properties of the infant-directed register appear to be central; before the availability of other forms of communication, the prosodic contours in these two domains are a primary means of transmitting emotional information (for a recent review, see Trehub, 2003).

Prosodic cues may also play a role in delineating the structural information that infants must learn to process in language and music. Prosodic cues are probabilistically correlated with structural boundaries such as clausal and phrasal units; for example, ends of clauses in speech are marked by syllable lengthening and a drop in pitch. The preferential listening methodology has been used assess the degree to which infants are aware of these correlations. By 7 months of age, infants listen longer to speech samples in which pauses fall at clause boundaries than to speech samples in which pauses are placed clause-internally (Hirsh-Pasek, Kemler Nelson, Jusczyk, & Cassidy, 1987). Interestingly, this is only the case when the speech samples are derived from infant-directed speech; the more muted prosodic structure of adult-directed speech does not facilitate the detection of structural boundaries by infants (Kemler Nelson, Jusczyk, Hirsh-Pasek, & Cassidy, 1989). The results of these and similar experiments (see Jusczyk, 1997, for a review) suggest that infants detected whether or not the pauses matched with the other prosodic boundary markers in the speech. Similar results emerge from studies using musical materials (Mozart minuets); infants listen longer to musical passages where pauses are placed at the ends of phrases rather than in the middles of phrases (Jusczyk & Krumhansl, 1993; Krumhansl & Jusczyk, 1990). Analysis of the musical materials suggests that the same prosodic markers function at musical phrase boundaries as observed with linguistic materials: a decline in pitch and a lengthening of the final note. It remains an open question whether infants are using the same mechanism to detect these parallel cues across domains, or whether instead they have learned about these similar prosodic properties independently.

Linguistic prosody and musical structure within a given culture may also relate in a more specific fashion. It is known that languages vary in their overall patterns of prosodic stress. Similarly, certain nationalistic styles of music are distinguished by their usage of certain characteristic rhythmic themes. Might the stress patterns of one's native language influence a composer's style, even for instrumental works? Patel and Daniele (2003) performed a quantitative analysis of a small corpus of nationalistic music from Britain and France, two countries whose natives speak languages that are rhythmically quite distinct; British English tends to have a high degree of syllable-to-syllable variability in syllable length, whereas French syllables are somewhat more evenly timed. Using a measure of rhythmic variability first developed for language, the authors discovered that British and French music differ in much the same manner and direction as do British and French speech. Although the study compared only a strictly delimited set of samples taken from two cultures, the methodology is quite interesting and paves the way for future empirical comparisons of speech and music.

THE GRAMMATICAL STRUCTURE OF LANGUAGE AND MUSIC

As important as segmental and suprasegmental cues are for learning both musical and linguistic systems, the real power of these systems comes from their infinitely combinatorial nature. Both types of systems contain a wealth of culture-specific, nuanced rules for well-formed strings that must be learned before adult-level comprehension can occur (Chomsky, 1957; Lerdahl & Jackendoff, 1983). Although detecting the statistical distributions of sounds affords clear benefits to those engaged in learning surface properties of a linguistic or musical system, it has been less obvious how one might use similar statistical information to learn abstract grammatical rules about what to do with concrete nouns or seventh chords. Indeed, for many years, the dominant theoretical position in linguistic theory has been that infants come prewired with a "universal grammar," a dedicated linguistic system containing a combination of innate knowledge and toggle switches for universal aspects of possible native languages, such as possible types of word order, which are set as the native language is learned (Chomsky, 1981).

More recently, there has been interest in directly investigating the ability of infants to infer grammatical structure from linguistic input. Marcus and colleagues showed that infants exposed to a corpus of short sentences following a simple pattern (e.g., AAB) will reliably attend more to novel sentences which fail to conform to the pattern (e.g., ABA), indicating that by age 7 months, humans are capable of pattern induction (Marcus, Vijayan, Bandi Rao, & Vishton, 1999) . Slightly older infants have demonstrated similarly impressive abilities with more complex strings of individually uttered words generated by a finite-state grammar (Gomez & Gerken, 1999). Impressive though this learning is, the artificial pauses between words make the task an easier one than natural language learning, where the stimulus stream is unsegmented. Can infants use the output of a statistical parser to accomplish a similar task? In a recent experiment, Saffran and Wilson (2003) exposed 12-month-old infants to a finite-state artificial grammar, presented in sets of fluent sentences, where the only cues to word boundaries were statistical. Infants were then tested with novel sentences that were either grammatical or ungrammatical. Importantly, the first-order transitional probabilities between syllables were identical for both grammatical and ungrammatical sentences. Thus, in order to succeed at this task, infants first had to segment the words in the new language and then learn the permissible orderings of these words. Nevertheless, infants showed a preference for grammatical over ungrammatical sentences, indicating that multilevel statistical learning is possible in a short time frame. These results are mirrored by evidence that infants are aware of some of the aspects of the grammatical structure of their native language early in the second year of life (e.g., Hirsh-Pasek & Golinkoff, 1996).

