Emotion
2013, Vol. 13, No. 3, 434–449

© 2013 American Psychological Association
1528-3542/13/$12.00 DOI: 10.1037/a0031388

# Universal and Culture-Specific Factors in the Recognition and Performance of Musical Affect Expressions

Petri Laukka
Stockholm University

Tuomas Eerola
University of Jyväskylä

Nutankumar S. Thingujam
Sikkim University

Teruo Yamasaki
Osaka-Shoin Women's University

Grégory Beller
Institut de Recherche et Coordination Acoustique/Musique (IRCAM)

We present a cross-cultural study on the performance and perception of affective expression in music. Professional bowed-string musicians from different musical traditions (Swedish folk music, Hindustani classical music, Japanese traditional music, and Western classical music) were instructed to perform short pieces of music to convey 11 emotions and related states to listeners. All musical stimuli were judged by Swedish, Indian, and Japanese participants in a balanced design, and a variety of acoustic and musical cues were extracted. Results first showed that the musicians' expressive intentions could be recognized with accuracy above chance both within and across musical cultures, but communication was, in general, more accurate for culturally familiar versus unfamiliar music, and for basic emotions versus nonbasic affective states. We further used a lens-model approach to describe the relations between the strategies that musicians use to convey various expressions and listeners' perceptions of the affective content of the music. Many acoustic and musical cues were similarly correlated with both the musicians' expressive intentions and the listeners' affective judgments across musical cultures, but the match between musicians' and listeners' uses of cues was better in within-cultural versus cross-cultural conditions. We conclude that affective expression in music may depend on a combination of universal and culture-specific factors.

*Keywords:* cross-cultural, emotion recognition, in-group advantage, music performance, music feature extraction

*Supplemental materials:* http://dx.doi.org/10.1037/a0031388.supp

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The question whether the expressiveness of one culture's music can be recognized by listeners from different musical cultures has been the source of much discussion (e.g., Higgins, 2012; Walker, 1996). On the one hand, music as a phenomenon is present in virtually all cultures (Merriam, 1964), and the basic building blocks of music, such as pitch and durations, seem to be organized according to largely universal principles (Brown & Jordania, in press; Higgins, 2006), which, in turn, are partly determined by the properties of the human auditory system (McDermott & Oxenham, 2008). On the other hand, it has been suggested that "music takes as many forms as culture" (Cross, 2008, p. 149), and many studies report culture-specific diversity with regard to perception and cognition of structural elements of music (Morrison & Demorest, 2009; Stevens, 2012).

Affect and emotion occupy a special place among music's expressive qualities, as evidenced by philosophical theories maintaining that an important value of music lies in its capacity to express emotions (e.g., Budd, 1985)—a view commonly shared by both music performers (Laukka, 2004) and music listeners (Juslin & Laukka, 2004). A large body of research suggests that musicians can communicate certain emotions to listeners by varying aspects of their performances in relatively emotion-specific ways (Juslin & Timmers, 2010). Several structural elements of music (e.g., mode) are also related to listeners' affective perception of musical works (Gabrielsson & Lindström, 2010). For example, happiness is often conveyed by fast tempo, staccato articulation, moderately high sound level, high

This article was published Online First February 11, 2013.
Petri Laukka, Department of Psychology, Stockholm University, Stockholm, Sweden; Tuomas Eerola, Department of Music, Finnish Centre of Excellence in Interdisciplinary Music Research, University of Jyväskylä, Jyväskylä, Finland; Nutankumar S. Thingujam, Department of Psychology, Sikkim University, Gangtok, India; Teruo Yamasaki, Faculty of Psychology, Osaka-Shoin Women's University, Nara, Japan; Grégory Beller, Institut de Recherche et Coordination Acoustique/Musique (IRCAM), Paris, France.
The research was supported by the Swedish Research Council through grant 2006-1360 to Petri Laukka.
Correspondence concerning this article should be addressed to Petri Laukka, Stockholm University, Department of Psychology, 10691 Stockholm, Sweden. E-mail: petri.laukka@psychology.su.se

pitch, consonant harmony, and a major mode, whereas sadness is conveyed by slow tempo, legato articulation, low sound level, low pitch, dissonant harmony, and a minor mode.

