CHAPTER 27

# CROSS-CULTURAL SIMILARITIES AND DIFFERENCES

WILLIAM FORDE THOMPSON AND LAURA-LEE BALKWILL

Over the past century, research on the relation between music and emotion has overwhelmingly focused on the perceptions and experiences of Western listeners in response to Western tonal music. This focus has largely been a pragmatic consequence of the challenges of carrying out research in non-Western contexts, but it has left many important questions unanswered. To what extent are emotional aspects of music similar across cultures? Are there general principles that might account for the connection between music and emotion in all (or most) cultures, or is that connection unique to each culture, and perhaps non-existent in some cultures? How might we conduct a cross-cultural study of emotion and music without the conclusions being corrupted by cultural biases?

In this chapter, we review empirical studies of music and emotion that involve a cross-cultural comparison, and we outline prevailing views on the implications of such studies. Cross-cultural investigations can provide important insights into the contribution of universal and cultural associations between music and emotion. They can also be used to validate psychological theories of music and emotion—which are overwhelmingly supported by research that privileges Western tonal music.

Reflecting the current state of research in this area, our focus is on the *perception* of emotion in music (decoding) rather than the induction of affective states by music. The question of whether individuals from different cultures have comparable affective experiences in response to music is largely unknown, but is a rich area for future research (see Chapter 6, this volume). We also discuss cross-cultural studies of emotional prosody. *Speech prosody* refers to the vocal qualities of speech and includes intonation (pitch variation), stress, and timing. It signals points of emphasis, indicates a statement or question, and conveys emotional connotations (Darwin, 1872; Frick, 1985; Juslin & Laukka, 2003). Prosodic communication of emotion can occur independently of verbal comprehension (Kitayama & Ishii, 2002).

We discuss research on speech prosody for two reasons. First, music and speech prosody share important acoustic attributes and may use a common 'code' for emotional communication (Juslin & Laukka, 2003). Second, current theory and evidence on cross-cultural emotional decoding is more advanced for prosodic materials than for musical materials and, as such, may provide a model for future cross-cultural research on music and emotion. We do not provide a detailed examination of ethnographic studies of specific musical traditions, as such discussions may be found in Chapters 6 and 7 (this volume).

We begin by discussing some theoretical implications of research on cross-cultural commonalities in the association between music and emotion. We note that cognitive, ethnomusicological, and sociological approaches complement each other by providing different levels of explanation and different perspectives on the concept of music. Section 27.2 reviews the central questions arising from cross-cultural research on emotion. In Section 27.3, we outline the *cue-redundancy model*, developed to account for cross-cultural similarities and differences in the expression and recognition of emotion in music. The model accounts for the balance of culture-transcendent and culture-specific emotional cues across musical genres. It can be used to frame research questions, and to communicate, compare, and integrate empirical findings.

Section 27.4 outlines a broader framework for summarizing existing data on emotional communication, referred to as *fractionating emotional systems* (FES). FES extends the cue-redundancy model by accounting for similarities and differences in emotional communication not only across cultures but also across the auditory channels of music and speech prosody. FES also accounts for the process of enculturation that permits the gradual division of musical and prosodic emotional coding systems as well as the emergence of distinctive systems across cultures. Section 27.5 provides a review of cross-cultural studies of music and emotion, while section 27.6 reviews cross-cultural studies of emotion in speech. Section 27.7 identifies some future prospects for the cross-cultural study of music and emotion.

## 27.1 Music as a cross-cultural construct

In order to investigate music as a cross-cultural construct, one must confront the tasks of defining *music* on the one hand and *culture* on the other. Merriam (1964) characterized music as having three aspects: sound, behaviour, and concept. As sound, music can be defined as a class of auditory signals that are produced by performers and perceived by listeners. As behaviour, music is associated with tangible activities (e.g. performance, dance, ritual) that are often essential to music experience, and that can be subjected to rigorous psychological, social, and historical analyses. As concept, music is construed as having specific functions within any social group (Clayton, 2001; Cross, 2006; Dissanayake, 2001). To a large extent, cross-cultural music cognition has focused on music as sound, but there is increasing awareness that music is a multimodal phenomenon. The behavioural aspect of music cannot be treated as a distinct level of analysis, but is inseparable from perceptions and experiences of music. Indeed, the visual input from viewing the facial expressions and gestures of a music performer can profoundly influence a listener's emotional responses to music (Thompson, Graham, & Russo, 2005; Thompson, Russo, & Quinto, 2007).

The construct of *culture* has also been scrutinized and debated. It refers to the set of behaviours, beliefs, social structures, and technologies of a population that are passed down from generation to generation. It includes social conventions related to art, dress, manner, dance, music, religion, ritual, and morality. Like music, cultures can be examined on multiple levels, some of which are nested within others. For example, within any cultural environment (e.g. Northern India) there are many subcultures (Hebdige, 1979). Each subculture, in turn, is associated with its own distinctive *habitus*: the habits, beliefs, skills, schemas, and preferences that are tacitly acquired and that shape perceptions, behaviours, and experiences (see Chapter 6, this volume). The term culture is not equivalent to 'country' or 'continent', and is not always associated with any one geographical region. Moreover, most individuals do not 'belong' to a single culture. More typically, numerous cultural and sub-cultural influences can be detected for any one individual, colliding, merging, and making temporary appearances when the appropriate context arises. Thus, although cross-cultural studies have often involved an examination of music materials and practices within specific geographical regions (e.g. Europe, Japan, Northern India), it should be recognized that the construct of culture does not restrict studies to such strategies.

Most ethnomusicologists are sensitive to such challenges, and are dubious of attempts to characterize nations as singular cultures and compare them with one another. In order to appreciate this disciplinary perspective, it may be useful to note that 'ethnomusicology' was known as 'comparative musicology' at the turn of the twentieth century. This name was abandoned for three reasons. First, the 'comparative method' as practised at the beginning of the twentieth century was steeped in an implied cultural hierarchy. Many scholars at the time regarded Western art music as a cultural pinnacle, and they interpreted music from other cultures as earlier or more primitive stages of development in the evolution of music. Like the common distinctions between 'hunter-gatherer', 'agrarian society', and 'nation state', comparative musicologists at the time were preoccupied with characterizing the 'stage' of musical development for each culture. Second, any comparison involves the establishment of criteria for evaluating musical practices, and the question of which criteria to adopt raises thorny methodological issues. How does one compare cultures objectively without inadvertently using criteria that privilege the culture of the researcher? Most contemporary ethnomusicologists have concluded that there are no neutral criteria. Third, ethnomusicologists resolved that each culture should be understood on its own terms. The very act of comparing musical cultures is therefore suspect. The change in name from comparative musicology to ethnomusicology reflected these concerns.

Ethnomusicologists often point to the work of Alexander Ellis as a seminal event in the birth of their discipline ('On the musical scales of various nations'; Ellis, 1885). At the time Ellis was working, many scholars believed that different musical cultures were progressing along a developmental path that approached Western European art music. Evidence of this could purportedly be found in the instruments of other cultures whose tuning was considered crude approximations of the more developed Western scale. The 'crude tuning' of these instruments was regarded as tacit evidence of the rudimentary listening skills of people in other cultures. Working at the British Museum, Ellis measured the tuning of instruments from different cultures, and observed that the varied tunings in other cultures were not merely poor approximations of Western tuning. Rather, he found that different instruments pointed to unique and stable tuning systems that were categorically 'different' from the Western scale. The instruments were not merely technologically naïve efforts along some path toward the Western scale, but sophisticated technological efforts in their own right. Ellis's results thereby conflicted with notions of a cultural hierarchy.

Although the abandonment of notions of cultural hierarchies was a welcome development, the abandonment of 'comparative' approaches altogether by ethnomusicology has been unfortunate. A full understanding of the cognitive basis of music is not possible unless similarities and differences across cultures in the perception, experience, and production of music are taken into account. Comparisons across cultures are difficult methodologically and susceptible to researcher bias, but renouncing all comparative research is hardly a productive response to such challenges. Ellis's own achievements depended on comparative measurements within and between cultures, and in recent years researchers have employed reciprocal or counterbalanced methods in which test materials and subjects are recruited from two or more cultures. These methods may not eliminate cultural bias entirely, but the empirical data provide useful grist for theorizing about the cognitive basis of music.

There are several approaches to the cross-cultural study of music, including ethnomusicological, anthropological, sociological, and cognitive. Among the most influential sociological approaches is that of Alan Lomax (1975, 1976, 1978). His method, called *cantometrics*, involved characterizing song styles from different cultures using rating scales for features such as intensity, tempo, rhythmic complexity, interval width, embellishment, register, and tension. The procedure allowed Lomax to analyse and compare song styles cross-culturally. This research led him to conclude that the emotional properties of music are central to understanding cross-cultural similarities and differences in musical behaviours: 'music somehow expresses emotion; therefore, when a distinctive and consistent musical style lives in a culture or runs through several cultures, one can posit the existence of a distinctive set of emotional needs or drives that are somehow satisfied or evoked by this music' (Lomax, 1962, p. 425). Put simply, song style mirrors and reinforces cultural style.

*Cross-cultural music cognition* may be defined as the exploration of similarities and differences in cognitive processes and emotional experiences for music across cultures, and can be used to differentiate universal and culture-specific determinants of mature forms of music understanding. In general, cross-cultural music cognition has a more restricted focus than ethnomusicology, which applies research strategies adapted from cultural anthropology, sociology, and other disciplines in order to understand specific musical systems and traditions within a social or cultural context (see Chapter 6, this volume).

All cross-cultural research has a specific set of challenges above and beyond those that are part of any research endeavour. The difficulties of identifying geographical correlates of culture, of isolating one culture from other cultural influences, and the potential for researcher bias have already been noted. How effectively a study addresses such pitfalls and challenges is an important factor in interpreting its findings. Of critical importance in cross-cultural research is an awareness of one's own cultural perspectives and how they can bias every facet of the research, from the question being asked, to the methodology employed, to the analysis strategy employed, to the interpretation of observations. Consulting with members of the cultures under investigation is one important way of increasing an awareness of one's biases.

Perhaps the greatest challenge to studies of the expression and recognition of emotion across cultures is establishing cross-cultural equivalence of conceptual and empirical variables. Much has been written about the use of emotion labels to conceptualize the communication of emotions (see Wierzbeca, 1992, for a review). Typically, researchers employ the strategy of back-translation to create an equivalent set of instructional materials in the language of each culture of interest. However, even when linguistic equivalence has been achieved, conceptual equivalence may still prove elusive. For example, the word for anger in English and the word for anger in Japanese may not be interpreted by members of these cultures with the same degree of intensity or as signifying the same situations or behaviours. Therefore, establishing equivalence must be handled with care and sensitivity to cultural mores.