Our knowledge about how infants and young children learn the grammar of their native musical system is more limited, but available data indicate that development of this knowledge is slower, emerging somewhat later in ontogeny. We know that experienced listeners have a hierarchical internal representation of relevant structures within their musical idiom that govern expectations of what is to come next (for a review, see Krumhansl, 1990). Using a probe methodology in which subjects are asked which of a set of stimuli best completes a given context, Western listeners preferentially end pieces on the tonic, less frequently on other notes in the tonic chord, still less frequently on other notes within the scale, and rarely on notes outside the diatonic context (Krumhansl, 1990). No evidence has yet been provided that infants use similar information in responding to musical stimuli (though see Saffran, 2003b for preliminary evidence that 8-month-old infants treat tonal and atonal materials differently). However, by 5 years of age, Western children show some knowledge of Western tonal structure (Trehub et al., 1986), and by 7 years this knowledge is comparable to an adult's (Speer & Meeks, 1988). The reason for the slower pace of this type of grammar acquisition is unclear. Two possibilities are that (1) infants are exposed to fewer examples of musical phrases than linguistic ones (which seems likely, though this has not been shown quantitatively), and (2) the practical communicative benefits of knowledge about tonal structure are fewer than the benefits of structural linguistic knowledge.

Advances in neuroscientific techniques have contributed to our understanding of syntactic processing in both domains (for a recent review of these techniques, see Tervaniemi & van Zuijen, 1999). Using electroencephalography (EEG), Osterhout and Holcomb (1992, 1993) found that words that are difficult to integrate into their surrounding sentences elicit a positive brain potential 600 ms after their onset. This P600 component (also referred to as the syntactic positive shift, or SPS) appears to be insensitive to other errors, such as semantically inappropriate word substitutions. In an exciting extension of this line of research, Patel, Gibson, Ratner, Besson, & Holcomb (1998) showed that a comparable P600 response is elicited by unexpected musical events. In this study, subjects heard six kinds of stimuli: sentences containing target words that were easy, difficult, or impossible to integrate meaningfully, and chord sequences containing target chords that were in the key, in a nearby key, and in a distant key. P600 responses emerged for all four types of anomalies, with larger peaks corresponding to more severe integration difficulty. Moreover, the responses were quite similar for speech and music stimuli of comparable difficulty. Others have found further similarities in early neural processing of linguistic and musical syntax, both of which make use of Broca's area and its right-hemisphere homologue. In particular, syntactically incongruous words and out-of-key chords tend to elicit an early negative component, stronger on the left in response to speech stimuli (ELAN: Hahne & Friederici, 1999) and on the right for musical stimuli (ERAN: Koelsch, Gunter, Friederici, & Schröger, 2000; Maess, Koelsch, Gunter, & Friederici, 2001). Importantly, this response has been demonstrated in nonmusicians, indicating that explicit training is not necessary for this level of implicit knowledge to develop.

Meaning in Language and Music

Although lower-level parallels between music and spoken language are relatively easy to discern, the relationship between the two at the level of semantics is less obvious and is likely where the systems diverge most strongly. Certainly a sonata does not carry referential meaning in the same way that a sonnet does. However, music can and does often elicit strong, predictable emotional responses from people who may vary by culture. The nature of these mappings from sound to response, and the means of their construction, is of interest to the cognitive scientist; thus a comparison of the two "meaning-systems," loosely defined, may be instructive.

In the case of music, the "meaning" that adult listeners give to phrases is most strongly related to the emotional responses they generate. We know that one of the basic building blocks for this is present from early infancy; several studies have found that infants as young as 2 months old, like adults, prefer consonance to dissonance (Trainor & Heinmiller, 1999; Trainor, Wu, Tsang, & Plantinga, 2002; Zentner & Kagan, 1998). This preference has long been posited as an important underpinning of emotion in music (e.g., Helmholtz, 1895). In addition, research has demonstrated infant preferences for higher-pitched music (Trainor & Zacharias, 1998), which often correlates with positive emotional judgments (e.g., Juslin & Laukka, 2003). Furthermore, as with lullabies (Trehub, Unyk, & Trainor, 1993a, 1993b), some emotional content in music is recognizable cross-culturally (Balkwill & Thompson, 1999), indicating a possible set of musical universals of emotion (for a detailed review, see Juslin & Laukka, 2003).