The relations between the strategies that musicians use to convey various expressions, and listeners' perceptions of the affective content of the music, can be illustrated by Brunswik's (1956) lens model, which is a widely used framework for studying human judgment processes. As shown in Figure 1, a performer (encoder) can express affect by using a set of probabilistic (i.e., uncertain) and partly redundant cues, which are then utilized by listeners (decoders) to judge the expression of the performance. Each cue in itself is not a perfectly reliable indicator of the expressed affective character, and decoders therefore have to combine many cues for successful communication to occur. Also, because each cue is partly redundant, more than one way of using the cues may lead to accurate communication. Such flexibility further allows performers to exchange different cues for another, and makes it possible to convey affect expressions on a wide range of instruments and employing idiosyncratic and culture-specific expressive styles (Juslin, 2000).

Music has many important acoustic attributes in common with speech, and many of the cues associated with musical expression (e.g., pitch, loudness, timbre, and tempo) are also important for vocal expression (i.e., expression of affect through the nonverbal aspects of speech). Many authors have noted that musical and vocal expressions could share a common expressive "code" and evidence from comprehensive reviews support this hypothesis (Juslin & Laukka, 2003; Scherer, 1995). Thus, the tendency to hear emotion in music may partly stem from similarities between the acoustic attributes of the music and aspects of emotional human behavior.[1] Scherer (1986) proposed that vocal expressions largely result from physiological responses (e.g., autonomic activation) to the appraisal of emotional situations and their resulting effect on the speech production apparatus. This implies that aspects of vocal affect expressions can be expected to exhibit phylogenetic continuity, and many studies have reported accurate cross-cultural communication of vocal expressions (e.g., Pell, Paulmann, Dara, Alasseri, & Kotz, 2009; Sauter, Eisner, Ekman, & Scott, 2010; Scherer, Banse, & Wallbott, 2001; Thompson & Balkwill, 2006; see Juslin & Laukka, 2003, for a review). Thus, to the extent that musicians express affect using a nonverbal code that derives from vocal expression, and to the extent that vocal expressions are universal, we might expect to find universality also for affective expressions conveyed by music.

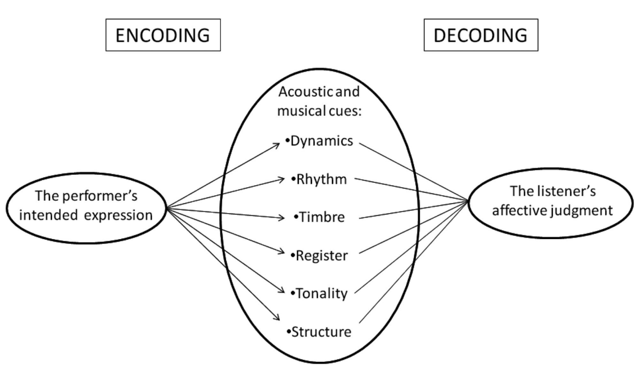


*Figure 1.* A lens model of musical communication.

Most studies on affect expression in music report data collected from a culturally narrow range of musicians and listeners (typically Western art music judged by Western listeners), and evidence for cross-cultural communication remains limited (for reviews, see Davies, 2011; Thompson & Balkwill, 2010). Nevertheless, there is tentative evidence that people can understand the affective expression of musical excerpts from unfamiliar musical traditions. For example, Balkwill and Thompson (1999) let 30 Western listeners judge the expression of 12 Hindustani ragas intended to express anger, happiness, peacefulness, and sadness, and found that the listeners were sensitive to the intended expression of the ragas. Also, Fritz et al. (2009) showed that listeners from a native African population—who were naïve to Western music—were able to recognize fear, happiness, and sadness from Western music with accuracy above chance.