Ethnomusicologists and anthropologists have drawn attention to many nuances of difference in musical understandings across cultures, but evidence from cross-cultural music cognition suggests that music experiences are also constrained in important ways by the nature of our physical environment, the structure of the auditory system, and evolved strategies of perceptual and cognitive processing (Huron, 2006; Patel, 2008; Thompson & Schellenberg, 2006). Such constraints provide a foundation upon which processes of enculturation lead to additional layers of culture-specific understandings of music. They are an important source of commonalities in music cognition across cultures, and may exert a powerful influence on emotional responses to music. They do not necessarily lead to musical 'universals' if environment-specific or culture-specific influences overwhelm their impact. However, the effects of constraints can sometimes lead to universal, or near-universal, aspects of music cognition or experience, giving rise to striking commonalities across cultures. The following observations have been proposed as likely candidates for musical universals:

- a processing advantage for music built on a small number of discrete pitch levels that are spaced unevenly (e.g. the major scale);
- greater sensitivity to pitch contour than to exact intervals;
- perceived similarity of pitches separated by an octave;

- greater sensitivity to pitch relations than to absolute pitch values;
- sensitivity to sensory consonance and dissonance;
- a tendency to perceive sequences of pitches that are proximal in pitch as part of the same group;
- a processing advantage for music that contains a regular temporal pattern of stress; and
- a tendency to perceive pitches as having different levels of stability, with one pitch class often acting as a point of reference for other pitch classes.

Such commonalities can be explained with reference to one or more cognitive, environmental, genetic, and biophysical constraints, which may be collectively referred to as *system constraints*. The similarity across cultures of sung phrase lengths, for example, may arise in part because human short-term memory has limited capacity and would be burdened by excessively long phrases (an evolved feature of cognitive architecture). In addition, the ability to produce a melodic phrase vocally is constrained by the biophysics of singing, in that air resistance, breathing apparatuses, and oxygen requirements of humans are similar across cultures.

Experience and learning invariably build upon, expand, and in some cases can counter the constraints imposed by cognitive, environmental, genetic, and biophysical factors. For example, there is considerable diversity in musical *scales* across cultures, including the amount of tonal material and the number of scale notes (e.g. 12 tones in the Western chromatic scale, seven tones in the Western major, Byzantine, Gypsy and Hungarian scales, five tones in Celtic, West African and Indonesian scales), the intervals formed by scale notes, and the tuning of scale notes (Patel, 2008). Indeed, the influence of experience and learning is vast and leads to patterns of thought, behaviour, and aesthetic sensibility that are often unique to specific subcultures, as documented by the *thick descriptions* provided by ethnomusicologists and anthropologists (Geertz, 1973; Nettl, 1983). When differences in musical systems, behaviours, and experiences are observed from one culture or subculture to the next, they can usually be interpreted as effects of enculturation.

Cross-cultural music cognition is a valuable strategy for understanding the complex interplay between early predispositions and enculturation, but other scientific approaches are also essential to a full understanding of the issues involved. In particular, developmental approaches, neuroscientific data, and evolutionary psychology are particularly important for interpreting the nature of culture-transcendent and culture-specific sources of musical understanding (see Cross, 2003; Livingstone & Thompson, in press; and Chapters 5, 12, and 23, this volume).

Considerable caution must be taken when attempting to draw conclusions from cross-cultural similarities in musical practices and structures. In particular, observations of apparent 'universals' in music have often been mistakenly interpreted as implicating the involvement of genetic (innate) or 'hard-wired' biological determinants of musical systems and behaviours. There are important qualifications that must accompany this interpretation. First, genetic constraints on musical behaviours that result from a process of *phylogenesis* represent just one of several potential system constraints. Commonalities in musical behaviours across cultures also arise from biophysical constraints, such as oxygen consumption requirements, or similarities across cultures in acoustic environments, such as the tendency for sequences of pitched sounds emanating from a single source to exhibit pitch proximity. Second, many apparent constraints on musical systems may not be specific to music. Some may arise as a by-product of general auditory processing or the limits of short-term memory (Justus & Hutsler, 2005; Patel, 2008). Third, commonalities in music behaviours across cultures frequently arise from transcultural diffusion and interaction. In the twenty-first century, it is difficult to find individuals who have not been exposed to a large number of musical styles and tuning systems from a wide range of cultures and subcultures. Therefore, arguments for innately determined constraints on musical systems are compelling only when they are grounded in evolutionary biology, when other potential constraints have been ruled out, or when accompanied by direct genetic evidence. Indeed, many phenomena that appear to be innate may prove to be learned, because learning takes place in an interaction between organisms and physical environments that is ubiquitous around the world. As such, many apparent 'universals' of music probably have no genetic basis.

Conversely, differences in musical practices across cultures do not imply that genetic or innate factors play no role in such practices. Rather, cultural differences result from *ontogenetic* processes, which are themselves products of natural selection that permit cognitive mechanisms to adapt to environmental conditions. Such processes are guided by knowledge and expectancies, and their operation changes with experience (Heyes, 2003). Ontogenetic processes account for differences across cultures in the communication of emotion in music, but they do not guarantee such differences because some environmental and biophysical influences on musical systems are ubiquitous and can lead to musical universals in the absence of any 'innate' influences on music cognition. This conception of phylogenetic and ontogenetic influences has the potential to resolve some of the abiding concerns surrounding the study of similarities and differences in the musical practices of different cultures.

The connection between music and emotion may be examined at several levels of analysis relevant to cross-cultural comparisons, with some levels relying more on phylogenetic determinants and other levels relying more on ontogenetic processes. Environmental signals that remain stable across geographical regions and over long periods of evolutionary time are most economically handled by phylogenetic determinants. For example, the startle response that follows an unexpected and intensely loud sound is instinctive; it is a genetically determined reflex that requires no process of learning. Other phylogenetically-determined mechanisms include early feature detection and orienting reflexes. More generally, psychological processes, including those associated with emotional responses to music, vary in the extent to which they permit adaptive modification in response to changing environmental conditions.

Environmental signals that are unstable across time or geographical location are most effectively handled by ontogenetic processes. Indeed, ethnomusicologists have revealed that there is a high degree of semiotic variability in our auditory environment, such that the role of learning and enculturation is indispensable to full (adult) understanding of the many connections between music and emotion. For example, the increased tension that is experienced by a North American listener following the occurrence of a nondiatonic tone results from an ontogenetic process: it requires a process of enculturation to a diatonic tonal system and is manifested only after a certain stage of development. Indeed, children are not sensitive to differences in mode until after the age of five (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001).

Phylogenetic determinants are responsible for predispositions that shape initial perceptions of musical attributes, such as early preferences for consonance over dissonance, and they exert culture-transcendent forces on the connection between music and emotion (Dowling & Harwood, 1986). Ontogenetic processes build upon such predispositions and permit cultural fractionation to occur in the emotional connotations of music across musical traditions, cultures, styles, and time periods. Examples include the banning of the tritone interval from Western tonal music in sacred music during the middle ages, and the characterization of certain kinds of music as socially or emotionally dangerous at various points in history (e.g. music composed in the mixolydian scale in 350 BC, ragtime in the 1900s, heavy metal in the 1980s). Mature forms of musical understanding reflect a network of cognitive processes that combine phylogenetic and ontogenetic sources of adaptive features. Adaptive features that have a phylogenetic source were favoured by natural selection and are specialized to respond to highly stable features in the environment. Adaptive features that have an ontogenetic source were not favoured by natural selection but are generated in the course of development, and vary across cultures.

All musical systems implicate a dense combination of emotional and structural signals, some of which are culture-transcendent and handled by phylogenetically determined cognitive mechanisms, others that are culture-transcendent but nonetheless handled by ontogenetically determined cognitive mechanisms, and still others that are historically and culturally specified and again handled by ontogenetically determined cognitive mechanisms. Phenomena that are culturally transcendent are often identified by experimental strategies developed in cognitive science. Behaviours that are historically and culturally specified are often identified through ethnographic strategies, commonly applied in ethnomusicology, sociology, and anthropology.

## 27.2 Emotion as a cross-cultural construct

Emotions are, in an important sense, biological mechanisms of 'preparedness' (Darwin, 1972; Ekman, 1992; Huron, 2006; Lazarus, 1991; Tooby & Cosmides, 1990), and this functional property of emotions has important implications for cross-cultural comparisons. Emotional states often have tangible biological functions, facilitating survival responses such as fighting, fleeing, or freezing. Emotional experiences and behaviours are observed in all cultures, and may be interpreted as universal responses to significant events. Some events are significant regardless of one's cultural environment, and may generate consistent emotional responses across cultures. Being in imminent danger of attack by a wild animal is likely to generate heightened arousal and fear in any individual, whether from Bali, China, Finland, or Canada. Many events acquire meaning through social construction, however, and an understanding of their significance requires knowledge of cultural traditions. The convention of bowing as a signal of respect and politeness in some cultures can be very specific: bowing too briefly or for too long can have significant connotations that elicit a strong emotional response.

According to Matsumoto (1989), emotions are biologically programmed, but the process of learning to perceive and express emotions is dependent on enculturation. Categories of emotion are also influenced by cultural factors. Broad dimensions such as valence and arousal may be applicable across cultures, while specific emotional categories such as the Indian *rasa* '*adbhutam*' (meaning wondrous) may not translate well across cultures (Deva, 1973; Russell, 1994). Moreover, the notion that typical interactions involve the communication of a singular definable emotion may be unrealistic. In many if not most contexts, people experience varying combinations of emotions, and musical expressions of human experience often reflect such mixed emotions. If decoding individual emotions across cultures is difficult, decoding expressions of mixed emotions is an even greater challenge. To our knowledge, investigations of sensitivity to mixed emotions across cultures have yet to be conducted.

Landmark studies by Ekman and colleagues suggested that facial emotion recognition is universal for a number of basic emotions (Biehl et al, 1997; Ekman, 1970, 1972; Ekman, Sorenson, & Friesen, 1969). Ekman also emphasized cultural differences in display rules for emotional communication, but it has nevertheless been suggested that many of the early studies of emotional communication underestimated the role of enculturation by relying on forced-choice response formats and posed expressions (Russell, 1994), and because the researchers chose to focus on agreement across cultures rather than disagreement. The challenge of interpretation inherent in such studies is exemplified by a study that revealed that members of Bahinemo tribes perceived all faces of Americans as 'angry' (Sorensen, 1975). Such findings underscore the difficulty of decoding emotions across cultural boundaries. Indeed, some scholars have even argued that certain emotions are unique to particular cultures, such that cross-cultural decoding is essentially impossible (Briggs, 1970; Lutz, 1988).

