The study of music and emotion presents enormous challenges from a developmental perspective. For example, the dominant method for studying the perception of emotion in music requires listeners to assign emotion labels to musical excerpts, usually from a limited set of alternatives (e.g. happy, sad, angry, fearful). Obviously, this approach, which depends on mastery of the emotion lexicon, musical conventions, and emotional displays, is inappropriate for very young listeners. Our goal in the present chapter is to explore the emotional ramifications of music for those whose abilities, interpretations, and response patterns differ dramatically from those of adults, notably infants and young children. As a result, we limit our coverage to periods in which conventional emotional interpretations or responses to music are absent or incomplete—infants, primarily, and young children, secondarily. In our view, insight into the affective musical cues in the infant’s musical environment and infants’ responses to
such cues can illuminate early perceptual and learning biases that provide a foundation for subsequent musical development.

We conceive of music as a form of communication, one that originates in vocal affect. Moreover, we endorse Bachorowski and Owren’s (2003) view that the principal function of affective vocal signals is to influence listeners’ affect, attitudes, and behaviour in ways that are favourable to the signaller, such influences having shaped the signalling process over evolutionary time. We acknowledge, as they do, that some affective responses are a direct consequence of signal acoustics (e.g. amplitude, variability), with others being indirect consequences of experience (e.g. familiarity, prior associations). From this perspective, composers and performers attempt, implicitly or explicitly, to modulate listener affect by capitalizing on domain-general cues (e.g. tempo, amplitude), culture-specific cues (e.g. shared customs and history), music-specific cues (e.g. major/minor mode, harmony), or specific knowledge of the intended audience. We take issue with the view that musical performances convey discrete emotional messages that can be described by conventional emotional labels (Juslin & Laukka, 2003). Instead, we contend that the affective intentions of performers are more general and the responses more variable, reflecting listeners’ musical experiences as well as their life experiences.

In the pages that follow, we begin by describing musical elements in the infants’ environment, which arise primarily from mothers’ (or primary caregivers’) speech and singing. We outline differences in these modes of expression, their consequences for infants, and possible links to the origins of music. We also consider infants’ responsiveness to aspects of non-vocal music that have affective consequences for adults.

Our consideration of maternal speech to pre-verbal infants may seem unusual to those who view music through the narrow lens of music as practised in contemporary mainstream cultures. We aim, instead, for an approach to music and emotion that has developmental, historical, and cross-cultural resonance. The prevailing wisdom is that music making and music listening in the distant past were ubiquitous, with the forms of such music being tied to specific contexts and functions. Accordingly, we include the structured vocal tones of melodious maternal speech to avoid value judgements grounded in music as we know it today. Judgements of musical status are subjective and highly variable both across cultures (e.g. designating some music as primitive) and within cultures, as evidenced by reactions to hip-hop, post-industrial music, and the infamous 4'33” by John Cage.

After describing the music available to infants and their responsiveness to it, we summarize evidence of children’s gradual acquisition of conventional interpretations and expressions of emotion in music. Initially, children rely largely on expressive cues shared by speech and music (e.g. tempo, loudness, pitch level). Later, they add music-specific cues to their interpretive and expressive repertoire. Children’s intense interest in music and their sensitivity to its emotional connotations, even in the context of social, cognitive, or sensory disabilities, lend credence to the centrality of music in childhood.
23.2 MUSIC AND CAREGIVING

The musical experiences of infants are largely intimate ones, involving vocal behaviours on the part of the principal caregiver that are aimed at affect regulation and interpersonal attunement. In early infancy, arousal is labile, positive affect is infrequent, and the limited self-soothing repertoire (e.g. sucking) is ineffectual for high arousal levels. In general, caregivers respond sensitively, not only to alleviate distress but also to promote infant attention and positive affect (Thompson, 1994). They do so by combining stylized vocalization, touch, movement, and visual gestures in unique ways.

23.2.1 Musical conversations with infants

The ubiquity of speech to pre-verbal infants may seem odd until one reflects on the irrelevance of what is said and the significance of how it is said. The surface features of infant-directed (ID) speech contrast markedly with those of adult-directed (AD) speech in ways that make the former sound much more musical than the latter (Fernald, 1991; Trehub & Trainor, 1998). Prototypical ID speech is distinguished by its elevated pitch, increased pitch range, distinctive pitch contours, slow tempo, rhythmicity, and repetitiveness, which have been documented in numerous languages and cultures (e.g. Fernald et al, 1989). These features apply primarily to playful rather than soothing ID vocalizations, which are characterized by low pitch, falling pitch contours, and very slow tempo. Aspects of the ID speech style are also evident in women's emotionally charged utterances to familiar adults (Trainor, Austin, & Desjardins, 2000) and pets (e.g. Burnham, Kitamura, & Vollmer-Conna, 2002), highlighting acoustic correlates of the speaker's arousal level and social regulatory goals.