Studies on cross-cultural communication of (mainly) facial expressions have documented evidence for an *in-group advantage*—whereby individuals perform better when judging emotional stimuli from their own culture versus other cultures (Elfenbein & Ambady, 2002). It is important to use a balanced design—where stimuli from each culture are judged by individuals from each culture—when testing a possible in-group advantage, otherwise cultural effects cannot be separated from other group effects. However, previous studies of musical affect recognition have relied on designs where judges from different cultures rate music from one culture (Adachi, Trehub, & Abe, 2004; Kwoun, 2009; Wieczorkowska, Datta, Sengupta, Dey, & Mukherjee, 2010; Zacharopoulou & Kyriakidou, 2009), or judges from one culture rate music from different cultures (Balkwill, Thompson, & Matsunaga, 2004; Hoshino, 1996; Keil & Keil, 1966).[2]

According to the *dialect theory* of emotion (Elfenbein, Beaupré, Lévesque, & Hess, 2007), the in-group advantage results from subtle cultural differences in expression style, and the lower recognition of out-group stimuli has been directly linked to cultural differences in expression style in studies of facial expressions (Dailey et al., 2010; Elfenbein et al., 2007). In a similar vein, Thompson and Balkwill (2010; Balkwill & Thompson, 1999) proposed that affective expression in music is achieved through a combination of two partly redundant sources of information, namely, universal and culture-specific cues. Universal cues refer to elements of music that can be understood without musical enculturation, and listeners should be able to infer affect from unfamiliar musical cultures to the extent that universal cues to affect are present. When culture-specific cues to affect are present, however,

[1] In this article, we will focus on the hypothesis that music can express affect by imitating expressive features of affective speech because this view has received considerable empirical support (e.g., Juslin & Laukka, 2003). However, it should be noted that expression in music may also derive from symbolic similarities with other forms of human emotional behavior such as "the gait, attitude, air, carriage, posture, and comportment of the human body" (Davies, 2006, p. 182).

[2] A few studies have used balanced designs, where judges from Western and Eastern cultures judge both Western and Eastern music (Darrow, Haack, & Kuribayashi, 1987; Gregory & Varney, 1996), but none of these assessed emotion recognition accuracy and also confused expressed emotion with experienced emotion (see Davies, 2011, for an insightful critique of this literature). Also, Fritz et al. (2009) included an experiment where Western and Mafa listeners rated the valence of both Western and Mafa music, but they did not assess emotion recognition accuracy in that experiment.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

they are best understood by listeners who are familiar with the particular musical style, and may thus contribute to an in-group advantage in recognition accuracy. Taken together, this emphasizes the need to study both the encoding and decoding sides of the communication chain in order to determine if affective character is expressed and interpreted similarly or differently across cultures. Importantly, cultural differences may exist (a) in the way different affective characters are portrayed (resulting in different cue configurations), (b) in the way members of different cultures perceive these features and the inferences they draw from them, or (c) in both encoding and decoding (Scherer, Clark-Polner, & Mortillaro, 2011).

Previous cross-cultural music studies have mainly focused on decoding and few studies have investigated cultural effects on encoding. Bowling, Sundararajan, Han, and Purves (2012) recently compared acoustic cues in Indian and Western music and speech, and reported that positive/excited emotion was encoded using larger melodic/pitch intervals and more high-frequency spectral energy when compared with negative/subdued emotion in both speech and music across cultures. A few studies have also attempted to link decoding and encoding by relating listener judgments of emotion with subjective ratings of performance cues (Adachi et al., 2004; Balkwill & Thompson, 1999; Balkwill et al., 2004; Zacharopoulou & Kyriakidou, 2009). However, no previous studies have investigated the associations between encoding and decoding using objective acoustic measures.