A number of researchers have emphasized the significance of such cross-cultural differences (Matsumoto, 1989; Mesquita & Frijda, 1992; Mesquita et al, 1997). This research has led to the speculation that many cultures have distinctive norms for displaying and decoding emotions, or even culture-specific emotions. Such norms may function to maintain positive social interactions, and are especially apparent for the expression and recognition of negative emotions, which have the greatest potential to disrupt social interactions (Matsumoto, 1989). The role of cultural norms can also explain the phenomenon of ethnic or in-group bias. For example, when emotional speech is recorded from an individual, decoding accuracy is often highest by members of the same cultural group from which the utterance was recorded (Kilbride

& Yarczower, 1983; Mesquita & Frijda, 1992). Although the study of ethnic bias has received considerable attention, a full understanding of the phenomenon is unlikely to emerge soon, because results are inconsistent across studies, emotional channels, and cultures under consideration (Elfenbein & Ambady, 2002).

Differences in emotion recognition across cultures may also arise from language differences. For example, some languages have highly expressive terms for certain emotional concepts and lack terms for expressing other emotional concepts. Such differences in language apply to the expression of emotion across communication channels. Note, however, that the lack of an emotional term does not imply the absence of that emotion in a culture, but reflects a lack of emphasis on a particular subordinate category (e.g. frustration) within a broader emotion category (e.g. anger) in the social discourse of that culture (Boucher, 1979; Johnson-Laird & Oatley, 1989; Russell, 1991).

In addition, each communication channel is associated with channel-specific properties that can sometimes interfere with cross-cultural emotional communication. With respect to speech prosody, differences in emotional decoding accuracy across cultures can also arise because culture-transcendent prosodic cues are masked differentially by attributes of voice quality that are unfamiliar to non-speakers of that language. As an example, pitch changes involved in tonal languages like Mandarin may confuse or mask the interpretation of emotional intent to individuals who are unfamiliar with that language. Similarly, difficulties in emotional decoding of music can arise because unfamiliar conventions of composition and performance overwhelm impressions of the music, masking culture-transcendent emotional cues.

## 27.3 The cue-redundancy model

A basic property of music that is shared by all cultures is that it involves multiple features that must be analysed by the auditory system. Even silence is registered by the brain and contrasted with the possibility of sound. One or more acoustic features such as intensity, tempo, pitch height, timbre, sensory dissonance, rhythmic complexity, and harmony may be present in the music, and each feature varies with the emotional connotations of the music. Just as instrument membranes have resonant frequencies that allow them to vibrate in sympathy with different sounds, the various acoustic features of music can be said to 'resonate' to varying degrees with the emotional character of the music. In turn, the auditory receptors for these music features themselves resonate with the emotional character of the music.

Each emotion is correlated to some extent with different types of changes in acoustic features, but many of these associations are moderate or even weak, and so no individual feature can be relied upon to determine an intended emotion. Rather, the full set of features available must be evaluated in a probabilistic manner, with different features weighted according to the strength of their association with each emotion (see Chapter 14, this volume). A yet-to-be-discovered mechanism that summarizes the available evidence from multiple emotional cues determines the emotion that is ultimately perceived and experienced. Because cues overlap in their emotional implications, the process of emotional decoding has the advantage of cue redundancy, allowing emotional communication to occur through convergent evidence and appraisals based on 'family resemblances' rather than on defining features (Rosch & Mervis, 1975; Wittgenstein, 1953).

Thus, for example, one might imagine two musicians from disparate cultures (A and B) each composing a piece of music on the theme of joy. They would probably compose in a scale they know to be associated with joy in their culture. Musician A might use intricate harmonies, a simple rhythm, and compose for a high-pitched wind instrument that is instantly recognized by her people as 'the joy flute'. Musician B might compose an unaccompanied melodic line, a complex rhythm, and compose for a stringed instrument that is widely associated with joy by his people. They might both perform their pieces at a fast tempo and moderate amplitude.

Although there would probably be many surface differences between the two compositions, each piece would contain psychophysical cues (e.g. rate, intensity, pitch height) that have emotional connotations across cultures, and culture-specific cues (e.g. cadences, specific instruments, harmonic progressions) that enhance the recognition process for members of the same culture. Balkwill and Thompson (1999) defined *psychophysical cues* as 'any property of sound that can be perceived independent of musical experience, knowledge, or enculturation' (p. 44). However, psychophysical cues are not merely the result of innate processes but also reflect other system constraints, as described earlier. For example, music that communicates high-arousal emotions may be characterized by attributes that reflect the increased oxygen requirements associated with high-arousal states. Because the connection between oxygen requirements and arousal states is ubiquitous, certain attributes of high-arousal music are likely to occur across cultures. The presence of such attributes is unlikely to be dictated by direct genetic encoding, but may emerge indirectly as a result of biophysical constraints.

Examples of psychophysical cues include sound intensity, rate (tempo), melodic complexity, melodic contour, pitch range, rhythmic complexity, dynamics, and timbre. Listeners have the advantage of 'cue redundancy' because most music communicates emotion through multiple redundant cues (e.g. Juslin, 2000; 2001; Thompson & Robitaille, 1992). Listeners who share cultural experiences with the singer/songwriter have additional benefits of cue redundancy, whereas listeners from another culture must rely on psychophysical cues in order to recognize the emotional connotation of the music.

Figure 27.1 illustrates the *cue-redundancy model* (CRM) proposed by Balkwill and Thompson (1999). A central tenet of the CRM is that listeners can appreciate affective qualities of unfamiliar music by attending to psychophysical cues. Consciously or intuitively, composers and performers draw upon these culture-transcendent cues as well as culture-specific conventions in order to express emotion in music. Listeners, in turn, attend to either or both sources of emotional meaning. When psychophysical

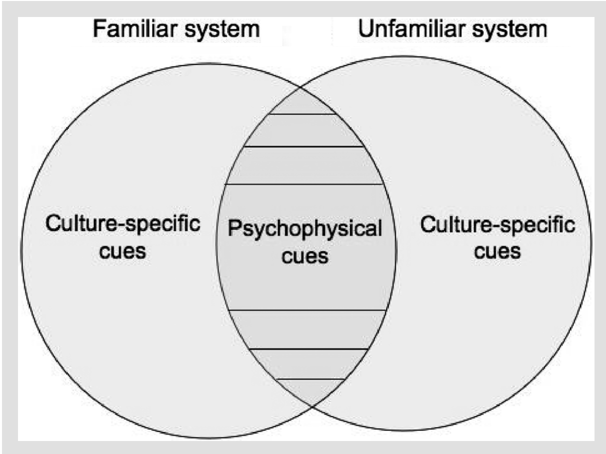


**Fig. 27.1 The cue-redundancy model (CRM), originally proposed by Balkwill & Thompson (1999).**

cues in the music are not overwhelmed by culture-specific cues, listeners from outside that culture are more likely to be able to decode the emotional intent.

Emotional decoding is affected by familiarity with the conventions of the tonal system and sensitivity to psychophysical cues. Listeners who are familiar with a musical style should find it relatively easy to decode emotional meaning in that music, because they can draw from both culture-specific and psychophysical cues. As such, in-group advantages are a basic prediction of the CRM. However, because studies of in-group advantages have yielded inconsistent results for other channels of emotional communication, their validity for emotional messages in music should be subjected to rigorous cross-cultural evaluation.

When cultural-specific cues are unfamiliar or absent, listeners may still attend to psychophysical cues such as tempo and intensity. These cues provide listeners with a general understanding of the intended emotion even for unfamiliar musical styles. Psychophysical cues are especially powerful signals because their interpretation requires no knowledge of musical conventions. Indeed, cues such as tempo and intensity have emotional significance in other channels such as speech prosody, animal calls, and alert signals.

The CRM can also be extended to the domain of speech prosody. Vocal expressions of anger are often marked by fast tempo and greater intensity; vocal expressions of joy are often fast in tempo and have a large pitch range, and vocal expressions of sadness are typically slow in tempo and low in pitch register (Juslin & Laukka, 2003). When speech prosody and music convey the same emotion, they also tend to share many of the same psychophysical cues (Ilie & Thompson, 2006). These commonalities suggest that emotions are communicated through psychophysical cues that not only transcend cultural boundaries but are manifested in different channels of communication.

## 27.4 Fractionating emotional systems

The concept of fractionating emotional systems (FES) extends the implications of the CRM, providing a broader framework for summarizing and interpreting cross-cultural data on emotional communication. Like the CRM, the FES framework acknowledges culture-transcendent and culture-specific influences on the emotional dimension of music and prosody. Ontogenetic processes allow for enculturation to occur, giving rise to culture-specific and domain-specific emotional cues and thereby differentiating or 'fractionating' emotional systems in different cultures and domains. Phylogenetic processes provide a point of convergence between emotional channels and cultural groups; ontogenetic processes provide a mechanism for distinguishing and focusing emotional channels and cultural groups.

Psychophysical cues are represented as a foundation for emotional communication that is common across cultures and auditory domains (music and speech prosody). This foundation is presumed to result from phylogenetic adaptations and other system constraints (ubiquitous environmental or biophysical influences), and is an important source of commonalities in emotional communication across cultures and auditory domains. For example, evidence suggests that infants prefer infant-directed song to infant-directed speech, just as they prefer infant-directed speech to affectively neutral speech (Nakata & Trehub, 2004; Singh et al, 2002). Such preferences may reflect early sensitivity to psychophysical cues along with an instinctive preference for positive emotional messages.

Emotional communication is maximally adaptive when it involves a combination of phylogenetic processes that handle stable acoustic features in the environment (e.g. the correlation between sound intensity and object proximity) and flexible ontogenetic processes for handling emotional signals that vary across environments. As displayed in Figure 27.2, enculturation builds culture-specific conventions and experiences of music from the foundations provided by phylogenetic contraints. The fanning out of triangular shapes represents the differentiating effects of ontogenetic processes, which promote the development of culture-specific experiences of the links between music and emotion, and between speech prosody and emotion.