ID speech occurs in a pseudo-conversational context. Mothers pose frequent questions with rising pitch contours, an acoustic form that succeeds in attracting infant attention and heightening arousal (Fernald, 1991). Coos, gurgles, and smiles are treated as credible infant turns, and the conversation proceeds so long as the infant shows signs of continuing engagement. There are notable cross-cultural differences. In contrast to American mothers, who often promote infant excitement and exuberance—positive affect coupled with high arousal—Japanese mothers rely on falling pitch contours in their quest for calm, contented infants—positive affect coupled with low arousal (Morikawa, Shand, & Kosawa, 1988).

23.2.2 Unique melodies in maternal speech

Universal melodies are thought to figure prominently in spoken social regulatory messages to infants (Bryant & Barrett, 2007). In line with linguistic conventions, the term
*melody* is typically used to describe intonation or pitch contours. Musical conventions, by contrast, dictate more precise criteria, including specific intervals and rhythms.

Because verbal content is irrelevant to pre-verbal infants, it is not surprising that mothers often repeat the same stereotyped utterances on different occasions (e.g. *How's my big girl? Do you want me to change your diaper?*). Detailed comparisons of such repeated utterances (i.e. identical verbal content) reveal that they are rhythmically similar, but substantially different in their pitch level and tempo (Bergeson & Trehub, 2002). If maternal speech is largely about form rather than content, it may be more fruitful to focus on melodic repetitions. Indeed, mothers use a small set of stable melodies or interval sequences in their repeated interactions with infants, with the component intervals being unrelated to conventional musical intervals (Bergeson & Trehub, 2007). Although some common pitch contours (e.g. rising) are evident across mothers, their component intervals differ across mothers. In other words, each mother seems to use unique or *signature* tunes that are likely to promote maternal voice recognition and infant attachment. Perhaps it is not surprising, then, that infants distinguish their mother's voice from a stranger's voice only when the vocal renditions are normally intoned (Mehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978).

### 23.2.3 The impact of musical speech

Infants are in no position to state their preferences, but they can show them in experiments designed specifically for that purpose. In typical preference procedures, one auditory pattern continues to play so long as infants look at a display in one location, and a contrasting auditory pattern plays so long as they look at the same display in another location. Because infants' visual attention controls the duration of exposure to each pattern, the distribution of attention over time reflects their attention or preference for one pattern over the other.

In this manner, infants have demonstrated their preference for ID over AD speech (e.g. Fernald, 1985; Werker, Pegg, & McLeod, 1994), for approving over disapproving ID utterances (Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990), and for affectively positive AD utterances over affectively neutral ID or AD utterances (e.g. Singh, Morgan, & Best, 2002). The implication is that positive affect underlies infants' enhanced attention to ID speech. Because the ID or affectively positive speech in these experiments was higher in pitch, greater in pitch range, louder, and more variable, the preference could be attributable to greater acoustic salience. A stronger test of the preference for speech with positive affect would necessitate the separation of positive affect or other expressive cues from acoustic variability.

In the early days of life, natural speech samples seem to be necessary for a demonstrable ID preference (Cooper & Aslin, 1994). By 4 months of age, however, the isolated pitch contours of ID speech are preferred to those of AD speech (Fernald & Kuhl, 1987). As infants begin to understand the verbal content of speech, their preference for ID prosody wanes (Hayashi, Tamekawa, & Kiritani, 2001; Newman & Hussain, 2006), but does not disappear.
The characteristic vocal affect of ID speech seems to have favourable consequences on perception and memory. For example, 7-month-old infants more readily extract regularities from syllable sequences spoken with ID prosody than with AD prosody (Thiessen, Hill, & Saffran, 2005). Muted affect in the speech of depressed mothers results in diminished learning in early infancy (Kaplan, Bachorowski, Smoski, & Hudenko, 2002). Positive vocal affect also has social consequences. After 5-month-olds are exposed to audiovisual samples of two women using ID or AD speech, they prefer a photograph of the ID speaker to one of the AD speaker (Schachner & Hannon, 2008). Enhanced learning in the context of ID speech may be mediated by arousal or mood, just as preferred music enhances children's and adults' performance in various domains (Schellenberg & Hallam, 2005; Schellenberg, Nakata, Hunter, & Tamoto, 2007; Thompson, Schellenberg, & Hussain, 2001).

In short, the affectively charged utterances of maternal speech influence infants in ways that are favourable to the speaker, which is consistent with Bachorowski and Owren's (2003) functional account of vocal affect. Maternal utterances fulfill caregiving goals by capturing and maintaining infant attention or soothing and comforting infants, as required. Aside from intuitive adjustments to accommodate infants' perceptual needs and preferences, mothers generate unique speech melodies that differentiate their utterances from those of others, with the likely consequence of enhanced infant attachment.