In the present study, our first aim was to investigate if emotions and related states can be communicated through music within and across different musical cultures. By using a balanced design, we present the first test of whether affect recognition is more accurate for culturally familiar music than for unfamiliar music. Musicians from three distinct musical traditions (Swedish folk music, Hindustani classical music, and Japanese traditional music) were instructed to perform short pieces of music to convey various expressions to listeners, and the resulting performances were evaluated by listeners from each culture (Sweden, India, and Japan). We hypothesized that familiarity with a specific musical culture would give rise to an in-group advantage, with more accurate affect recognition for familiar versus unfamiliar music. Most previous research on music and emotion has been conducted using Western music and, for the sake of consistency, we also included Western classical music as a control condition. Because Western music is ubiquitous in large parts of the world, and because musical structural knowledge can be implicitly acquired through mere exposure to a particular type of music (e.g., Huron, 2006; Tillmann, Bharucha, & Bigand, 2000; Wong, Roy, & Margulis, 2009), we expected listeners from all cultures to be familiar with Western classical music. Consequently, we expected cross-cultural differences in affect recognition to be small or nonexistent for this condition. Our second aim was to relate encoding and decoding in a lens-model analysis—using state-of-the-art analysis of musical features—in order to investigate how consistently musicians and listeners make use of various cues across musical cultures. Here we hypothesized that the matching between encoders' and decoders' uses of cues (see Figure 1) would be less accurate for cross-cultural, compared with within-cultural, conditions.

Whereas most previous studies focused on a limited number of emotions, we included a larger than usual selection of emotions and emotion-related states relevant to musical expressivity, based on a careful reading of the literature on music and emotion (e.g., Collier, 2002; Gabrielsson & Juslin, 2003; Hevner, 1936; Juslin & Laukka, 2004; Laukka, 2004; Wedin, 1969; Zentner, Grandjean, & Scherer, 2008). First, we included basic emotions—anger, fear, happiness, and sadness (Ekman, 1992)—which have been shown to be accurately communicated through both music and speech in previous cross-cultural studies (Balkwill et al., 2004; Fritz et al., 2009; Juslin & Laukka, 2003). We further included affection, peacefulness, and solemnity because previous within-cultural music studies have reported accurate recognition for these expressions (e.g., Gabrielsson & Juslin, 1996; Vieillard et al., 2008). Finally, we included humor, longing, and spirituality—which are considered important expressive qualities of music—although no previous studies have investigated if they actually can be conveyed by musical means. Based on the similarities between music and speech, we expected basic emotions to be accurately decoded both within and across cultures. The other selected expressions have rarely or never been included in cross-cultural studies on vocal expression, and it is thus an open question whether they can be musically conveyed across cultures or not.

# Method

## Musical Stimuli

Twelve professional musicians from four distinct musical cultures (Hindustani classical music, Japanese traditional music, Swedish folk music, and Western classical music) took part in the study (3 musicians/culture; mean age = 38.8 years; 5 women). The Swedish and Western classical performers were violinists, whereas the Indian and Japanese musicians performed on the sarangi and kokyū, respectively. The sarangi is a bowed-string instrument of India that plays an important part in Hindustani classical music (Bor, 1987), and the kokyū is a traditional Japanese instrument similar to a bowed lute (Hughes, 2001); both are held vertically in front of the chest. We chose to use bowed-string instruments because they allow the musicians to use a particularly wide range of expressive means. All musicians were well-established performers on their respective instrument, with extensive musical experience (years of experience as a professional musician—Hindustani musicians, $M = 20.3$ years; Japanese musicians, $M = 13.3$ years; Swedish musicians, $M = 28.3$ years; Western classical musicians, $M = 14.0$ years).

The musicians were instructed to express 11 different emotions and related states: affection, anger, fear, happiness, humor, longing, peacefulness, sadness, solemnity, spirituality, and neutral. It should be noted that this selection was influenced by empirical and theoretical work on musical expressivity (e.g., Gabrielsson & Juslin, 2003) rather than current psychological theories of emotion. Anger, fear, happiness, and sadness are generally considered basic emotions (e.g., Ekman, 1992), and are often conceptualized as rapid adaptive reactions to certain goal-relevant changes in our environment that consist of several components, including cognitive appraisal, physiological response, and expressive behavior. However, many of the other included expressions (especially spirituality) are hard to classify based on current psychological theorizing on emotion and we therefore refer to them collectively as "nonbasic" affective states, based on the premise that they denote

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

states with some sort of affective valence but that there is not sufficient evidence to consider them basic emotions (Matsumoto & Hwang, 2012).