Early in development, before enculturation has had the opportunity to exert a significant influence, there are few differences between individuals across cultures and few differences in the cues that must be used to decode emotional meaning in music and speech. With development, enculturation becomes increasingly evident, and leads to sensitivity to culture-specific and domain-specific emotional cues. This process

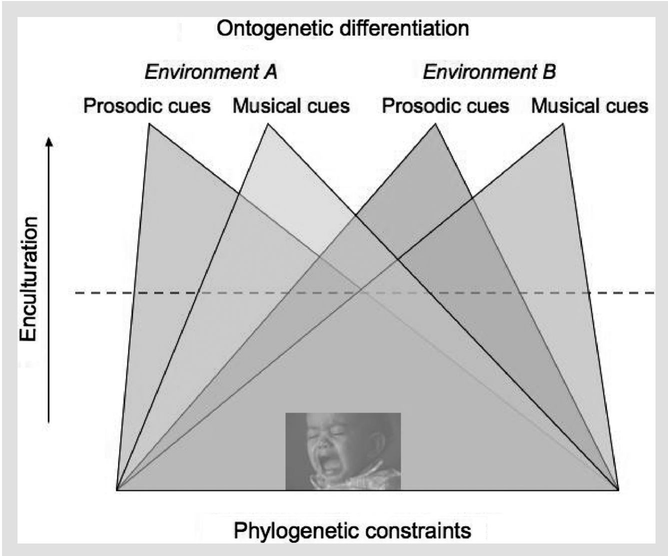


**Fig. 27.2 An illustration of the concept of fractionating emotional systems (FES). The cue-redundancy model addresses the distribution of psychophysical and culture-specific cues for an enculturated individual, and may be understood as a cross-section of FES (dotted line).**

permits the partial fractionation of different systems of emotional communication across cultures and emotional communication systems.

Throughout development, psychophysical cues bootstrap learning in complex ways. For language acquisition, they interact with prosodic cues to segmentation and syntactic relationships, and may be recruited to draw attention to statistical regularities in the transitional probabilities among acoustic events (Palmer & Hutchins, 2006; Saffran, Aslin, & Newport, 1996). Moreover, different languages and musical styles permit different degrees of variation in psychophysical cues. For example, the use of pitch variation for emotional communication is constrained in tone languages and in musical styles that reply on percussive instruments. Thus, the role of each psychophysical cue for emotional communication varies as a function of the learning environment.

The cue redundancy model may be understood as a cross section of FES at any one stage of development, from early development to full maturity. Early in development, an individual will be predominantly sensitive to psychophysical cues that transcend cultures and emotional channels. This stage would be represented in the CRM as a relatively large proportion of overlap in the cues used for emotional communication. The process of enculturation leads to an increasing contribution of culture-specific and domain-specific cues to emotional meaning, and is represented in the CRM by an increased proportion of such cues.

Nonetheless, a basic premise of FES is that psychophysical cues continue to exert an influence on emotional communication throughout the lifespan, providing a means for cross-cultural emotional communication and for commonalities between music and speech prosody in their use of emotional cues.

## 27.5 Cross-cultural studies of emotion and music

Cross-cultural research on emotion and music includes studies that directly assess the capacity of individuals to interpret the emotional meaning of music across cultures, as well as studies that provide indirect evidence for cross-cultural commonalities in the association between music and emotion. The latter studies include developmental approaches and investigations of tension and complexity. These three sources of evidence will be reviewed in turn.

### 27.5.1 Direct assessments

Gundlach (1932) conducted one of the early psychological studies of emotional communication that considered music outside the Western classical tradition. His aim was to determine whether or not there are 'objective characteristics' of music in aboriginal cultures that form the basis of communicating specific moods. He analysed archives of songs collected from several North American aboriginal tribes. Songs were chosen for their correspondence with important aspects of aboriginal life (healing songs, love songs, and war songs) that were likely to express a specific emotion. Analyses of pitch, range, tempo, and rhythm revealed that war songs were typically low in pitch, fast in tempo, and wide in pitch range, whereas love songs were usually high in pitch, slow in tempo, and had a moderate pitch range. Healing songs often had the fewest number of rhythmic changes, whereas war songs had the greatest number of rhythmic changes. Gundlach concluded that there were measurable objective differences between these songs as a function of their emotional character.

In a later study, Gundlach (1935) presented a phrase from 40 different musical pieces to groups of listeners with varying levels of musical training. Their task was to select from a list of terms the verbal descriptor that best characterized the mood expressed by the composer, or to fill in their own description if they could not find an appropriate term. Agreement among participants was observed for each phrase (mean agreement $\geq 0.75$). Analyses of tempo, rhythm, intervals, range, intensity, mean pitch, and melody range revealed several 'objective' features of phrases associated with emotional communication. However, the extent to which such features can be decoded across cultures was not determined.

Morey (1940) extended this work by presenting samples of Western classical music (Schubert, Davies, Handel, and Wagner) to 11 male members of the Loma, a tribe in Liberia, West Africa. The pieces were selected on the basis that they conveyed emotions such as fear, reverence, rage, and love. Participants were teachers and students at the Holy Cross Mission School at Bolahun. They were asked to indicate which emotions were expressed by each music selection. Morey observed that the music often did not appear to elicit any emotions—or at least they did not evoke emotional judgements that are typical of Western listeners. In another experiment, Morey presented his music samples to 20 members of the Zealua Loma tribe outside the school. He reported that they did not find the music interesting and many of them left while it was still playing.

With ubiquitous media and increasing globalization in the twenty-first century, Western tonal music is so pervasive that it is rapidly losing its value as music stimuli in cross-cultural investigations. Research that involves Western tonal music therefore requires participants from extremely isolated cultures who are naïve to this music. Because radio and television are almost universally available, it is reasonable to wonder whether such cultures exist. Tom Fritz, together with Stefan Koelsch and colleagues, performed a cross-cultural study with participants from a native African population known as the Mafa (Fritz, Sammler, & Koelsch, 2006; Fritz et al, in press). The Mafa are one of approximately 250 ethnic groups that make up the population of Cameroun. They are culturally isolated and live in the Extreme North Province in the Mandara mountain range. The remote Mafa settlements do not even have electrical supply, and are inhabited by individuals who pursue a traditional lifestyle and have not been exposed to Western music. Twenty German and 21 Mafa listeners judged the emotions conveyed in a sample of Western music. The music excerpts, which varied in duration (9–15 s), mode, tempo, pitch range, tone density, and rhythmic regularity, were composed to express the emotions happy, sad, and scary.

Music stimuli were classified using depictions of facial expressions from the Ekman archive (happy, sad, scary) (Ekman, 1976). Three Mafa participants could not recognize the facial expressions on the two-dimensional paper and were thus excluded from the test. Both German and Mafa listeners recognized all three emotions in the music samples at levels that were above chance performance (33 per cent). Among German listeners, recognition of happy music (99 per cent) was better than recognition of sad music (93 per cent) or scary music (81 per cent). Similarly, among the Mafa listeners, recognition of happy music (65 per cent) was better recognized than recognition of sad music (49 per cent) or scary music (48 per cent). This advantage for happy music did not arise because the two negative emotions were confused with one another: errors for each target emotion were distributed equally among the other two categories. Interestingly, the capacity to decode emotional expressions was significantly correlated with the degree of appreciation of the compositions. The authors proposed that 'meaningfulness' in music (clear emotional expression) leads to increased appreciation.

Hindustani music offers an ideal stimulus for cross-cultural comparisons, because of its strong association with emotional communication, and because it is easy to locate individuals who are unfamiliar with this music. Central to Hindustani music is the *raga-rasa* system in which each *raga* (set piece) is associated with one or more *rasas* (moods or essences). Each raga is also associated with a distinctive colour, season, and time of day. Deva and Virmani (1975) compared the intended mood of Hindustani ragas with mood ratings made by Indian listeners. In one experiment, four excerpts from ragas (roughly two minutes each) were played to 37 listeners, who were asked to choose from a list of mood adjectives the emotion that they believed was expressed. The authors reported that the perceptions of listeners often matched the intended mood of the ragas. In a second experiment, 228 listeners were asked to rate excerpts of a raga on expressed mood, associated colour, season, and time of day. Listeners judged the raga ('*Bhairav*') as dominant in empathy and compassion, and as reflecting vitality, positive affect, courage, and tranquillity. The majority of listeners associated the raga with the colour white or yellow, the season of autumn, and the time of day as early morning or early evening. According to Deva and Virmani, ratings generally approximated the intended mood of the raga.

In a later study, Gregory and Varney (1996) examined whether Western listeners who are unfamiliar with Hindustani music can interpret the emotional connotations of that music. Gregory and Varney asked British residents of Western and Indian heritage to judge the emotional messages implied in commercially recorded excerpts of Hindustani ragas, Western classical music, and Western new age music. After hearing each excerpt, listeners selected adjectives from a list developed by Hevner (1936; see Chapter 14, this volume). Western and Indian listeners were sensitive to intended emotions in Western music, but not Hindustani ragas. The finding suggests that emotional cues in unfamiliar music are not always interpretable, and suggests that culture-specific properties of music can mask culture-transcendent emotional cues.

Balkwill and Thompson (1999) performed a similar study, but observed far better emotional decoding of Hindustani music by Western listeners. They asked Canadian listeners to judge the emotional content of field recordings of Hindustani ragas, and to rate the presence of several psychophysical attributes in the music. Recordings were obtained of solo performances of Hindustani ragas that were explicitly intended to evoke specific emotions. Excerpts intended to convey joy/*hasya* were assigned high ratings of joy; excerpts intended to convey sadness/*karuna* were assigned high ratings of sadness; and excerpts intended to convey anger/*raudra* were assigned high ratings of anger. High ratings of joy were associated with high ratings of tempo and low ratings of melodic complexity. High ratings of sadness were associated with low ratings of tempo and high ratings of melodic complexity. Anger ratings were not significantly associated with ratings of tempo or complexity, but ragas performed on stringed instruments were rated much higher in anger than those performed on the flute.

A subsequent study conducted in Japan extended this work and provided additional evidence for the cue-redundancy model. Whereas Balkwill and Thompson (1999) examined judgements by Canadian listeners of Hindustani music, Balkwill, Thompson, and Matsunaga (2004) examined judgements by Japanese listeners of Japanese, Western, and Hindustani music. Music samples consisted of emotive music by two Canadian musicians in Toronto, Ontario, Canada, and six Japanese musicians in Sapporo, Hokkaido, Japan. The music provided by the Canadian musicians was a compilation of improvisations on each target emotion. Each musician was asked to draw upon their musical experience and repertoire and to focus on the task of conveying each of the three specified emotions. This procedure was adopted to minimize extra-musical associations that many listeners may have formed between past pairings of familiar music with their own experiences. Because Japanese listeners are generally well acquainted with classical and popular Western music, the use of improvisations also served to decrease the level of familiarity with the Western music materials for this listener group.

The music provided by the Japanese musicians was a compilation of several genres of traditional composed music (*gagaku*, *shinto* and *minyo*). Each musician was asked to choose pieces from their repertoire that they felt would best convey each target emotion. These two stimulus sets, along with the Hindustani music previously collected (from Balkwill & Thompson, 1999) were then edited to create short excerpts (mean duration = 30 s) intended to evoke anger, joy, or sadness.