However, adult responses to specific pieces of music are complex and most likely influenced by a variety of other factors as well. For instance, many adults report having strong physiological reactions to certain musical pieces and to particular sections within them, including tears, heart acceleration, and "chills" or "shivers down the spine" (Sloboda, 1991). These emotional responses are differentiable physiologically (Nyklícek, Thayer, & Van Doornen, 1997) and electrocortically (Schmidt & Trainor, 2001), and, in the case of the "chills" reaction, are linked to increased blood-flow in brain regions associated with emotion, motivation, and arousal (Blood & Zatorre, 2001). One oft-cited explanation for emotional responses to music is that listeners are experiencing alternating tension and relaxation in response to violation and confirmation of expectations (Meyer, 1956). As discussed earlier, at least some of these expectations must be culturally defined and are thus learned. In addition, emotional responses in adults may result from cultural associations with particular musical gestures. An example of this is the tendency for Western listeners to associate the major mode with happiness and the minor mode with sadness. Although this distinction is reliable in adults (Crowder, 1985) and children as young as 3 years old (Kastner & Crowder, 1990), 6-month-old infants fail to show a preference for either (Crowder, Reznick, & Rosenkrantz, 1991). In addition, in one study investigating affective associations of pieces of music, Nawrot (2003) found that infants looked longer at happy faces during the presentation of music judged "happy" by adults, but did not look longer at sad faces while music judged to be "sad" played. Whether these null results imply that infants cannot make these fine distinctions or merely that infants of this age have not yet learned the cultural associations of the modes, remains unclear.

What is the best parallel for this kind of meaning-building in the domain of language? Obviously, at the lexical level, no good correlate exists. However, human utterances carry meaning not only lexically, but also paralinguistically, through the use of intonation; it is here that we may find useful comparisons. As discussed earlier, a good deal of research exists pertaining to the exaggerated prosodic contours characteristic of infant-directed speech. In addition to cognitive and attentive benefits, some have suggested that a major function of infant-directed speech is emotional communication and bonding. Acoustic data showing few differences between infant-directed speech and emotional adult-directed speech lend support to this idea (Trainor, Austin, & Desjardins, 2000). One could easily make a similar argument for infant-directed music; certainly, some characteristics pertain to both, particularly infants' preference for higher pitched (happier, friendlier) utterances. However, whereas these infant responses to paralinguistic features are presumed to access some universals of human emotion, many complex adult responses to music would appear to require enculturation. Is there a useful linguistic parallel for this level of emotional communication? Perhaps a good place to look for such a link would be poetry, which, like music, makes use of basic prosodic cues but requires cultural and syntactic knowledge for full appreciation. Some theoretical work has been done comparing music and poetry as abstract cognitive domains (Lerdahl, 2003), but to our knowledge, no research has contrasted the ways in which they access emotion or the ways in which these emotional cues become available to developing children.

A final note on emotional meaning in music and speech: Recent neuroimaging work indicates that responses to nonlinguistic human vocal sounds are strongest in the right superior temporal area (Belin, Zatorre, & Ahad, 2002), near areas that have been implicated in processing of musical pitch in other studies (Zatorre, 2003). Whether this indicates a meaningful overlap between the two domains has yet to be seen. However, it at least lends plausibility to accounts of musical and linguistic evolution that emphasize emotional communication through prosody as a primary forebear of both systems.

How Do We Learn It?

MEMORY FOR LANGUAGE AND MUSIC

For successful learning to occur, young learners must be able to represent musical experiences in memory, permitting the subsequent accumulation and manipulation of knowledge. Interestingly, infants are remarkably adept at representing their auditory experiences in long-term memory. Jusczyk and Hohne (1997) assessed the linguistic long-term memory abilities of 7-month-old infants by repeatedly exposing them to brief stories. Following a 2-week retention interval, during which the infants did not hear these stories, they were tested to see whether the words from the stories were retained in long-term memory. The infants preferred to listen to a list of words taken from the stories compared with a new list of words, suggesting that they remembered the words last heard several weeks ago.

An analogous study using musical materials suggests similar abilities exist in infant musical memory (Saffran, Loman, & Robertson, 2000). Infants were exposed at home to CD recordings of Mozart piano sonata movements, played daily for 2 weeks. Following a 2-week retention interval, during which the infants did not hear these musical selections, they were tested on passages from the familiar pieces compared with novel passages drawn from other Mozart piano sonatas, performed by the same pianist. These infants were compared with a control group, consisting of infants who had never heard any of these pieces. As expected, the control group showed no preference for excerpts from the familiar versus the novel sonatas. However, the experimental group evidenced effects of their previous exposure to these pieces, showing a significant difference in listening preferences to the familiar versus the novel sonatas. Subsequent experiments demonstrated that the infants were not merely remembering random snippets of the music, but instead had represented aspects of the overall structure of the piece, with expectations regarding the placement of particular passages (Saffran et al., 2000). These results suggest that infants' musical memory may be as nuanced as their linguistic memory.