Each musician chose one brief excerpt of music for each intended expression, from his or her respective musical tradition, and then performed the piece in a way that would convey the intended expression to a listener. The samples of Swedish folk music consisted of traditional tunes, mainly from the northern Uppland tradition, whereas the Hindustani excerpts were different ragas or sections of ragas, such as alap, jor, and jhala. The Japanese stimuli consisted of pieces from the kokyū repertoire, mainly from the Edo period in Japan (17th to 19th centuries), and the Western classical stimuli were culled from composers such as J. S. Bach, Bartok, Beethoven, John Cage, Saint-Saëns, Stravinsky, Tchaikovsky, and Vivaldi.[3]

Recording sessions were conducted in professional music studios, and the recording level was first optimized for each performer and then kept the same during the whole recording session. The Swedish recordings were conducted in Uppsala, the Indian recordings in Pune, the Japanese in Osaka, and the Western classical musicians were recorded in Paris, France. Instructions were translated from English into the musicians' native language using a translation/back-translation procedure. Each musician produced one performance for each intended expression, which yielded a total of 132 musical stimuli (12 musicians × 11 emotions/related states). The length of the stimuli in general varied between 30 s to 1 min.

## Participants and Procedure

Recognition experiments were conducted in three countries, and we utilized a balanced design where native student participants from each culture judged the expression of all musical stimuli (Sweden, $n = 30$, mean age = 27.2 years, 15 women; India, $n = 30$, mean age = 23.9 years, 15 women; Japan, $n = 27$, mean age = 22.4 years, 13 women). In a forced-choice design, the judges were instructed to choose, from among the 11 intended expressions, one label that best represented the expression conveyed by each musical excerpt. Responses were scored as correct if the response matched the intended expression of the musical excerpt. Although there are well-documented concerns regarding the forced-choice methodology (e.g., Frank & Stennett, 2001; Russell, 1994), it was adopted here for consistency with previous research on the in-group advantage in cross-cultural emotion recognition (e.g., Elfenbein et al., 2007).

Experiments were conducted individually using MediaLab software to present stimuli and record responses. The presentation order of the different stimulus sets, as well as the order of the stimuli within each culture, was randomized. The participants listened to stimuli through headphones with constant sound levels. The length of one experimental session was 1 to 1.5 h. Response choices were translated from English into the participants' native languages using a translation/back-translation procedure. Having finished the recognition experiment, the participants were also asked to rate—on scales ranging from 1 (*not at all familiar*) to 7 (*very familiar*)—how familiar they were with Swedish folk music, Hindustani classical music, Japanese traditional music, and Western classical music.

## Musical Feature Extraction

Each musical stimulus was analyzed using the MIR toolbox (Eerola, Lartillot, & Toiviainen, 2009; Lartillot & Toiviainen, 2007), which is a widely used software dedicated to the extraction of music-related features from audio recordings. The tool analyzes the acoustic signal using short overlapping frames (typically ranging from 25 ms to 100 ms, with 50% overlap), and the results are then summarized across frames and presented as means or standard deviations. The acoustic signal was filtered, using predefined filtering (e.g., spectral or temporal) for each feature, before extraction. The raw values for each feature and musician were further $z$-transformed before being entered into the statistical analyses in order to control for differences in baseline values between musicians.

For this study, we selected 26 features to represent aspects of music that have previously been associated with affect expression (see Gabrielsson & Lindström, 2010; Juslin & Laukka, 2003), and these features are briefly described in Table 1. We refer to Eerola (2011) for a detailed description of pertinent literature and pointers for each feature. The features can be broadly divided into *acoustic cues*, which represent basic acoustic-perceptual dimensions shared between music and speech (i.e., register/pitch, dynamics/loudness, timbre, and tempo) and which show similar affective connotations in both domains (e.g., Juslin & Laukka, 2003), and *musical cues*, which represent higher-level music-specific constructs such as tonality and musical structure. In a way of summary, the assumed overlap between speech and music for each feature is also given in Table 1.