23.2.4 Maternal songs

Singing is another means by which mothers and other caregivers across cultures regulate infant affect. For soothing and inducing sleep, they use lullabies, moving rhythmically and synchronously as they sing. In general, lullabies are characterized by slow tempo, melodic, rhythmic, and syllabic repetition, and smooth, falling pitch contours (Unyk, Trehub, Trainor, & Schellenberg, 1992). Collectively, these features generate a distinctive quality that enables adult listeners to differentiate unfamiliar foreign lullabies from non-lullabies, even when the songs are matched on culture and tempo (Trehub, Unyk, & Trainor, 1993a). Cultures that idealize calm, contented infants and maintain almost continuous physical contact with infants (Morelli, Rogoff, Oppenheim, & Goldsmith, 1992) make extensive use of lullabies (Trehub & Trainor, 1998). Those that value infant vitality and expressiveness accord priority to play songs (Trehub et al, 1997).

23.2.5 Maternal singing style

Aside from a distinctive genre of songs for infants, mothers also use a distinctive performing style, which enables naïve adult listeners to distinguish ID performances from informal, non-ID versions of the same song both within and across cultures (Trehub, Unyk, & Trainor, 1993b; Trehub et al, 1997). Greater identification accuracy for same-culture performances implies culture-specific as well as culture-general features of the
ID singing style. Listeners also distinguish genuine ID performances (i.e. those sung directly to infants) by mothers and fathers from those in which parents attempt to reproduce their performances in the absence of infants (Trehub et al., 1997). In other words, the presence of an infant listener contributes to the expressiveness of parents’ musical performances, in part, by providing an appropriate communicative context and, in part, by boosting the singers’ level of arousal or engagement. Elevated arousal also contributes to the exuberance of pre-school and young schoolchildren’s singing in the presence of their infant siblings (Trehub, Unyk, & Henderson, 1994).

In contrast to ID speech, which permits considerable flexibility in the use of intonation patterns, especially with pre-verbal listeners, ID singing generally involves well-known songs with defined pitch patterns and rhythms. As a result, maternal singers, who are typically untrained, have a limited range of expressive devices at their disposal. In general, their ID performances are higher in pitch level (1–2 semitones) and slower in tempo than their non-ID performances (Trainor, 1996; Trehub et al., 1997). Pitch level is also higher and words less clearly articulated when mothers perform a song for their infant than for their pre-school child (Bergeson & Trehub, 1999).

Maternal singing exhibits greater temporal stability than informal solo singing (i.e. no audience). Unlike solo singers, mothers forego expressive timing in favour of strict temporal regularity (Nakata & Trehub, 2008). This strategy is interesting in light of infants’ enhanced processing of temporally regular music (Trehub & Hannon, in press). Perhaps to compensate for their lack of expressive timing, mothers give special emphasis to their dynamic accents. Specifically, pitch height and loudness are highly correlated in maternal singing, but not in solo singing (Nakata & Trehub, 2008).

ID singing is clearly distinguishable from non-ID singing, but the differences are much more modest than those between ID and AD speech. For example, ID speech is approximately 4–5 semitones higher than AD speech (Fernald, 1991), but ID singing is 1–2 semitones higher than non-ID singing (Trainor, 1996). Moreover, the dynamic range of ID speech, from whisper to shriek, greatly exceeds that of typical AD speech. In the case of singing, dynamic changes in ID renditions are more gradual than those in non-ID renditions (Nakata & Trehub, 2008). In many respects, then, ID singing is more predictable and less acoustically salient than ID speech.

Perhaps the most distinguishing feature of ID from AD singing—but the most elusive from a measurement perspective—is its vocal timbre, which conveys the singer’s intense emotional engagement with the infant audience. Mothers’ tendency to smile while singing to infants alters the shape of their vocal tract and the resulting vocal quality (Tartter & Braun, 1994). When naïve listeners attempt to distinguish ID from non-ID samples of singing, they often justify their judgements of ID singing on the basis of its ‘smiling sound’, ‘soft voice’, ‘sense of involvement’, or ‘warm voice’ (Trehub et al., 1997). These timbral differences may arise from micro-variations in frequency and amplitude that prevail in ID singing (Trainor, Clark, Huntley, & Adams, 1997).
23.2.6 Impact of maternal singing

Infants 6 months of age exhibit clear preferences for ID singing over non-ID singing (Trainor, 1996; see also Figure 23.1). In principle, this preference could be attributable to familiarity with the ID singing style, but comparable preferences have been demonstrated in hearing newborns with deaf, non-verbal parents (Masataka, 1999). The presumption is that elevated pitch level contributes to this preference (Trainor & Zacharias, 1998), but vocal timbre may be the critical feature. High pitch (relative to the singer’s usual pitch level) may generate a vocal timbre that is suitable for arousing but not for soothing ID songs. For example, infants prefer lower-pitched versions of expressively sung lullabies (Volkova, Trehub, & Schellenberg, 2006), perhaps because such performances are consistent with the singer’s soothing intentions.