Other recent studies investigating infant long-term memory for music similarly suggest that infants' auditory representations are quite detailed. For example, infants can represent more complex pieces of music, such as Ravel piano compositions, in long-term memory (Ilari & Polka, 2002). Moreover, the content of infants' memories include some extremely specific aspects of musical performances. Ten-month-olds represent acoustic patterns drawn from the specific performances with which they were previously familiarized (Palmer, Jungers, & Jusczyk, 2001). Six-month-old infants remember the specific tempo and timbre of music with which they are familiarized, failing to recognize pieces when they are played at new tempos or with new timbres, although recognition is maintained when pieces are transposed to a new key (Trainor et al., 2002). It thus appears that infant representations are extremely specific, not affording the opportunity to generalize to include changes in tempo or timbre. This level of specificity must change with age, as either a function of experience or development, else listeners would not be able to recognize familiar music played on different instruments or at different rates. It should be noted, however, that the ability to remember specific performance characteristics like key, tempo, and timbre is not lost completely during development (Levitin, 1994, 1996; Palmer et al., 2001; Schellenberg, Iverson, & McKinnon, 1999; Schellenberg & Trehub, 2003). The ability to encode music abstractly complements the capacity to engage in absolute encoding.

Similar shifts in specificity obtain for linguistic materials. For example, 7.5-month-old infants include talker-specific cues in their representations of spoken words; they have difficulty recognizing words when they are spoken in new voices, whereas 10.5-month-olds do not (Houston & Jusczyk, 2000). However, even younger infants are able to ignore talker-specific properties under other circumstances—in particular, infants readily exhibit vowel normalization, categorizing individual exemplars according to vowel identity despite differences in speaker sex (Kuhl, 1979, 1983). Infants thus appear to process linguistic auditory events at multiple levels of detail simultaneously. We see similar abilities to track multiple levels of information in the domain of pitch perception; in some tasks, infants appear to track absolute pitches, with no evidence of relative pitch representations

(Saffran & Griepentrog, 2001), whereas slight task manipulations lead infants to focus on relative pitch representations (Saffran, Reeck, Niehbur, & Wilson, 2004). These findings are reminiscent of results with an avian species—starlings—who can switch from relying on absolute pitch cues to using relative pitch cues when necessitated by the structure of the task (MacDougall-Shackleton & Hulse, 1996).

More insight into the development of auditory memory is being provided by recent work using EEG with young infants. Important components of adult ERP responses are seen even shortly after birth (Kushnerenko, 2003), including the mismatch negativity (MMN), a preattentive measure of auditory change detection that is detected when a sequence of repetitive standard stimuli is interrupted by an infrequent one that deviates from the standard on a particular criterion of interest. The apparently short duration of the memory traces leading to the MMN have made infant research somewhat more difficult than studies using this method with adults (Cheour, Ceponiene, et al., 2002); however, some interesting results have nonetheless been obtained. Cheour et al. (1998) have demonstrated that between the ages of 6 months and 1 year, infants' processing of phonemic contrasts changes, consistent with prior behavioral data. In their study, they presented infants with one standard Finnish vowel, one deviant Finnish vowel, and one deviant Estonian vowel. They found that at 6 months, infants' EEG traces display a tendency to respond more strongly when an infrequent stimulus is more acoustically distinct—in this case, the Estonian vowel—whereas by 1 year, they exhibit larger MMNs to the less-distinct but phonemically different Finnish vowel (Cheour et al., 1998). Learning such contrasts is possible even in the youngest infants. Less than a week after birth, Finnish infants exposed to the vowel contrast /y/ versus /y/i/ while sleeping showed an MMN-like response to the less frequent sound, whereas those with no exposure or unrelated exposure showed no such response—indicating that newborn auditory memory is sufficient to permit the learning of new phonetic contrasts without any possibility of conscious attention (Cheour, Martynova, et al., 2002). To our knowledge, no comparable infant MMN research has been done involving musical stimuli. Given that the requisite MMN to pitch change is observable in young listeners, this is a fertile field for further investigation of infant memory.