# Results

## Familiarity Ratings

The participants' familiarity ratings were analyzed with a 4 (musical culture) × 3 (listener culture) mixed measures ANOVA. Significant effects of listener culture, $F(2, 84) = 5.56$, $p = .005$, $\eta_p^2 = .12$, and musical culture, $F(3, 252) = 24.10$, $p < .0001$, $\eta_p^2 = .22$, were qualified by a significant interaction, $F(6, 252) = 70.26$, $p < .0001$, $\eta_p^2 = .63$. Pairwise comparisons indicated that all listener groups rated music from their own culture (in-group music) and also, as expected, Western classical music as more familiar than music from the remaining two cultures (out-group music): Indian listeners, $ts(29) \geq 4.27$; Japanese listeners, $ts(26) \geq 5.28$; and Swedish listeners, $ts(29) \geq 8.06$; all $ps < .0002$. Swedish and Japanese participants further rated in-group music and Western classical music as equally familiar, whereas Indian listeners rated in-group music as significantly more familiar than Western classical music ($t_{29} = 9.53$, $p < .0001$). Table 2 lists mean values of rated familiarity for all culture pairings.

[3] It is beyond the scope of this article to describe in detail the musical structure and performance practices of the included musical traditions, but this information is available from other sources (e.g., Burkholder, Grout, & Palisca, 2009; Jersild & Ramsten, 2001; Malm, 2000; Ruckert, 2004).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
*Brief Description of the Acoustic and Musical Features*

| Feature type | Description |
|---|---|
| Dynamics | Applicable to both music and speech |
| RMS *M*, *SD* | The amplitude of the sound, calculated as the root mean square energy of the signal |
| Low energy | Percentage of frames within the segment having less-than-average energy |
| Rhythm | Applicable to both music and speech |
| Attack time *M* | The time duration between the initial note onset time and its peak time, i.e., a measure of how percussive or soft the sound attack is |
| Tempo *M* | Musical tempo (beats per minute) estimated by applying autocorrelation techniques to the amplitude envelope of the waveform |
| Pulse clarity[a] | How clear and stable the beat in music is; obtained by applying autocorrelation function to the amplitude envelope |
| Event density *M* | Number of estimated note onsets per second |
| Timbre | Applicable to both music and speech |
| Spectral centroid *M* | The geometric centre of the spectral frequency; a high value indicates prevalence of high frequencies, which makes the sound more sharp and less soft |
| Spectral entropy *M* | A measure of the complexity of the spectrum |
| Roughness | Estimation of the sensory dissonance |
| MFCC 2–7 | Mel frequency cepstral coefficients, i.e., descriptors of timbral structure across the frequency scale |
| Spectral flux *M* | Variation between the consecutive spectral frames |
| Register | Applicable to both music and speech |
| Salient pitch *M*, *SD* | The prevalent pitch in Hz established by chromagram-based methods |
| Tonality | Applicable to music |
| Key clarity *M* | Key clarity refers to the maximum correlation of so-called key profiles, which represent the stability of the pitch-classes in a given key |
| Mode *M* | Degree to which the segment is in a major key rather than a minor key, based on the key profiles |
| HCDF *M* | Harmonic Change Detection Function, which describes the amount of change of the tonal centroid |
| Structure | Applicable to music |
| Spectral novelty | "Novelty" refers to the degree of temporal repetition of any particular feature, such as spectrum across time based on detection of edges within the diagonal of the self-similarity matrix |
| Rhythm novelty | Degree of temporal repetition of rhythm, calculated using an autocorrelation vector to represent local rhythm structures |
| Tonal novelty | Degree of tonal repetition, calculated using a chromagram-based vector of pitch-classes to represent tonal structures |
| Register novelty | Degree of register repetition, calculated using an unwrapped chromagram vector to represent the local register |

*Note.* $M$ = mean, $SD$ = standard deviation.
[a] Not previously indicated as a shared cue with vocal expression.