One hundred and forty-seven Japanese listeners (76 women, 71 men, mean age = 23.7) rated the degree of anger, joy, sadness, complexity, loudness, and tempo in each sample of music from Western, Japanese, and Hindustani stimulus sets. As expected, Japanese listeners were sensitive to the intended emotion of music from Japanese, Western, and Hindustani music, and their judgements were associated with the presence of psychophysical cues. High ratings of joy were associated with a fast tempo and a simple melody; high ratings of sadness were associated with a slow tempo and a complex melody; and high ratings of anger were associated with high intensity and a complex melody. The findings confirmed that Japanese listeners were sensitive to the emotions communicated in familiar and unfamiliar music, and their sensitivity was associated with the perception of psychophysical cues that transcended cultural boundaries.

More recently, we compared the judgements of Canadian and Japanese listeners for the same music stimuli (Balkwill, 2006). One hundred and thirty-nine Western listeners (82 women, 57 men, mean age = 25.72) participated in the study. As with Japanese listeners, the Canadian group was remarkably sensitive to intended emotions expressed in Japanese, Western, and Hindustani music, and their judgements were associated with the presence of psychophysical cues. Both groups were able to recognize music intended to evoke anger, joy, and sadness in all three tonal systems, and judgements for every emotion were significantly associated with at least one of the psychophysical attributes of complexity, intensity, tempo, and timbre.

An interesting difference observed between the Canadian and Japanese listener groups was the number of psychophysical cues associated with some emotion judgements. For example, in the case of anger ratings the degree of perceived complexity, tempo, and intensity were all significant predictors for Japanese listeners, while for Canadian listeners, only degree of perceived intensity was significant. Previous research involving descriptions of visual stimuli have indicated that Japanese participants tend to process scenes more holistically, reporting more background stimuli than North American participants, who tend to focus on primary focal elements (Masuda & Nisbett, 2001). Whether similar differences between Japanese and North American groups in the quantity and quality of psychophysical cues associated with emotion recognition in this study can be attributed to cultural differences in attention focus or cognitive style is an intriguing question (see Nisbett, Peng, Choi, & Norenzayan, 2001, for a review). Further experiments with music and other auditory stimuli more clearly focused on this question may yield some interesting contributions to this area of research.

### 27.5.2 Developmental evidence

There is little doubt that long-term exposure to the music of one's culture has a powerful effect on our perceptions and experiences of music (Hannon & Trainor, 2007). Memory for music is better as a function of increasing age (Krumhansl & Keil, 1982; Trainor & Trehub, 1994) and of familiarity or conventionality of the music itself (Cuddy, Cohen, & Mewhort, 1981). Music processing is also enhanced by formal music training (Lynch & Eilers, 1991).

However, there is also considerable evidence that basic auditory skills are well developed at birth, and such skills are similar in all humans across cultures. Infants possess a number of perceptual skills for music that are remarkably similar to those of listeners who have had years of exposure to music. These apparent predispositions exert a significant influence on music experiences throughout the lifespan, including emotional responses to music and speech (Trehub, 2000, 2003; Trehub & Nakata, 2002). As examples, infants show evidence of processing advantages for contour patterns over precise intervals, for consonant over dissonant sounds, and for temporal pattern relations over absolute durations (Trehub & Hannon, 2006). Because these skills are observed very early in development, it is possible that their emergence reflects innate biases rather than environmental stimulation. In so far as such early skills do not rely heavily on processes of enculturation, they may form the basis for many important cross-cultural similarities in the perception of music.

It should be emphasized that it is not possible to rule out explanations of early auditory skills based on rapid adaptation of the auditory system to music exposure. To illustrate, the A1 tonotopic frequency map is fully developed (i.e. equivalent to adult organization) in rats two days after they are born (de Villers-Sidani, Chang, Bao, & Merzenich, 2007). If newborn rats are exposed to broadband noise for up to 90 days, no tonotopic map is evident. But within two days of turning off the noise, the A1 tonotopic map will be fully formed. Two conclusions may be drawn from these observations: (1) exposure to pitched tones is necessary for tonotopic perceptions to emerge, and (2) the rat learns pitch with only two days of auditory exposure. Clearly, the effects of learning can be unexpectedly rapid, such that one might be forgiven for assuming that the musical dispositions of newborns are innately determined.

An important influence on the emotional connotation of music is the relative balance of consonance and dissonance. In Western tonal music, consonance usually signals stability and positive emotional valence, whereas dissonance signifies tension, instability, and negative emotional valence. Musicologists have suggested that dissonance is understood and heard somewhat differently across different musical traditions, cultures, styles, and time periods, but evidence suggests that infants begin life with predispositions for processing consonance and dissonance, and these predispositions may remain influential into adulthood. Specifically, infants are highly sensitive to the distinction between consonance and dissonances, and even newborn infants prefer music that is consonant to music that is dissonant.

Zentner and Kagan (1996) exposed 4-month-old infants to consonant and dissonant versions of two melodies. Infants looked longer at the source of sound and moved less when consonant versions of melodies were played than when dissonant versions of melodies were played. Moreover, when the dissonant version was played, infants were more likely to fret and turn away from the music source. The researchers argued that infants are biologically predisposed to prefer consonance to dissonance (Crowder, Reznick, & Rosenkrantz, 1991; Trainor & Heinmiller, 1998).

Infant-directed singing (lullaby) is prevalent across cultures (Trehub, Unyk, & Trainor, 1993; Unyk, Trehub, Trainor, & Schellenberg, 1992), and plays many important functions in infant development, such as the acquisition of attentional skills, the maintenance of emotional stability, and the enhancement of infant–caregiver bonds (see also Chapter 23, this volume). These functions have implications for the survival of the infant, and are thought to have a biological basis. A number of researchers have proposed that all music may have ultimately originated from this form of musical activity, which conferred a survival advantage for ancestral populations. Unlike other primates, infants cannot cling to their mother's fur while the mother forages with both hands. Singing would have allowed mothers to soothe their infants (keeping them quiet to avoid predators) while foraging. If such conjectures are correct, then the diverse musical systems across all cultures should have a common foundation that is reflected in infant-directed songs. Moreover, cross-cultural similarities should be particularly evident for the emotional qualities of music.

In support of this view, Trehub and colleagues demonstrated that infants are naturally attracted to lullabies. Nakata and Trehub (2004) found that 6-month-old infants look longer at audio-visual recordings of their mother if she is singing a lullaby than if she is speaking, suggesting that lullabies generate positive affective states in infants. Moreover, cortisol levels (associated with stress) dropped significantly following infant-directed song, indicating that lullabies are naturally soothing.

Infant-directed song contains many of the emotional cues that are observed in infant-directed speech, including relatively high pitch, slow tempo, and simple and repeating contours (Fernald, 1991; Fernald & Simon, 1984). Is there a general connection between the emotional cues used in music and speech prosody? Evidence by Thompson, Schellenberg, and Husain (2004) suggests that there is. Musically trained and untrained adults were given a test of sensitivity to emotional prosody for phrases spoken in their own language (English) or in a foreign language (Tagalog, a language of the Philippines). The test required listeners to classify the emotional connotation of each utterance as happy, sad, fearful, or angry. Musically trained adults were significantly better than untrained adults at classifying the emotions, including those expressed in the unfamiliar language of Tagalog. That is, cross-cultural sensitivity to vocal emotions was enhanced for musically trained adults. The results suggest that psychophysical cues to emotion are instantiated in both music and speech prosody. Thus, long-term training in music gives rise to enhanced sensitivity to such cues, whether they occur in familiar music, unfamiliar music, familiar speech, or unfamiliar speech.

To address the possibility that the trained adults were naturally more sensitive to vocal emotions irrespective of their training, the researchers assessed the sensitivity of 6-year-olds to emotional prosody after one year of music lessons, drama lessons, or no lessons. The results revealed that both drama and keyboard lessons resulted in enhanced sensitivity to emotional prosody, including prosodic materials from an unfamiliar language (Tagalog). Thus, drama lessons, with their emphasis on learning to express different emotions, had a 'direct' effect on learning, whereas music lessons had a 'transfer effect' to the domain of speech prosody. One explanation for these findings is that music and speech prosody implicate the same acoustic cues to emotion. Music training may not only enhance sensitivity to culture-specific properties of a musical system; it may also engage and refine neural processes that respond to universal acoustic cues to emotion, thereby enhancing sensitivity to emotions in music and speech across cultures.

### 27.5.3 Tension and stability

Musical tension provides another source of emotional meaning in music. Tension can arise from several sources, including sensory dissonance, violations of expectations, complexity, and tonal instability. When we listen to music from an unfamiliar genre, our levels of uncertainty are generally higher, giving rise to perceived complexity and tension. Huron (2006) used a 'betting paradigm' to evaluate the level of uncertainty of continuations in a melody. Balinese and American musicians listened to an unfamiliar traditional Balinese melody and provided judgements, via a betting game, of their ability to anticipate future pitches. Not surprisingly, uncertainty was greater for American listeners than for Balinese listeners, but the levels of uncertainty also showed a different trajectory for the two groups. For Balinese listeners, the level of uncertainty was maximal in the middle of the melody, and minimal near the beginning and end of the melody. No such rising and falling pattern was observed for American listeners, who simply showed a subtle decline in uncertainty as the melody progressed.

In a series of studies of the *implication-realization* (I-R) model (Narmour, 1990), American and Chinese listeners were asked to provide ratings of continuation following fragments of British folk songs, Webern lieder, and Chinese pentatonic songs (Krumhansl, 1995; Schellenberg, 1996). Certain predictors based on the I-R model accounted for a significant proportion of the variance in ratings for the two listener groups and three musical styles, suggesting cross-cultural similarities in the formation of expectancies. For example, listeners tended to expect continuation tones that were close in pitch to preceding tones (see also Carlsen, 1981; Thompson & Stainton, 1998). If such expectancies are 'hard-wired', then they could provide a basis for emotional communication across cultures, whereby large changes in pitch in a melody universally give rise to heightened arousal (i.e. surprise) and increased tension.

Perceived tension in music can also be influenced by the overall complexity of the music. Cross-cultural comparisons suggest that a range of factors influence perceived complexity, and that these factors are weighted differently for judgements of familiar and unfamiliar music. Eerola, Himberg, Toiviainen, and Louhivuori (2006) asked Finnish and African listeners to judge the complex of Western and African folk songs. A range of melodic complexity measures was developed, including average interval size, entropy of interval distribution, entropy of note duration, rhythmic variability, note density, tonal ambiguity, and contour entropy (where entropy is a measure of uncertainty).