Learning Mechanisms for Language and Music

Once learners have received sufficient exposure to musical and linguistic systems, they must somehow derive structure across the corpus of specific experiences represented in memory. Various types of learning mechanisms have been implicated in this process. We focus here on two such mechanisms: rules and statistics. Rules require learners to abstract away from the specific items in their experience to discover underlying structure. The classic formulation of this process comes from Chomsky (1959), who noted that while no listener had ever heard the sentence “Colorless green ideas sleep furiously,” that sentence was nevertheless grammatical (compare it with the ungrammatical “Furiously green sleep ideas colorless”). Similar ideas have been advanced to explain certain aspects of music cognition, including expectancies and decisions about well-formedness and grouping (Lerdahl & Jackendoff, 1983; Narmour, 2000). An experimental demonstration of this type of process is the study by Marcus et al. (1999) mentioned earlier. In this study, 7-month-old infants were exposed to sentences like “wo fe fe,” “ti la la,” and “bi ta ta.” They were then tested on novel sentences that followed the ABB rule, such as “ko ga ga,” versus novel sentences that violated the ABA rule, such as “ko ga ko.” The hallmark of rule-based learning is to have abstracted away from the particular elements in the input to recognize “grammatical” sequences that have not been heard before; the infants in Marcus’ study achieved this after just a few minutes of exposure.

Another learning mechanism that has received growing attention is statistical learning: detecting patterns of sounds, words, or other units in the environment that cue underlying structure (for a recent review, see Saffran, 2003a). The environment contains massive amounts of statistical information that is roughly correlated with various levels of structure. For example, the probabilities with which syllables follow one another serve as cues to word boundaries; syllable sequences that recur consistently are more likely to be words than sequences that do not (compare the likelihood that “pre” is followed by “ty” to the likelihood that “ty” is followed by “bay”, as in the sequence “pretty baby”). These statistics are readily captured by young infants; 8-month-olds can discover word boundaries in fluent speech, after just 2 minutes of exposure, based solely on the statistical properties of syllable sequences (e.g., Aslin, Saffran, & Newport, 1992; Saffran, Aslin, & Newport, 1996).

Similar statistical learning abilities appear to be used for sequences of musical tones. For example, infants can discover boundaries between “tone words” by tracking the probabilities with which particular notes occur (Saffran & Griepentrog, 2001; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, 2003b). Even complex aspects of a musical system, such as the tonal hierarchy of traditional Western music, are reflected in the statistics of the input (Budge, 1943), implying that they may be available for statistical learning by humans or even by neural networks (Bharucha, 1991). These results suggest that at least some aspects of music and language may be acquired via the same learning mechanism. In some ways, this is not surprising given other facts about music and language. For example, pitch plays a critical role in many of the world's languages; these "tone languages" (such as Mandarin, Thai, and Vietnamese) use pitch contrastively, such that the same syllable, spoken with a different pitch or pitch contour, has an entirely different meaning. This use of pitch is upheld by adult users of tone languages, who are vastly more likely to maintain highly specific pitch representations for words than are their counterparts who speak nontone languages such as English (Deutsch, 2002).

Conclusions

We conclude by returning to the issue of modularity. Although we and other researchers have drawn some perhaps interesting parallels between the faculties of language and music during development, it is vital to note that there remains a good deal of neurological evidence for cortical separation of these functions in adults, and some imaging evidence as well. It has long been observed by clinicians that injury of the left temporal lobe often results in language impairment of various kinds; in contrast, such aphasia is less often seen with damage exclusively to the right temporal lobe. Complementing this literature, Isabel Peretz and colleagues (Peretz & Coltheart, 2003) have demonstrated the existence of an acquired form of amusia—a music-specific processing deficit—that is related to damage to the right, but not to the left, temporal lobe. Research with normal adults has often indicated hemispheric lateralization of these functions as well. Electro- and magnetoencephalographic data has shown canonical lateralization for early auditory discrimination of chords and syllables as measured by the MMN (Tervaniemi, 2001; Tervaniemi et al., 1999). A more detailed picture has been given by high-spatial-resolution techniques like positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), which also implicate greater involvement in musical and linguistic processing for right and left auditory cortex, respectively (Tervaniemi et al., 2000), possibly due to respective right- and left-hemisphere advantages for spectral and temporal variation (Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002). Intracerebral work with epileptic patients indicates that even the basic tonotopic organization of auditory cortex may differ in the two hemispheres (Liégeois-Chauvel, Giraud, Badier, Marquis, & Chauvel, 2003). How can we reconcile the apparent contradiction between the neurological data that suggest modularity and some of the behavior results reviewed earlier that suggest parallels between the linguistic and musical systems?