## Affect Recognition Accuracy Within and Across Cultures

Recognition rates (proportion of judgment accuracy) were analyzed with a 3 (listener culture: Indian, Japanese, and Swedish) × 4 (musical culture: Hindustani classical, Japanese traditional, Swedish folk, and Western classical music) × 11 (expression) mixed measures ANOVA. There was no significant effect of listener culture, $F(2, 84) = 1.99$, $p = .14$, $\eta_p^2 = .05$, indicating that the listener groups overall performed with similar level of accuracy. However, accuracy varied across musical cultures, $F(3, 252) = 55.64$, $p < .0001$, $\eta_p^2 = .40$, such that Western classical music ($M = 0.31$) received significantly higher overall accuracy than Swedish ($M = 0.24$) and Indian music ($M = 0.22$), both of which, in turn, were better recognized than Japanese music ($M = 0.18$); $ts(86) \geq 3.38$, $ps < .0011$. Recognition rates also varied across expressions, $F(10, 840) = 45.06$, $p < .0001$, $\eta_p^2 = .34$, to the effect that anger ($M = 0.37$), fear ($M = 0.34$), happiness ($M = 0.29$), humor ($M = 0.33$), and sadness ($M = 0.32$) were better recognized, $ts(86) \geq 4.87$, $ps < .0001$, than affection ($M = 0.16$), longing ($M = 0.18$), neutral ($M = 0.18$), peacefulness ($M = 0.19$), and solemnity ($M = 0.14$); and all expressions, with the exception of solemnity, were better recognized than spirituality ($M = 0.10$), $ts(86) \geq 3.96$, $ps < .0002$.

All main effects were also qualified by significant interactions, and Table 2 shows the recognition rates as a function of listener culture, musical culture, and intended expression. The interaction between listener culture and intended expression, $F(20, 840) = 2.62$, $p = .00045$, $\eta_p^2 = .06$, indicated that hit rates for specific expressions varied between listener groups. The main trends were that Swedish listeners showed higher recognition rates for fear and longing than the other groups, whereas Indian listeners showed higher rates for sadness but lower rates for solemnity. The interaction between musical culture and expression, $F(30, 2520) = 17.57$, $p < .0001$, $\eta_p^2 = .17$, in turn indicated that hit rates for specific expressions also varied between musician groups. As shown in Table 2, Western classical music received the highest hit rates for most expressions, but happiness and humor were best recognized from Swedish music. Indian music was particularly effective for conveying anger and sadness, whereas Japanese music received its highest hit rates for fear.

The interaction between listener culture and musical culture, $F(6, 252) = 9.85$, $p < .0001$, $\eta_p^2 = .19$, is the main effect of interest for analyses of cultural differences in affect recognition. Indian and Swedish listeners performed significantly better overall when judging in-group compared with out-group music, $ts(29) \geq 3.88$, $ps < .0006$ (see Table 2). However, Japanese listeners did not perform better for in-group compared with out-group music. We observed no group differences in accuracy between in-group music and the control condition (Western classical music) for Indian and Swedish listeners, but Japanese listeners performed significantly better for Western classical music than for Japanese music ($t_{26} = 5.79$, $p < .0001$). As hypothesized, we found evidence for an overall in-group advantage (calculated as the mean difference between in-group and out-group accuracy across conditions), with higher accuracy in within-cultural ($M = 0.25$) versus cross-cultural ($M = 0.20$) conditions. A separate $t$ test (dependent samples) comparing each listener's overall in-group accuracy to

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 2
*Mean Recognition Accuracy Rates and Familiarity Ratings As a Function of Intended Expression, Musical Culture, and Listener Culture*