Comparisons across studies reveal more sustained attention to vocal music (Trainor & Zacharias, 1998; Volkova et al, 2006) than to instrumental music (Plantinga & Trainor, 2005; Saffran, Loman, & Robertson, 2000). Greater engagement at the time of initial exposure to music should result in richer and more enduring memory for that music. Thus, although infants exhibit long-term memory for the tunes (i.e. relative pitch patterns) of synthesized folk melodies (Trainor, Wu, & Tsang, 2004), they fail to remember their pitch level (Plantinga & Trainor, 2005), as they do for expressively sung lullabies (Volkova et al, 2006). These findings confirm the importance of the voice (Vouloumanos & Werker, 2004, 2007) and of ecologically valid musical material for infants.

For the most part, the effects of ID music (and speech) have been examined with recorded auditory stimuli from unfamiliar women (i.e. the mothers of other infants). This contrasts with infants’ usual experience of music as a familiar, multimodal stimulus. In fact, live maternal singing has important consequences for infant arousal (Shenfield, Trehub, & Nakata, 2003). The same type of singing can attenuate arousal levels for infants with higher initial levels and elevate arousal for infants with lower initial levels, raising the possibility that maternal singing optimizes infant arousal.

23.2.7 Relative impact of speech and singing

In contrast to repeated ID utterances, which typically differ in pitch level and tempo, ID songs are performed on different occasions at nearly identical pitch level and tempo (Bergeson & Trehub, 2002). In addition to specific ID songs being more stable than specific ID utterances in terms of timing, pitch patterning, and dynamics, the sound patterns of music in general are much more predictable than are those of speech. If acoustic salience underlies auditory preferences in early life, then infants should prefer ID speech to ID singing. Instead, they exhibit considerably greater attentiveness and engagement to audiovisual episodes of maternal singing than to comparable episodes of maternal speech (Nakata & Trehub, 2004). Infants’ differential attentiveness to maternal singing is consistent with its suggested role as an arousal optimizer (Shenfield et al, 2003).
Fig. 23.1 Four photographs of an infant that indicate a gradation of emotional expressions from neutral to full smile during maternal singing.
23.2.8 Keeping in touch with infants

Just as Bachorowski and Owren (2003) have argued that selective pressures favoured vocal signals that influenced listener affect, Falk (2004) has argued for comparable selective pressures on ancestral mothers. She contends that productive activities such as foraging required mothers to release babies from their cradling arms, putting a premium on vocal signals that could maintain infant contentment in the absence of body contact or even visual contact.

Although maternal touch is a highly effective modulator of arousal (Montagu, 1986), its incidence is higher in Eastern cultures that make less use of ID speech than Western cultures do (Morikawa et al, 1988). This raises the possibility that different modes of maternal behaviour (e.g. ID speech, touch) could substitute for one another in managing infant arousal or affect. When mothers are briefly restricted from touching their 6-month-old infants, their ID speech becomes more expressive, as reflected in its elevated pitch and increased pitch range (Nakata & Trehub, 2002). This apparent compensation for restricted touch seems necessary for maintaining infant attention during maternal speech episodes. By contrast, mothers’ singing style remains unchanged during bouts of restricted touch, yet infants’ attention is comparable whether maternal touch is present or absent. The implication is that singing requires no special enhancement for situations in which caregivers are at a distance, perhaps even out of sight. These findings are consistent with Falk’s (2004) claims of expressive maternal vocalization as a means of ‘keeping in touch’ with infants. It is fair to say, then, that infants are touched by expressive ID speech and singing.

23.3 CROSS-MODAL CORRESPONDENCES IN SPEECH AND SINGING

We know that adults automatically encode identity-specific cues from speech, and there are indications that infants do so as well (Houston & Jusczyk, 2003). For adults, individual differences in expressive timing (Lander, Hill, Kamachi, & Vatikiotis-Bateson, 2007) and articulatory style (Lachs & Pisoni, 2004; Rosenblum, Smith, Nichols, Hale, & Lee, 2006) provide cross-modal cues to the identity of unfamiliar speakers. When adults are exposed to a scripted utterance followed by dynamic but silent visual displays of two unfamiliar speakers, they succeed in matching the correct video to the previously heard speaker (Kamachi, Hill, Lander, & Vatikiotis-Bateson, 2003), but their level of accuracy is modest. Accuracy is greater when the samples are drawn from natural ID speech (Trehub & Bricic, 2008), which attests to the individuality of expressive cues in conversational speech (Lander et al, 2007).
Remarkably, 6-month-old infants succeed in linking silent videos of unfamiliar women to previously heard samples of ID speech (Trehub, Plantiga, & Brcic, 2009). They fail to match person-specific cues across modalities when the samples are drawn from different songs, succeeding, however, with different excerpts from the same song (Trehub et al, 2009). This finding may reflect mothers’ use of different songs for different expressive intentions or the absence of a unique singing style in untrained singers. Nevertheless, it is clear that infants and adults automatically extract cues to the identity of unfamiliar speakers and singers—in some situations, at least—and that such identity cues are bimodal or multimodal.