One proposal, put forth by Patel (2003), is that a distinction must be made between the processing resources used by a cognitive faculty and the content that the processing creates. According to this perspective, general auditory processing mechanisms responsible for pattern analysis are involved in the perception of both speech and music. However, the vast stores of knowledge pertaining to these separate domains may be stored in separate places in the brain. Patel argues that when neurological patients present with what appear to be domain-specific deficits in speech or music, what has actually been lost is not the processing capacity, but the knowledge required to engage it as a mature comprehender or producer. On this hypothesis, basic similarities between infant speech and music learning mechanisms would be expected. To test this hypothesis, Patel suggests a more careful examination of neuropsychological patients who present with disorders apparently specific to one domain.

Another way of looking at this controversy is that a distinction must be made between the putative modularity of mechanisms used to learn in different domains and the evidence for modularity of functions in the mature learner. Although the data supporting separate cortical regions subserving some aspects of musical and linguistic processing in adults are overwhelming, it is still quite plausible that young children may bring some of the same skills to bear on learning in each domain. The brains of young children are quite plastic and show a remarkable ability to reorganize in the event of head trauma, which suggests that, whatever the arguments for functional localization in adults, it is not fixed in children. Furthermore, differences between the brains of musicians and nonmusicians have already been demonstrated (e.g., Schlaug, 2003), and it is tempting to conclude from this that experience has a profound effect on cortical organization. However, this hypothesis requires further testing, perhaps through a systematic investigation of less experienced brains. To date, relatively few imaging studies have been done with young children, in part because of the risks associated with PET. Luckily, with the advent of less invasive techniques like fMRI, it has become possible to see whether imaging results showing modularity in adults can be replicated in children. Efforts in this direction have been aided by a recent study by Kang and colleagues, which showed that standard methodological procedures for handling adult fMRI data, such as standardizing it to a common stereotactic space, are adequate for working with child imaging data (Kang, Burgund, Lugar, Petersen, & Schlaggar, 2003). Early results suggest that some language-related functions do show age-related organizational differences that are unrelated to performance level (Schlaggar et al., 2002). However, more detailed research must be done using auditory musical and linguistic stimuli in order to better understand the modularity issue as it pertains to music and language.

The theoretical issue of modularity aside, it is also the case that metaphor plays a powerful role in directing our thinking and suggesting new insights. Whether or not music and language share common ancestry or circuitry, thinking about them as related functions may still be quite helpful in generating novel hypotheses that can help us to better understand them as separate domains. It is our hope that our review of the relevant linguistic and musical issues will help to inspire productive developmental research toward this end.[1]

References

Aslin, R., Jusczyk, P., & Pisoni, D. B. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In D. Kuhn & R. Siegler (Eds.), *Handbook of Child Psychology* (5th ed., Vol. 2, pp. 147–198). New York: Wiley.

Aslin, R., Saffran, J., & Newport, E. (1992). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321–324.

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.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13,* 17–26.

Bergeson, T., & Trehub, S. (2002). Absolute pitch and tempo in mothers' songs to infants. *Psychological Science, 13,* 72–75.

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

Blood, A., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences, 98,* 11818–11823.

Budge, H. (1943). *A study of chord frequencies: Based on the music of representative composers of the eighteenth and nineteenth centuries.* New York: Teachers' College, Columbia University.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1,* 351–353.

Cheour, M., Ceponiene, R., Leppänen, P., Alho, K., Kujala, T., Renlund, M., Fellman, V., & Näätänen, R. (2002). The auditory sensory memory trace decays rapidly in newborns. *Scandinavian Journal of Psychology, 43,* 33–39.

Cheour, M., Martynova, O., Näätänen, R., Erkkola, R., Sillanpää, M., Kero, P., Raz, A., Kaipio, M.-L., Hiltunen, J., Aaltonen, O., Savela, J., & Hämäläinen. (2002). Speech sounds learned by sleeping newborns. *Nature, 415,* 599–600.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior. Language, 35,* 26–58.

Chomsky, N. (1981). *Lectures on government and binding.* Dordrecht: Foris.

Cooper, R., & Aslin, R. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61,* 1584–1595.

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

Crowder, R., Reznick, J. S., & Rosenkrantz, S. (1991). Perception of the major/minor distinction: V. Preferences among infants. *Bulletin of the Psychonomic Society, 29,* 187–188.

Cutting, J., & Rosner, B. (1974). Categories and boundaries in speech and music. *Perception and Psychophysics, 16,* 564–570.

---

[1] Preparation of this manuscript was supported by grants from NICHD (HD37466) and NSF (BCS-9983630) to JRS, and a University of Wisconsin Graduate Fellowship and a Beinecke Brothers Memorial Fellowship to EMM. We thank Erik Thiessen for helpful comments on a previous draft.

DeCasper, A., & Fifer, W. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208,* 1174–1176.