Overall patterns of judgements were similar for the two groups and the two styles of music, supporting the idea that complexity is a culture-transcendent construct. However, culture-specific influences on judgements of complexity were also evident. Whereas the style of the Western folk songs was familiar to both groups, the style of the African folk songs was familiar to only the African group. Complexity measures reflected these differences in familiarity, with differences in the predictive power of complexity variables differing between groups for the African folk songs. Taken together, the results suggest that melodic complexity transcends cultural boundaries, but musical enculturation still plays an important role.

Tension is also inversely related to tonal stability (Lerdahl & Krumhansl, 2007). Cross-cultural studies of tonal stability suggest that in the absence of knowledge of a tonal system, the relative stability of tones (tonal hierarchies) formed by non-native listeners were influenced mainly by pitch distribution. When Western listeners provide goodness-of-fit judgements for probe tones presented in the context of North Indian music (Castellano, Bharucha, & Krumhansl, 1984) and Balinese music (Kessler, Hansen, & Shepard, 1984), their judgements are primarily influenced by the frequency of occurrence of pitches. Similarly, when Western listeners provide judgements of North Sami yoiks the influence of Western schematic knowledge is strongly evident; when indiginous Sami listeners judge the same materials there is relatively little evidence for an influence of Western schematic knowledge (Krumhansl et al, 2000).

## 27.6 Cross-cultural studies of emotion and prosody

Because emotion is expressed in music and speech prosody using very similar cues, cross-cultural studies of speech prosody are highly relevant to cross-cultural investigations of music and emotion (Ilie & Thompson, 2006; Juslin & Laukka, 2003; Patel, 2008). Indeed, tempo, intensity, and timbre affect judgements of emotion in both music and speech. For example, vocal expressions of anger are often marked by fast tempo and greater intensity (in the case of 'hot' anger); vocal expressions of joy are often fast in tempo with a higher pitch range; and vocal expressions of sadness are typically slow in tempo and low in pitch (Bachorowski & Owren, 1995; Frick, 1985; Scherer, 1986; Scherer & Oshinsky, 1977; Thompson & Balkwill, 2006).

Several investigations of emotional prosody have converged on the view that listeners are highly sensitive to the prosodic expression of anger, joy, and sadness (Juslin & Laukka, 2003). Conversely, a relatively low success rate has been reported for the recognition of fear and disgust in speech. Emotional sensitivity for prosodic materials is similar to emotional sensitivity for musical materials, although the two modalities differ in the emotions that are most commonly conveyed. Anger is commonly communicated through speech prosody, and it is often the best recognized emotion for spoken stimuli (e.g. Banse & Scherer, 1996). Interestingly, anger is also recognized rather well in music (Juslin & Laukka, 2003), even though this emotion is communicated in only certain genres of music, and composers have difficulty expressing anger with pitch or rhythmic structure (Thompson & Robitaille, 1992).

Although the CRM was developed to account for cross-cultural studies of music and emotion, it readily applies to emotional prosody, suggesting that a broader framework is needed to account for the emergence of emotional systems. The concept of fractionating emotional systems assumes that psychophysical cues provide a common point of departure for the development of emotional communication in music and speech prosody. According to FES, individuals should be able to recognize emotions expressed prosodically even when utterances are spoken in an unfamiliar language. Moreover, an ethnic bias should be observed for decoding emotional messages expressed prosodically, because of the presence of culture-specific cues.

Prosodic cues for emotion are well documented for English and some Western European languages (German, Dutch). Joy is expressed with a comparatively rapid speaking rate, high average pitch, large pitch range, and bright timbre; sadness is conveyed with a slow speaking rate, low average pitch, narrow pitch range, and low intensity; anger is expressed with a fast speaking rate, high average pitch, wide pitch range, high intensity, and rising pitch contours; and fear is conveyed with a fast speaking rate, high average pitch, large pitch variability, and varied loudness (Juslin & Laukka, 2003; Scherer, 1986).

Analyses of recognition rates indicate that prosodic cues alone allow listeners to identify the emotion being conveyed, with identification rates of roughly four to five times the rates that would be predicted by guessing (Banse & Scherer, 1996; Frick, 1985; Scherer, 1979, 1986; Standke, 1992; van Bezooijen et al, 1983). Not all emotions are decoded equally well from speech prosody, however. For example, anger and sadness are typically decoded more reliably than joy and fear (Banse & Scherer 1996; Johnstone & Scherer 2000).

A large number of studies have revealed sensitivity to emotional prosody across cultures and in unfamiliar languages. Kramer (1964) demonstrated that English-speaking American listeners could identify vocal expressions of emotion rendered in English and Japanese (content-filtered) at a rate better than chance. Beier and Zautra (1972) found that American, Polish, and Japanese listeners were able to identify emotion in American English speech. Similar positive results were reported by McCluskey and Albas (1981) for their study of Canadian and Mexican children and adult listeners judging vocally expressed emotion provided by Canadian and Mexican actors. In some cases, listeners were more accurate in recognizing the emotion in the unfamiliar language than in their own. In an experiment with samples of emotive speech spoken in Dutch, judges from the Netherlands, Taiwan, and Japan were also able to recognize each intended emotion with better-than-chance accuracy (van Bezooijen et al, 1983), although there were differences in level of accuracy as a function of emotion and language. More recently, Scherer et al (2001) recorded German actors (two male, two female) portraying five emotions (joy, anger, sadness, fear, and neutral). The voice samples were elicited using a scenario approach in which the actors were given text descriptions of emotive scenarios and asked to act them out (e.g. death of a loved one) in the performance of two standard sentences. These sentences were designed to circumvent the issue of semantic meaning by combining phonemes from six different languages (German, English, French, Italian, Spanish, and Danish; e.g. '*Fee gott laish jonkill gosterr*').

By enlisting collaborators in nine countries (Germany, Switzerland, Great Britain, the Netherlands, Italy, France, Spain, the United States, and Indonesia), emotion judgements from groups of participants in each country were obtained ($n$ ranged from 32 to 70, total = 428; age 18–30). All materials were translated into the language of each country. Participants were asked to focus on the emotion rather than the content of the sentence. After each utterance, they were given six seconds to select up to two emotion labels on the list of five target emotions. They were urged to select ones they thought best described the emotion the actor was trying to convey.

The authors reported an accuracy of 66 per cent across all emotions and all countries. Rate of accuracy was affected by country, type of emotion, and gender of the speakers. Confusion matrices were generated to assess patterns of errors as well as accuracy. German listeners had the highest rates of accuracy (Cohen's Kappa = .67, $p < .001$). Listeners from the other European countries also did well, with accuracy ranging from .52 to .62 (Cohen's Kappa, $p < .001$). Listeners from the one non-European country, Indonesia, had the lowest accuracy rate (Cohen's Kappa = .39, $p < .001$). The highest recognition rate was associated with expressions of anger; the lowest recognition rate was associated with expressions of joy.

Elfenbein and Ambady (2002) conducted a meta-analysis of 97 experiments exploring cross-cultural recognition of emotion in visual and auditory modes. In each study analysed, emotional stimuli (speech, facial expression) from at least one culture were presented to members and non-members of that culture. Cross-cultural emotion recognition accuracy was lower in studies of prosody than in studies of facial expression and body language, although most studies of prosody reported better-than-chance rates of recognition. Very few cross-cultural studies of emotional prosody have involved non-Western languages.

Examination of non-Western languages is important, because there is a long history of close contact between speakers of Western languages that might explain similar uses of prosody. Similar uses of prosody in Swedish and Norwegian, for example, can be explained by the history of interaction between speakers of these languages, and would not implicate universal principles of emotional prosody. Similar uses of prosody in

Western and non-Western languages, on the other hand, would provide compelling evidence for universal principles of emotional prosody.

Elfenbein and Ambady (2003) proposed a *cultural proximity hypothesis* to predict how well people of different cultures recognize emotional expression. According to their hypothesis, members of cultures who share cultural elements such as degree of individualism or collectivism, power structure, and gender roles, should be more successful at decoding each other's emotional expressions than members of cultures that are less similar. The cultural proximity hypothesis predicts, for example, that Japanese people should be better at recognizing the emotional expressions of Chinese people than the emotional expressions of North Americans, because Japanese and Chinese cultures are more similar to each other on relevant dimensions than Japanese and American cultures. Conversely, it predicts that English-speaking listeners should find it difficult to decode emotional prosody in Japanese and Chinese speech. This prediction contrasts with that of the cue-redundancy model, which suggests that psychophysical cues associated with emotions will allow individuals to decode emotions across cultural boundaries, including individualist and collectivist cultural boundaries.

Thompson and Balkwill (2006) investigated how well certain prosodic cues (frequency, intensity, and event density) were associated with the ability of English speakers to recognize emotion in five languages: English, German, Mandarin Chinese, Tagalog, and Japanese. Recordings were obtained of semantically neutral sentences (e.g. 'The bottle is on the table') spoken in a way that communicated each of four intended emotions: joy, sadness, anger, and fear. The use of prosody was not exaggerated or dramatic, but merely typical for speakers of each language. The five languages represent two members of the Germanic branch of the Indo-European language family (English and German), one language from the Sino-Tibetan family (Mandarin-Chinese), one language from the Altaic family (Japanese), and one from the Austronesian family (Tagalog) (Katsiavriades & Qureshi, 2003). German and English are Western languages spoken in individualistic societies, whereas Japanese and Chinese are Asian languages spoken in collectivist societies. Tagalog has been influenced by both the colonizing languages of Spanish and English, as well as by the languages of its Asian neighbours, Japan, Korea, and China.

Three predictions were evaluated. First, assuming that sensitivity to emotional prosody is partially dependent on psychophysical cues, listeners should be able to decode emotional prosody in any language at rates higher than that predicted by chance. Second, based on the idea that certain prosodic cues of emotion are culture specific, listeners should have higher rates of recognition for emotional prosody in their own language. Third, in view of previous research on emotional prosody, we predicted that listeners should be able to decode anger and sadness more reliably than joy and fear in all five languages (Johnstone & Scherer 2000).

English-speaking listeners recognized the four emotions at a rate significantly better than chance in all five languages, achieving the highest recognition scores within their own language. The analysis of speech stimuli revealed several psychophysical cues associated with specific emotions, providing potential cues to the decoding of emotional meaning. Regression analyses confirmed the predictive power of these cues for emotive judgements. Across languages, the ability to decode emotions was associated with psychophysical cues such as mean frequency, intensity, and event density.