23.4 Speculations about Beginnings

It is possible that the precursors of music were affective vocal improvisations used in interpersonal contexts for social regulatory goals. We offer the highly speculative proposal that some of these affective communications became conventionalized and eventually elaborated into a code with discrete pitches and predictable timing that made synchronous activity possible. Although music continued to serve social regulatory goals, its reach extended from dyadic contexts to group contexts. With respect to language and its origins, there are speculations that multimodal, musical phrases, initially with holistic meanings, were subsequently subdivided into smaller meaningful units and combined, over time, in increasingly complex ways (Mithen, 2005; Wray, 1998, 2002). Some aspects of the ancestral vocal style may have been retained to serve as carriers of informational messages and, ultimately, spoken language.

Just as attachment to the primary caregiver provides the impetus for social-emotional development (Bowlby, 1968), affective responses to music may originate in infants’ reactions to the vocalizations of caregivers. As noted, infants exhibit enhanced attention and affectively appropriate responses to maternal speech and song. To some extent, comparable cues drive affective reactions to spoken and sung performances, but song is more effective at comforting infants and sustaining their attention, perhaps because it combines positive emotional tone along with patterns of pitch and timing that have stood the test of time.

Elements from affective vocalizations, whether sung or spoken, predominate in the infant’s musical environment. Most infants also experience some incidental exposure to non-vocal music, either by overhearing music intended for their parents or by parents’ deliberate use of children’s recordings. Exposure to recorded performances increases progressively during childhood. It is therefore useful to consider affective responses to non-vocal music in the early years, with the goal of identifying aspects of musical structure that trigger unlearned responses and others that depend on enculturation.
23.5 Responses to non-vocal music

Most studies of musical emotion have been restricted to Western listeners and to Western musical materials. As a result, it is difficult to determine whether affective responses to music stem from universal reactions to acoustic patterns, music-specific knowledge, or extra-musical associations (i.e., links between specific musical structures and emotional experiences). There are suggestions, however, that pan-cultural interpretations of emotion in music are cued by specific acoustic features (e.g., tempo, loudness) and by subjective judgments of complexity (e.g., negative valence associated with complexity ratings). North American listeners use such cues to discern the intended emotions of joy, sadness, and anger in Hindustani musical passages (Balkwill & Thompson, 1999). Japanese listeners also apply those labels comparably to Hindustani, Western, and Japanese musical passages (Balkwill, Thompson, & Matsunaga, 2004; see also Chapter 27, this volume). Because the rich musical stimuli in these studies differ along multiple dimensions, a simple assignment of cues to specific emotion judgments is impossible. Moreover, the ubiquity of Western music raises the possibility that Western expressive devices influence musical performances in other cultures. Developmental research provides a unique opportunity to examine the responsiveness of listeners whose musical enculturation is in progress.

23.5.1 Consonance matters

Simultaneous or harmonic combinations of tones are considered consonant if their pitches blend smoothly to yield a pleasant sound. Such intervals are considered dissonant if their component pitches blend poorly, leading to fluctuations in amplitude that sound unpleasant to human listeners across cultures (Butler & Dastan, 1968), but not to non-human primates (McDermott & Hauser, 2004). Although scholars consider the perceived consonance or dissonance of sequential or melodic intervals to result from musical exposure (e.g., Blacking, 1992), the wide cross-cultural distribution of the most consonant melodic intervals (Sachs, 1943), notably the octave, perfect fifth and perfect fourth, is unlikely to be coincidental. In other words, unlearned perceptual biases may have influenced the selection of melodic intervals across cultures.

The study of naïve listeners can address questions about the biological or cultural basis for consonance and dissonance and their associations with positive or negative valence. Infants (Schellenberg & Trehub, 1996a; Trainor, 1997) and young children (Schellenberg & Trehub, 1996a) show processing biases for consonant intervals or tone combinations. Specifically, they find consonant intervals more distinctive and more memorable than dissonant intervals. For example, they perceive the similarity of tones an octave apart (Dehany & Armand, 1984), and they show enhanced processing of melodies with prominent perfect fifths (e.g., Cohen, Thorpe, & Trehub, 1987; Trainor & Trehub, 1993).
Like adults, infants show distinct preferences for consonant over dissonant patterns in the newborn period and thereafter (Masataka, 2006; Trainor & Heinmiller, 1998; Trainor, Tsang, & Cheung, 2002; Zentner & Kagan, 1996). Because the consonance/dissonance comparisons in the preference studies involved isolated harmonic intervals or harmonic intervals embedded in a musical piece, it is unclear whether infants would show comparable preferences for unaccompanied melodies with consonant component intervals. Such preferences may be restricted to non-vocal contexts. The pleasing harmonic quality of the human voice may override potentially negative responses to dissonant melodic intervals articulated in speech or song.