Deutsch, D. (2002). The puzzle of absolute pitch. *Current Directions in Psychological Science, 11,* 200–204.

Eimas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171,* 303–306.

Elman, J., Bates, E., Johnson, E., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). Cambridge, MA: MIT Press.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8,* 181–195.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In J. H. Barkow & L. Cosmides (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). London: Oxford University Press.

Fishman, Y. (2001). Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology, 86,* 2761–2788.

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

Gomez, R., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11,* 194–205.

Helmholtz, H. L. F. von (1895). *On the sensations of tone as a physiological basis for the theory of music* (A. J. Ellis, Trans.) (3rd ed.). London: Longmans, Green, and Co.

Hirsh-Pasek, K., & Golinkoff, R. (1996). *The origins of grammar: Evidence from early language comprehension.* Cambridge, MA: MIT Press.

Hirsh-Pasek, K., Kemler Nelson, D., Jusczyk, P., & Cassidy, K. (1987). Clauses are perceptual units for young infants. *Cognition, 26,* 269–286.

Houston, D., & Jusczyk, P. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1570–1582.

Huron, D. (2003). Is music an evolutionary adaptation? In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 57–75). Oxford: Oxford University Press.

Ilari, B., & Polka, L. (2002). *Memory for music in infancy: The role of style and complexity.* Paper presented at the International Conference on Infant Studies, Toronto.

Jusczyk, P. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Jusczyk, P., & Hohne, E. (1997). Infants' memory for spoken words. *Science, 277,* 1984–1986.

Jusczyk, P., Rosner, B., Cutting, J., Foard, C. F., & Smith, L. B. (1977). Categorical perception of non-speech sounds by two-month-old infants. *Perception and Psychophysics, 21,* 50–54.

Jusczyk, P., Rosner, B., Reed, M., & Kennedy, L. (1989). Could temporal order differences underlie 2-month-olds' discrimination of English voicing contrasts? *Journal of the Acoustical Society of America, 85,* 1741–1749.

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,* 627–640.

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

Kang, H. C., Burgund, E. D., Lugar, H., Petersen, S., & Schlaggar, B. (2003). Comparison of functional activation foci in children and adults using a common stereotactic space. *NeuroImage, 19,* 16–28.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science.* Cambridge, MA: MIT Press.

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

Kemler Nelson, D., Jusczyk, P., Hirsh-Pasek, K., & Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16,* 55–68.

Kluender, K., Lotto, A., Holt, L., & Bloedel, S. (1998). Role of experience for language-specific functional mappings of vowel sounds. *Journal of the Acoustical Society of America, 104,* 3568–3582.

Koelsch, S., Gunter, T. C., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12,* 520–541.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch* (Vol. 17). New York: Oxford University Press.

Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science, 1,* 70–73.

Kuhl, P. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America, 66,* 1668–1679.

Kuhl, P. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6,* 263–285.

Kuhl, P., Andruski, J., Chistovich, I., & Chistovich, L. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277,* 684–686.

Kuhl, P., & Miller, J. (1975). Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science, 190,* 69–72.

Kuhl, P., & Padden, D. (1982). Enhanced determinability at the phonetic boundaries for the voicing feature in macacques. *Perception and Psychophysics, 32,* 542–550.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255,* 606–608.

Kushnerenko, E. (2003). *Maturation of the cortical auditory event-related brain potentials in infancy.* Unpublished doctoral dissertation, University of Helsinki, Helsinki.

Lerdahl, F. (2003). The sounds of poetry viewed as music. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music.* Oxford: Oxford University Press.

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

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics, 56,* 414–423.

Levitin, D. J. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception and Psychophysics, 58,* 927–935.

Liégeois-Chauvel, C., Giraud, K., Badier, J.-M., Marquis, P., & Chauvel, P. (2003). Intracerebral evoked potentials in pitch perception reveal a functional asymmetry of human auditory cortex. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 3–20). Oxford: Oxford University Press.

MacDougall-Shackleton, S., & Hulse, S. (1996). Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology, 110,* 139–146.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4,* 540–545.

Marcus, G., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283,* 77–80.

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

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* B101–B111.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29,* 143–178.

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

Narmour, E. (2000). Music expectation by cognitive rule-mapping. *Music Perception, 17,* 329–398.

Nawrot, E. S. (2003). The perception of emotional expression in music: Evidence from infants, children and adults. *Psychology of Music, 31,* 75–92.

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

Osterhout, L., & Holcomb, P. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31,* 785–806.

Osterhout, L., & Holcomb, P. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive Processes, 8,* 413–437.

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45,* 526–545.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10,* 717–733.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6,* 674–681.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition, 87,* B35–B45.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6,* 688–691.