We also assessed the ability of adult, native speakers of Mandarin Chinese and Japanese to identify intended emotions in English, Chinese, and Japanese speech samples (Balkwill & Thompson, 2005; Balkwill, 2006). Emotional decoding from these listeners was compared with the results from our English-speaking listeners. Again, semantically neutral sentences were used so that emotional intentions could be determined only by attending to prosodic aspects of speech.

Chinese and Japanese groups recognized anger, joy, and sadness at rates that were higher than chance performance in all five languages. Across languages and listener groups, the ability to decode emotions was associated with acoustic qualities such as mean frequency, intensity, and event density. Although both groups recognized fear above chance in English and Chinese, Japanese listeners' recognition of fear fell below chance in their own language. More data are needed to draw strong conclusions from such findings, but it is interesting to note that research on facial expressions indicates that Japanese judges also have difficulty decoding fear from Japanese facial expressions (Biehl et al, 1997; Matsumoto, 1992, 1996; Russell, Suzuki, & Ishida, 1993). It has been suggested that the emphasis in Japan on group cohesion and harmony discourages acknowledgement of emotions that threaten group harmony (Matsumoto, 1996).

AU: Bill and Laura-Lee pls check at proof stage whether this should be three or five.

We were able to address the cultural proximity hypothesis by comparing recognition rates for English, Chinese, and Japanese listeners. Support for the hypothesis would be evident if: (a) all listener groups are better at recognizing emotion in their own language than in other languages (even when verbal cues provide no information); (b) Chinese listeners are better than English-speaking listeners at recognizing emotion in Japanese speech; and (c) Japanese listeners are better than English listeners at recognizing emotion in Chinese speech. China and Japan have a long history of interaction between their peoples in times of peace and conflict (Rose, 2005). Emotional display rules in China and Japan are also similar in that negative emotions are considered inappropriate to display in public. North Americans, in contrast, have a higher tolerance of negative emotional displays (Eid & Deiner, 2001; Matsumoto, 1990).

Only our English-speaking listeners showed an in-group advantage across emotion categories. Chinese and Japanese listeners had an in-group advantage for the recognition of joy in their own language, but on the whole did not decode emotion better in their own language than in the English language. The high recognition of emotion in English speech by Chinese and Japanese listeners was not surprising, as all participants had exposure to English speech during their secondary school English-language training, as well as ongoing exposure through television, movies, and other media (Albas et al, 1976; Rosenthal et al, 1979; Scherer et al, 2001). That is, passive exposure to English speech among Japanese and Chinese listeners may have overwhelmed any decoding advantage arising from cultural proximity.

Across languages, judgements of joy were associated with greater range of fundamental frequency; judgements of sadness were associated with lower mean intensity, lower range of intensity, and lower event density; and judgements of anger were associated with greater mean and range of intensity, and greater event density.

Consistent with the cue-redundancy model, the type of acoustic parameter and the direction of association with emotion judgements were often similar in Canadian, Japanese, and Chinese listeners.

## 27.7 Summary and future prospects

Cross-cultural studies of music and speech prosody suggest that emotional communication occurs through a complex interplay between system-specific and psychophysical cues. Psychophysical cues have similar connotations across cultures (Balkwill & Thompson, 1999; Balkwill, Thompson, & Matsunago, 2004) and similar functions in music and speech prosody (Ilie & Thompson, 2006; Juslin & Laukka, 2003; Patel, 2008). Reliance on psychophysical influences wanes throughout the lifespan, with enculturation gradually leading to increased sensitivity to culture-specific and channel-specific cues. Enculturation can lead to an understanding of abstract cues that have no psychophysical basis, but it can also influence the interpretation of psychophysical cues themselves, overlaying their significance with culture-specific nuance or altering their interpretation entirely.

To summarize, the FES model contains three components, as follows:

1. *Phylogenetic base*. Infants start life with an understanding of basic psychophysical cues to emotion. Psychophysical cues are determined by system constraints that may include genetic factors, but may also be determined by ubiquitous environmental and biophysical constraints. Musical and prosodic channels of communication share these psychophysical cues.
2. *Ontogenetic process*. Throughout development is a process of refinement and fractionation of different emotional communication systems. This fractionation applies not only to the modalities of music and prosody (which fractionate from each other), but also to musical and prosodic systems *across cultures*. That is, the process of fractionation permits the emergence of culture-specific systems of emotional coding, some prosodic and others musical. System-specific cues are learned through exposure to the conventions of one's culture, giving rise to culture-specific understandings of emotional communication.
3. *Cross-cultural communication*. Psychophysical cues remain available for decoding at all stages of development, with culture-specific and domain-specific cues overlaid increasingly throughout development. When an individual attempts to decode an emotional message (from prosody or music) across cultural boundaries, it is usually optimal to focus attention on psychophysical cues while ignoring culture-specific cues. Culture-specific cues can mask the ability to attend to psychophysical cues, and emotional communication is weaker across cultures than within cultures because psychophysical cues provide only one part of the

full emotional code. Nonetheless, it generally enables above-chance decoding of emotional messages.

Research is needed to evaluate the capacity for enculturation to alter the emotional connotation of psychophysical cues entirely. Existing data suggest that this capacity is either limited or seldom realized: individuals at all stages of development are capable of decoding basic emotions such as anger, joy, and sadness in unfamiliar music or speech, because both channels of emotional communication contain a number of psychophysical cues that transcend cultural boundaries, including pitch level and variability, sound intensity, rate (tempo), timbre, and complexity. Difficulties in decoding do occur, but they usually arise because culture-specific conventions and cues mask the detection of psychophysical cues. Training listeners to selectively attend to psychophysical cues may lead to improved cross-cultural emotional communication in music and speech prosody.

Further cross-cultural studies are also needed to classify the many known emotional cues as either psychophysical and having a phylogenetic basis, or culture-specific and having an ontogenetic basis. Juslin and Laukka (2003) have identified a large number of musical and prosodic cues such as jitter, vibrato, formant structure, the presence of pauses, onset characteristics, and articulation. More recently, Juslin and Västfjäll (2008) proposed a set of mechanisms for understanding emotional induction by music (see Chapter 22, this volume). One such mechanism—the *brain stem* response—is likely to respond in similar ways across all cultures to psychophysical features such as high intensity. Nonetheless, the potential role of this mechanism for cross-cultural emotional communication has yet to be fully understood.

How might one account for the existence of culture-transcendent connections between psychophysical cues and emotions? Huron (2006) provided a number of evolutionary explanations, and Scherer's (1986) *component process model* describes physiological changes that occur in the vocal apparatus during the expression of emotions. For example, speaking of something that is deeply unpleasant often manifests as faucal and pharyngeal constriction and a tensing of the vocal tract. The acoustic outcome is higher-frequency energy. Based on this model, Banse and Scherer (1996) predicted several associations between the acoustic attributes of speech and emotional intent.

Differences between expression and recognition of emotion in music and speech are also of interest. Are some emotions easy to convey through speech prosody, but difficult to convey in music? One might speculate that because speech carries the weight of semantic as well as emotional meaning, it may be more susceptible to within-culture norms as well as in-group and out-group biases. The FES framework accounts for the dynamic interplay between psychophysical cues and culture-specific conventions throughout development, and complements the CRM and other models of emotional communication. For example, the Brunswikian lens model (Brunswik, 1956) provides another valuable framework for understanding the ability to decode emotional meaning from either speech prosody (Scherer 1982) or music (Juslin, 2000). Originally designed to describe visual perception, the model has been adapted to many types of human judgements (see also Chapter 17, this volume).

In a Brunswikian framework the intent of the encoder to express an emotion is facilitated by the use of a large set of cues that are probabilistic and partially redundant. Each cue by itself is not a reliable indicator of the expressed emotion, but combined with other cues, each one contributes in a cumulative fashion to the communication and recognition of emotion. The model incorporates the flexibility of the decoder to shift expectations from unavailable to available cues (Juslin, 2000). The CRM and FES are consistent with this model, but focus on the significance of psychophysical cues for the expression of emotions, their interaction with culture-specific convention that can enhance or hinder the ability of listeners to decode emotions across cultures, and the gradual fractionation of emotional communication systems across cognitive domains and cultural environments.

## Recommended further reading

1. Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–35.
2. Juslin, P. N., & Västfjäll, D. (2008a). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–75.
3. Reyes, A. (2009). What do ethnomusicologists do? An old question for a new century. *Ethnomusicology*, *53*, 1–17.

## References

Albas, D. C., McCluskey, K. W., & Albas, C. A. (1976). Perception of the emotional content of speech: A comparison of two Canadian groups. *Journal of Cross-Cultural Psychology 7*, 481–90.

Bachorowski, J., & Owren, M. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, *6*, 219–24.

Balkwill, L-L. (2006). *Perceptions of emotion in music across cultures*. Paper presented at Emotional Geographies: The Second International & Interdisciplinary Conference, May 2006, Queen's University, Kingston, Canada.

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.

Balkwill, L-L., & Thompson, W. F. (2005). *Recognition of emotion in five languages*. Paper presented at Auditory Perception Cognition and Action Meeting, November 2005, Toronto, Canada.

Balkwill, L-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, North Indian and Western music by Japanese listeners. *Japanese Journal of Psychological Research*, *46*, 337–49.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–36.

Biehl, M., Matsumoto, D., Ekman, P., & Hearn, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*, *21*, 3–21.

Bezooijen, R. van., Oto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology*, *14*, 387–406.

Boucher, J. D. (1979). Culture and emotion. In A. J. Marsella, R. G. Tharp, & T. V. Ciborowski (eds), *Perspectives on cross-cultural psychology* (pp. 159–78). San Diego, CA: Academic Press.

Briggs, J. L. (1970). *Never in anger*: *Portrait of an Eskimo family*. Cambridge, MA: Harvard University Press.

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

Carlsen, J. C. (1981). Some factors which influence melodic expectancy. *Psychomusicology*, *1*, 12–29.

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

Clayton, M. (2001). Introduction: towards a theory of musical meaning (in India and elsewhere). *British Journal of Ethnomusicology*, *10*, 1–18.

Cross, I. (2003). Music, cognition, culture and evolution. In I. Peretz & R. J. Zatorre (eds), *The cognitive neuroscience of music* (pp. 42–56). Oxford: Oxford University Press.

Cross, I. (2006). The origins of music: Some stipulations on theory. *Music Perception*, *24*, 79–82.

Crowder, R. G., Resnick, J. S., & Rosenkrantz, S. L. (1991). Perception of the major/minor distinction: V. Preferences among infants. *Bulletin of the Psychonomic Society*, *29*, 187–8.

Cuddy, L. L., Cohen, A., & Mewhort, D. J. (1981). Perception of structure in short melodic sequences. *Journal of Experimental Psychology*: *Human Perception and Performance*, *7*, 869–83.

Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*, B1–B10.

Darwin, C. (1872). *Expression of the emotions in man and animals*. London: John Murray.

Deva, B. C., & Virmani, K. G. (1975). *A study in the psychological response to ragas. (Research Report II of Sangeet Natak Akademi)*. New Delhi, India: Indian Musicological Society.

de Villiers-Sidani, E., Chang, E. F., Bao, S., & Merzenich, M. M. (2007). Critical period for spectral tuning defined in the primary auditory cortex (A1) in the rat. *Journal of Neuroscience*, *27*, 180–9.

Dissanayake, E. (2001). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (eds), *The origins of music* (pp. 389–410). Cambridge, MA: MIT Press.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. New York: Academic Press.

Eerola, T., Himberg, T., Toiviainen, P., & Louhivuori, J. (2006). Perceived complexity of western and African folk melodies by western and African listeners. *Psychology of Music*, *34*, 337–71.

Eid, M., & Diener, E. (2001). Norms for experiencing emotions in different cultures: Inter- and intranational differences. *Journal of Personality and Social Psychology*, *81*, 869–85.

Ekman, P. (1970). Universal facial expressions of emotion. *California Mental Health Research Digest*, *8*, 151–8.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (ed.), *Nebraska symposium on motivation* (pp. 207–83). Lincoln, NE: University of Nebraska Press.

Ekman, P. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

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

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotions. *Science*, *164*, 86–8.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*, 203–35.

Elfenbein, H. A., & Ambady, N. (2003). Cultural similarity's consequences: A distance perspective on cross-cultural differences in emotion recognition. *Journal of Cross Cultural Psychology*, *34*, 92–109.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*, 412–29.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Friederici, A. D., & Koelsch, S. (in press). Universal recognition of three basic emotions in music. *Current Biology*.

Fritz, T., Sammler, D., & Koelsch, S. (2006). How far is music universal? An intercultural comparison. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (eds), *9th International Conference on Music Perception and Cognition* (p. 88). Bologna, Italy: Bononia University Press.

Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The interpretation of cultures*: *Selected essays* (pp. 3–30). New York: Basic Books.

Gundlach, R. H. (1932). A quantitative analysis of Indian music. *American Journal of Psychology*, *44*, 133–45.

Gundlach, R. H. (1935). Factors determining the characterization of musical phrases. *American Journal of Psychology*, *47*, 624–43.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, *11*, 466–72.

Hebdige, D. (1979). *Subculture*: *The meaning of style*. London: Routledge.

Heyes, C. (2003). Four routes of cognitive evolution. *Psychological Review*, *110*, 713–27.

Huron, D. (2006). *Sweet anticipation*: *Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*, 319–29.

Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition & Emotion*, *3*, 81–123.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. M. Haviland-Jones (eds), *Handbook of emotions* (2nd edn, pp. 220–35). New York: Guilford Press.

Juslin, P. N. (2000). Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology*: *Human Perception and Performance*, *26*, 1797–813.

Juslin, P. N. (2001). Communicating emotion in music performance. A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (eds), *Music and emotion*: *Theory and research* (pp. 309–37). Oxford: Oxford University Press.

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

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–75.

Justus, T., & Hutsler, J. J. (2005). Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain-specificity. *Music Perception*, *23*, 1–27.

Katsiavriades, K., & Qureshi, T. (2003). Ten language families in detail. Retrieved from: http://www.Krysstal.com

Kessler, E. J., Hansen, C., & Shepard, R. N. (1984). Tonal schemata in the perception of music in Bali and the west. *Music Perception*, *2*, 131–65.

Kilbride, J. E., & Yarczower, M. (1983). Ethnic bias in the recognition of facial expressions. *Journal of Nonverbal Behavior*, *8*, 27–41.

Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition & Emotion*, *16*, 29–59.

Krumhansl, C. L. (1995). Music psychology and music theory: Problems and prospects. *Music Theory Spectrum*, *17*, 53–80.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. *Cognition*, *76*, 13–58.

Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory & Cognition*, *10*, 243–51.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford: Oxford University Press.

Lerdahl, F., & Krumhansl, C. L. (2006). Modeling tonal tension. *Music Perception*, *24*, 329–66.

Livingstone, S., & Thompson, W. F. (in press). The emergence of music from the Theory of Mind. *Musica Scientiae*.

Lomax, A. (1962). Song structure and social structure. *Ethnology*, *1*, 425–51.

Lomax, A. (1975). The good and the beautiful in folksong. *Journal of American Folklore*, *80*, 213–235.

Lomax, A. (1975). *Cantometrics: An approach to the anthropology of music. Audiocassettes and a handbook*. Berkeley, CA: University of California Media Extension Center.

Lomax, A. (1978). *Folk song style and culture*. Edison, NJ: Transaction Publishers.

Lutz, C. (1988). *Unnatural emotions: everyday sentiments on a Micronesian atoll and their challenge to western theory*. Chicago, IL: University of Chicago Press.

Lynch, M. P., & Eilers, R. E. (1991). Children's perception of native and nonnative musical scales. *Music Perception*, *9*, 121–31.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology*, *81*, 922–34.

Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology*, *20*, 92–105.

Matsumoto, D. (1990). Cultural similiarities and differences in display rules. *Motivation and Emotion*, *14*, 195–214.

Matsumoto, D. (1992). American-Japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology*, *23*, 72–84.

Matsumoto, D., Yoo, S. H., Hirayama, S., & Petrova, G. (2005). Development and validation of a measure of display rule knowledge: the Display Rule Assessment Inventory. *Emotion*, *5*, 23–40.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, *112*, 197–204.

Mesquita, B., Frijda, N. H., & Scherer, K. R. (1997). Culture and emotion. In J. W. Berry, P. R. Dasen, & T. S. Saraswathi (eds), *Handbook of cross-cultural psychology*, *Vol. 2*: *Basic processes and human development* (2nd edn, pp. 255–97). Needham Heights, MA: Allyn & Bacon.

Morey, R. (1940). Upset in emotions. *Journal of Social Psychology*, *12*, 333–56.

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*, 455–64.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures*: *the Implication-Realization Model*. Chicago, IL: Chicago University Press.

Nettl, B. (1983). *The study of ethnomusicology*. Chicago, IL: University of Illinois Press.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*, 291–310.

Palmer, C., & Hutchins, S. (2006). What is musical prosody? *The Psychology of Learning and Motivation*, *46*, 245–78.

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

Rose, C. (2005). *Sino-Japanese relations*: *facing the past*, *looking to the future*. London: Routledge.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. I., & Archer, D. (1979). *Sensitivity to nonverbal communication*: *the PONS test*. Baltimore, MD: Johns Hopkins University Press.

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, *110*, 426–50.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of cross-cultural studies. *Psychological Bulletin*, *115*, 102–41.

Russell, J. A., Suzuki, N., & Ishida, N. (1993). Canadian, Greek, and Japanese freely produced emotion labels for facial expressions. *Motivation and Emotion*, *17*, 337–51.

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

Schellenberg, E. G. (1996). Expectancy in melody: Tests of the implication-realization model, *Cognition*, *58*, 75–125.

Scherer, K. R. (1979). Non-linguistic vocal indicators of emotion and psychopathology. In C. E. Izard (ed.), *Emotions in personality and psychopathology* (pp. 493–529). New York: Plenum Press.

Scherer, K. R. (1982). Methods of research on vocal communication: Paradigms and parameters. In K. R. Scherer & P. Ekman (eds), *Handbook of methods in nonverbal behavior research* (pp. 136–98). Cambridge, UK: Cambridge University Press.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*, 143–65.

Scherer, K. R. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. Wagner & A. Manstead (eds), *Handbook of social psychophysiology* (pp. 165–97). New York: Wiley.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*, 331–46.

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, *3*, 365–94.

Sorenson, E. R. (1975). Culture and the expression of emotion. In T. R. Williams (ed.), *Psychological anthropology* (pp. 361–72). Chicago, IL: Aldine.

Standke, R. (1992). *Methods of digital speech analysis in research on vocal communication*. Frankfurt, Germany: Peter Lang.

Thompson, W. F., & Balkwill, L-L. (2006). Decoding speech prosody in five languages. *Semiotica*, *158*, 407–24.

Thompson, W. F., Graham, P., & Russo, F. A. (2005). Seeing music performance: Visual influences on perception and experience. *Semiotica*, *156*, 203–27.

Thompson, W. F., & Robitaille, B. (1992). Can composers express emotions through music? *Empirical Studies of the Arts*, *10*, 79–89.

Thompson, W. F., Russo, F. A., & Quinto, L. (in press). Audio-visual integration of emotional cues in song. *Cognition & Emotion*.

Thompson, W. F., & Schellenberg, E. G. (2006). Listening to music. In R. Colwell (ed.), *MENC handbook of music cognition & development* (pp. 72–123). Oxford: Oxford University Press.

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

Thompson, W. F., & Stainton, M. (1998). Expectancy in Bohemian folksong melodies: Evaluation of implicative principles for implicative and closural intervals. *Music Perception*, *15*, 231–52.

Tooby, J., & Cosmides, L. (1990). The past explains the present: Emotional adaptations and the structure of ancestral environments. *Ethology and Sociobiology*, *11*, 375–424.

Trainor, L. J., & Trehub, S. E. (1994). Key membership and implied harmony in Western tonal music: Developmental perspectives. *Perception & Psychophysics*, *56*, 125–32.

Trainor, L. J., & Heinmiller, B. M. (1998). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior & Development*, *21*, 77–88.

Trehub, S. E. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker, & S. Brown (eds), *The origins of music* (pp. 427–48). Cambridge, MA: MIT Press.

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience*, *6*, 669–73.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms. *Cognition*, *100*, 73–99.

Trehub, S. E., & Nakata, T. (2002). Emotion and music in infancy. *Musicae Scientiae*, *Special Issue 2001–2*, 37–61.

Trehub, S. E., Schellenberg, E. G., & Hill, D. S. (1997). The origins of music perception and cognition: A developmental perspective. In I. Deliège & J. A. Sloboda (eds), *Perception and cognition of music* (pp. 103–28). Hove, UK: Psychology Press.

Trehub, S. E., Unyk, A., & Trainor, L. J. (1993). Maternal singing in cross-cultural perspective. *Infant Behaviour and Development*, *16*, 285–95.

Unyk, A. M., Trehub, L. J., & Schellenberg, E. G. (1992). Lullabies and simplicity: A cross-cultural perspective. *Psychology of Music*, *20*, 15–28.

Wittgenstein, L. (1953/2001). *Philosophical investigations*. Oxford: Blackwell.

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