### 23.5.2 Timing matters

Intrinsic constraints may also influence responses to temporal aspects of music such as rhythm, metre, and tempo. Unpredictable or irregular rhythmic contexts compromise discrimination and memory in infants (Bergeson & Trehub, 2006; Hannon & Trehub, 2005; Trehub & Hannon, 2009) as well as adults (e.g. Jones, Johnston, & Puente, 2006). Temporal regularity may also have affective consequences. There are suggestions, for example, that infants prefer regular to irregular auditory sequences (Nakata & Mitani, 2005). For adults, unpredictable sound sequences can have unfavourable consequences on arousal (Herry et al, 2007), but expressive variations in timing may have more subtle and more favourable effects (Nair, Large, Steinberg, & Kelso, 2002).

### 23.6 Acquiring labels for musical emotions

The ability to label musical emotions obviously depends on mastery of the emotion lexicon and linking appropriate labels to the intended musical emotions. To study age-related changes in this ability, children are typically required to judge the emotional intentions of musical samples by selecting a depicted facial expression or verbal label from two or more alternatives (see Chapter 8, this volume, for examples). The results across several studies converge in suggesting that children as young as 4 identify expressive intentions in music such as happiness and sadness, but they confuse fear and anger (Dolgin & Adelson, 1990; Terwogt & van Grinsven, 1991).

There is limited information about the cues that underlie age-related changes in children's judgements. Adults reliably interpret a range of emotional intentions in music on the basis of tempo, loudness, pitch, mode (major/minor), and consonance/dissonance (Chapters 14 and 17, this volume). In forced-choice tasks, Western adults judge music in the major mode and with rapid tempo as happy and music in the minor
mode and with slow tempo as sad (e.g. Hevner, 1935; Peretz, Gagnon, & Bouchard, 1998). Mode differs in its usage across cultures and historical periods and is music-specific. By contrast, tempo is linked to arousal in speech and music across cultures. Accordingly, one would expect emotional interpretations based on tempo (fast or slow) to appear well before those based on modal contrast.

Most developmental studies of emotion perception in music have used orchestral excerpts from the classical repertoire (e.g. Cunningham & Sterling, 1988; Esposito & Serio, 2007; Nawrot, 2003), precluding the identification of specific cues underlying children's judgements. When tempo and loudness are held constant, children 8 and older associate the major mode with positive emotions and the minor mode with negative emotions (Gerardi & Gerken, 1995; Gregory, Worrall, & Sarge, 1996). When tempo and mode are varied systematically, 6- to 8-year-olds make conventional judgements of music as happy or sad on the basis of mode or tempo, 5-year-olds do so only on the basis of tempo, and 4-year-olds perform at chance levels on both dimensions (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001).

Some insight into children's understanding of musical cues to positive and negative affect is provided by their own expressive performances. When 4- to 12-year-old children sing a well-known song (‘Twinkle, Twinkle, Little Star’), their ‘happy’ renditions of the song are faster, louder, and at a higher pitch level than are their ‘sad’ portrayals (Adachi & Trehub, 1998). For the younger children, it is not at all clear that their contrastive performances arise from an understanding of the relevant expressive cues. In fact, requests for sad portrayals of songs often lead to confusion, hesitation, and highly unconventional performances. Nevertheless, child listeners within and across cultures successfully decode the intended emotion from performances of same-age children, and they do so more accurately than adults (Adachi & Trehub, 2000, Adachi, Trehub, & Abe, 2004).

In sung as opposed to instrumental music, words provide information about the singer's affective intentions. In most cases, the affective implications of the lyrics and performing style are consistent, making it difficult to disentangle their separate contributions to children's emotional interpretations. When 5- to 10-year-old children listen to novel excerpts sung in a stereotypically happy manner (fast tempo, major key) or sad manner (slow tempo, relative minor key), they judge the singer's feelings as happy when the lyrics depict positive events (e.g. ‘Dad gave me a new bike for my birthday’) and as sad when the lyrics depict negative events (‘I lost all my money on the way to the store’) (Morton & Trehub, 2007). Adults, by contrast, judge the singer's feelings on the basis of her expressive style. When the same tunes are sung with nonsense syllables rather than words, children's judgements match those of adults, confirming their ability to decode the non-verbal, expressive cues but underlining the dominance of verbal cues in their affective interpretations of songs.

In short, the ability to label discrete emotions in various types of music is evident in childhood, although the course of development of cue utilization remains unclear. Young children rely primarily on tempo and loudness when judging the intended emotion in musical performances (Adachi & Trehub, 2000; Adachi et al, 2004). Older children use pitch level, mode, and other cues arising from musical enculturation.
Multiple sources of information are available for the identification of emotions in music, but the relative weighting of cues is likely to change as children acquire greater understanding of the emotional implications of various musical structures. It remains to be determined, however, whether children’s judgements of emotion in open-ended contexts would be similar to those in forced-choice contexts.