Saffran, J. R. (2003a). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12,* 110–114.

Saffran, J. R. (2003b). Absolute pitch in infancy and adulthood: The role of tonal structure. *Developmental Science, 6,* 35–43.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., & Griepentrog, G. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology, 37,* 74–85.

Saffran, J. R., Johnson, E., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52.

Saffran, J. R., Loman, M., & Robertson, R. (2000). Infant memory for musical experiences. *Cognition, 77,* B15–B23.

Saffran, J. R., Reeck, K., Niehbur, A., & Wilson, D. (2004). Changing the tune: Absolute and relative pitch processing by adults and infants. Manuscript submitted for publication.

Saffran, J. R., & Wilson, D. (2003). From syllables to syntax: Multi-level statistical learning by 12-month-old infants. *Infancy, 4,* 273–284.

Schellenberg, E. G., Iverson, P., & McKinnon, M. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review, 6,* 641–646.

Schellenberg, E. G., & Trehub, S. (1999). Culture-general and culture-specific factors in the discrimination of melodies. *Journal of Experimental Child Psychology, 74,* 107–127.

Schellenberg, E. G., & Trehub, S. (2003). Good pitch memory is widespread. *Psychological Science, 14,* 262–266.

Schlaggar, B., Brown, T., Lugar, H., Visscher, K., Miezin, F., & Petersen, S. (2002). Functional neuroanatomical differences between adults and school-age children in the processing of single words. *Science, 296,* 1476–1479.

Schlaug, G. (2003). The brain of musicians. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music.* Oxford: Oxford University Press.

Schmidt, L., & Trainor, L. (2001). Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions. *Cognition and Emotion, 15,* 487–500.

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

Smith, L. B., Kemler Nelson, D., Grohskopf, L. A., & Appleton, T. (1994). What child is this? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition, 52,* 23–54.

Speer, J., & Meeks, P. (1988). School children's perception of pitch in music. *Psychomusicology, 5,* 49–56.

Tervaniemi, M. (2001). Musical sound processing in the human brain: Evidence from electric and magnetic recordings. In R. J. Zatorre & I. Peretz (Eds.), *The biological foundations of music* (Vol. 930, pp. 259–272). New York, NY: New York Academy of Sciences.

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *NeuroImage, 9,* 330–336.

Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping, 10,* 74–79.

Tervaniemi, M., & van Zuijen, T. L. (1999). Methodologies of brain research in cognitive musicology. *Journal of New Music Research, 28,* 200–208.

Thiessen, E., Hill, E., & Saffran, J. (2004). Infant-directed speech facilitates word segmentation. Manuscript submitted for publication.

Trainor, L. (1996). Infant preferences for infant-directed versus noninfant-directed playsongs and lullabies. *Infant Behavior and Development, 19,* 83–92.

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

Trainor, L., & Heinmiller, B. (1999). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior and Development, 21,* 77–88.

Trainor, L., & Trehub, S. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 394–402.

Trainor, L., & Trehub, S. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception and Psychophysics, 56,* 125–132.

Trainor, L., Wu, L., Tsang, C. D., & Plantinga, J. (2002). Long-term memory for music in infancy. Paper presented at the International Conference on Infant Studies, Toronto.

Trainor, L., & Zacharias, C. (1998). Infants prefer higher-pitched singing. *Infant Behavior and Development, 21,* 799–805.

Trainor, L. J., Tsang, C. D., & Cheung, V. H. W. (2002). Preference for sensory consonance in 2- and 4-month-old infants. *Music Perception, 20,* 187–194.

Trehub, S. (2003). Musical predispositions in infancy: An update. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 3–20). Oxford: Oxford University Press.

Trehub, S., Cohen, A., Thorpe, L., & Morrongiello, B. (1986). Development of the perception of musical relations: Semitone and diatonic structure. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 295–301.

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

Trehub, S., Unyk, A., & Trainor, L. (1993a). Adults identify infant-directed music across cultures. *Infant Behavior and Development, 16,* 193–211.

Trehub, S., Unyk, A., & Trainor, L. (1993b). Maternal singing in cross-cultural perspective. *Infant Behavior and Development, 16,* 285–295.

Werker, J., & Lalonde, C. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672–683.

Wilson, S., Wales, R., & Pattison, P. (1997). The representation of tonality and meter in children aged 7 to 9. *Journal of Experimental Child Psychology, 64,* 42–66.

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

Zatorre, R. J. (2003). Neural specializations for tonal processing. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 231–246). Oxford: Oxford University Press.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11,* 946–953.

Zatorre, R. J., Belin, P., & Penhune, V. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences, 6,* 37–46.

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