23.7 Expectations

Music theorists consider the internal or deep structure of music, which generates expectations in listeners, as the principal source of emotional responses to music (Huron, 2006; Meyer, 1956; Narmour, 1990; Chapter 21, this volume). Musical expectations refer to the predictions that listeners implicitly generate about upcoming musical events, along with the confirmation or violation of such predictions over the course of listening. Violations of expectation are thought to generate affective reactions such as surprise, anxiety, or tension (i.e. the need for continuation). Confirmations of expectation are thought to generate resolution, repose, or relief.

Some expectations are thought to arise from universal principles of perceptual organization, but others stem from implicit knowledge of style-specific principles (Krumhansl, 2002; Narmour, 1990), including mental representations of the tonal hierarchy (Krumhansl, 1990). Implicit knowledge of the tonal hierarchy leads some pitches (e.g. tonic or reference tone, dominant or perfect fifth above the tonic) to sound stable and others (e.g. leading tone, tones outside the key) to sound unstable or incomplete. Violations of tonal expectation lead to changes in subjective judgements of emotion, physiological responses, and neural responses (Krumhansl, 1997; Sloboda, 1991; Steinbeis, Koelsch, & Sloboda, 2006).

Culture-specific expectations are necessarily shaped by musical exposure. In recent years, there has been increasing interest in mechanisms that underlie the acquisition of implicit knowledge about regularities in speech or music. Remarkably, after limited exposure to a sequence of tones, 7-month-old infants extract regularities based on the conditional probabilities between tones (Saffran, Johnson, Aslin, & Newport, 1999). In principle, infants could also learn about the complex hierarchical structure of Western tonal music, including the tonal hierarchy, from statistical regularities in Western musical pieces (Huron, 2006). Adults with little or no musical training exhibit implicit knowledge of this structure, which attests to the impact of incidental exposure (Tillman, Bharucha, & Bigand, 2000). The developmental timetable for emerging tonal knowledge and its consequences for responding affectively to music are of particular interest.

Although infants show enhanced learning in the context of consonant intervals (e.g. Schellenberg & Trehub, 1996b) and temporal regularity (e.g. Trehub & Hannon, 2009), which are common across cultures, no comparable advantages are evident for
musical patterns that conform to Western tonal conventions (Lynch, Eilers, Oller, & Urbano, 1990; Trainor & Trehub, 1992; Schellenberg & Trehub, 1999). By 5 years of age, children show implicit knowledge of the component tones of a key (Trainor & Trehub, 1994), and by 7, they are sensitive to tonal stability and basic harmonic functions (Krumhansl & Keil, 1982; Schellenberg, Bigand, Poulin-Charronnat, Garnier, & Stevens, 2005; Speer & Meeks, 1988; Wilson, Wales, & Pattison, 1997).

Accordingly, children in the early school years could respond affectively to violations of expectation, as adults are presumed to do (Huron, 2006; Meyer, 1956). To date, however, no research has addressed this question. Even in the case of adults, it is unclear whether the tensions and resolutions that listeners perceive while listening to music result in distinctly affective experiences like those that arise from performance features (e.g. tempo, dynamics, timbre) or from extra-musical associations (e.g. Mendelssohn's Wedding March, national anthem). Perhaps the internal structure of music has its greatest affective impact on musically trained or highly experienced listeners who apprehend its aesthetic richness.

### 23.8 Developmental Disorders

Research on children with developmental disorders has the potential to shed light on engagement with music in the face of severe intellectual, perceptual, or social challenges. Autism is defined principally by deficits in social interaction, communication, and cognitive flexibility. Aside from isolated cases of autistic musical savants (e.g. Treffert, 1989; Young & Nettelbeck, 1995), autistic individuals, on average, have more accurate pitch perception (Heaton, 2005; Järvinen-Pasly & Heaton, 2007) and pitch memory (Heaton, 2003; Heaton, Hermelin, & Pring, 1998) than the general population. Given their pronounced social deficits, one might expect delays or difficulty in recognizing the affective connotations of music. Nevertheless, autistic children 7 to 15 years of age identify conventional affective connotations of melodies in the major and minor mode much like their non-autistic peers (Heaton, Hermelin, & Pring, 1999). In stark contrast to the diminished sociability of autistic children, individuals with Williams syndrome, who have pronounced cognitive and spatial deficits, exhibit enhanced sociability. These children show greater engagement in music and a greater range of emotional responses to music than do normally developing children (Don, Schellenberg, & Rourke, 1999; Levitin et al, 2004).

Deaf listeners with cochlear implants are another population of interest. Their prostheses provide degraded pitch resolution that is adequate for the perception of speech in quiet environments but woefully inadequate for the transmission of musical pitch patterns. As a result, postlingually deafened adults tend to find familiar music unpleasant or unrecognizable in the post-implant period (Lazaletta et al, 2007; Leal et al, 2003). By contrast, most congenitally deaf children who receive their implants by 2 or 3 years
of age find music highly engaging. They sing familiar songs with vim and vigour, preserving the timing but not the pitch contours of the songs (Nakata, Trehub, Mitani, & Kanda, 2006). Many child implant users listen regularly to pop music, which they recognize with or without the vocals but not from the melody alone (Vongpaisal, Trehub, & Schellenberg, 2006). They also recognize songs that accompany their favourite television programmes (Mitani et al, 2007). These child implant users successfully identify the valence (happy or sad) of musical pieces (piano renditions with multiple cues) from the Western classical repertoire (Hopyan-Misakyan, Gordon, Papsin, & Dennis, 2006), but the cues underlying their judgements have not been identified.

23.9 Conclusion

A number of scholars contend that musical emotions can be best understood as a process of sending and receiving discrete emotional messages (Juslin & Laukka, 2003). Although we concur with the perspective of music as communication, we dispute the notion that discrete emotional messages are involved. Instead, we view senders of musical messages as attempting to influence the affective state or actions of receivers, whether those receivers are infants, children, or adults. On the receiving end, we regard interest as the most basic and most important response to music, both for naïve listeners and for expert listeners. In the early years, intense interest in music motivates learning in this domain. For mature listeners, gradations of interest remain at the core of aesthetic emotions (Silvia, 2007).

Our contention is that affective responses to music arise within a social regulatory system in which caregivers use nonverbal vocal communication to modulate infants' arousal and engagement. Considerable social and musical enculturation is necessary before young listeners understand the conventions relating discrete emotional labels to music. Undoubtedly, the sound patterns of music give rise to a rich variety of affective experiences, with discrete emotional categories being among the many ways in which these experiences are described.

Although children eventually master the ability to interpret and express discrete emotions in music, they initially experience confusion about the cues that are relevant to specific emotional categories (Adachi & Trehub, 1998; Dalla Bella et al, 2001; Morton & Trehub, 2007). This confusion may reflect a broader problem, namely, the assumption that conventional emotion labels for music capture the essence of our emotional experiences. When adults are given a wide array of descriptors, they rarely select negative adjectives such as sad, angry, and fearful to describe their emotional responses to musical excerpts, opting instead for positive adjectives such as relaxed, happy, dreamy, enchanted, nostalgic, and touched (Juslin & Laukka, 2004, Table 4; Zentner, Grandjean, & Scherer, 2008, Table 2). Nevertheless, the same respondents select negative adjectives
to denote emotions expressed by the music, which implies a fundamental disconnect between affective reactions to music, on the one hand, and labelling expressive intentions, on the other.

Obviously, listeners are often moved by music, but there is no compelling evidence that they are moved to discrete states of happiness, sadness, anger, or fear. Instead, being moved or touched by music may involve changes in affect arising, in part, from unlearned responses to acoustic cues, which would generate convergence across listeners and, in part, from listeners’ personal and musical history, which should generate divergence. The indeterminate meaning of music, which gives listeners free rein to conjure up emotionally significant memories and fantasies, is likely to lead to further divergence.

Why, then, do listeners apply conventional emotion labels to music in typical experimental contexts? In general, they have few options because of the forced-choice format of responding. Even if they could respond freely, they might not choose to share their thoughts and feelings (e.g., imagining an amorous episode, wanting to escape from the laboratory without losing course credit or payment). It is also unlikely that experiences evoked by specific musical pieces in contextually appropriate settings would be elicited in the laboratory. Categorical labelling may result primarily from the demand characteristics of experiments on musical emotions. For situations in which listeners are genuinely moved by music, their feelings may be difficult, perhaps impossible, to describe in words. The default option is to use labels that are pervasive in everyday discourse about emotion.

In sum, we propose that musical emotions are best understood in the broad context of communication, with sound being used to influence the behaviour of others. Emotional responses arise, in part, from acoustic features such as amplitude, pitch variability, and temporal predictability and, in part, from domain-specific conventions and extra-musical associations. Exposure to emotionally meaningful music begins in the early days of life with the melodious speech and soft singing of the primary caregiver. Presumably, those sound patterns promote feelings of comfort, which are reinforced by the security of the dyadic context. They also enhance attention and arousal in ways that foster learning and strengthen interpersonal bonds. Such early responses to music may provide a foundation for subsequent emotional responses to music, which become more refined or differentiated over the course of development. Even in maturity, responses to music retain their basic social function of promoting emotional regulation, including self-regulation, and connections with others.¹

¹ Preparation of this article was assisted by funding from the Social Sciences and Humanities Research Council and the Natural Sciences and Engineering Research Council of Canada.
Recommended further reading


References