| | Recognition accuracy (%) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Western classical music | | | Hindustani classical music | | | Japanese traditional music | | | Swedish folk music | | | |
| Intended expression | Ind | Jap | Swe | Ind | Jap | Swe | Ind | Jap | Swe | Ind | Jap | Swe | In-group advantage |
| Affection | 23$^{\dagger}$ | 28$^{*}$ | 24$^{\dagger}$ | 19 | 15 | 8 | 2 | 9 | 8 | 23$^{\dagger}$ | 20 | 19 | 3 |
| Anger | 51$^{*}$ | 46$^{*}$ | 57$^{*}$ | 62$^{*}$ | 47$^{*}$ | 46$^{*}$ | 26$^{*}$ | 22$^{\dagger}$ | 21$^{\dagger}$ | 11 | 25$^{\dagger}$ | 28$^{*}$ | 8 |
| Fear | 39$^{*}$ | 40$^{*}$ | 58$^{*}$ | 40$^{*}$ | 30$^{*}$ | 10 | 32$^{*}$ | 32$^{*}$ | 57$^{*}$ | 18 | 25$^{\dagger}$ | 30$^{*}$ | 6 |
| Happiness | 41$^{*}$ | 36$^{*}$ | 32$^{*}$ | 12 | 15 | 10 | 7 | 14 | 11 | 58$^{*}$ | 54$^{*}$ | 53$^{*}$ | 1 |
| Humor | 42$^{*}$ | 40$^{*}$ | 57$^{*}$ | 22$^{\dagger}$ | 30$^{*}$ | 24$^{\dagger}$ | 14 | 16 | 13 | 39$^{*}$ | 47$^{*}$ | 57$^{*}$ | 4 |
| Longing | 24$^{\dagger}$ | 7 | 23$^{\dagger}$ | 18 | 16 | 24$^{\dagger}$ | 11 | 15 | 24$^{\dagger}$ | 10 | 10 | 32$^{*}$ | 6 |
| Neutral | 29$^{*}$ | 31$^{*}$ | 28$^{*}$ | 22$^{\dagger}$ | 6 | 9 | 7 | 15 | 11 | 20$^{\dagger}$ | 22$^{\dagger}$ | 14 | 5 |
| Peacefulness | 24$^{\dagger}$ | 17 | 22$^{\dagger}$ | 18 | 14 | 28$^{*}$ | 11 | 20 | 21$^{\dagger}$ | 16 | 21$^{\dagger}$ | 19 | 0 |
| Sadness | 42$^{*}$ | 52$^{*}$ | 31$^{*}$ | 62$^{*}$ | 15 | 23$^{\dagger}$ | 39$^{*}$ | 22$^{\dagger}$ | 26$^{*}$ | 13 | 28$^{*}$ | 29$^{*}$ | 14 |
| Solemnity | 12 | 17 | 29$^{*}$ | 10 | 14 | 3 | 10 | 23$^{\dagger}$ | 13 | 7 | 10 | 26$^{*}$ | 10 |
| Spirituality | 6 | 6 | 9 | 16 | 12 | 13 | 9 | 12 | 13 | 8 | 7 | 8 | 1 |
| Overall accuracy | 30 | 29 | 34 | 27 | 19 | 18 | 15 | 18 | 20 | 20 | 25 | 29 | 5 |
| Familiarity | 2.63 | 4.48 | 4.73 | 5.40 | 2.00 | 2.20 | 1.43 | 4.30 | 2.07 | 1.27 | 2.33 | 4.63 | 2.89 |

*Note.* $N = 90$ observations/cell for Indian and Swedish listeners, and $N = 81$ observations/cell for Japanese listeners. The in-group advantage was calculated as the difference between in-group accuracy (i.e., when expressors and perceivers come from the same culture) and out-group accuracy (i.e., when expressors and perceivers come from different cultures). Note that recognition rates for the control condition (Western classical music) were not included in the calculation of the in-group advantage. Asterisks and daggers indicate the chance level of binomial tests conducted on the proportion of participants who chose each expression for a given target expression.
$^{*}$ Chance level = 20%, $p < .05$
$^{\dagger}$ Chance level = 9.09%, $p < .05$ (Bonferroni corrected).

their out-group accuracy indicated that this difference was statistically significant, $t_{86} = 4.60$, $p < .0001$, $d = .69$.

Finally, we observed a significant three-way interaction between listener culture, musical culture, and expression, $F(60, 2520) = 3.11$, $p < .0001$, $\eta_p^2 = .07$, indicating that the size of the in-group advantage varied as a function of expression and culture conditions. As shown in Table 2, the in-group advantage was largest for sadness, solemnity, anger, longing, and fear, whereas practically no advantage was observed for happiness, peacefulness, and spirituality. Pairwise $t$ tests comparing each listener's in-group versus out-group accuracy were conducted separately for each expression, and these tests showed that the in-group advantage was statistically significant for sadness ($t_{86} = 4.03$, $p = .0001$, $d = .56$), solemnity ($t_{86} = 3.34$, $p = .0012$, $d = .51$), and anger ($t_{86} = 2.01$, $p = .048$, $d = .33$), and marginally significant for longing ($t_{86} = 1.98$, $p = .051$, $d = .26$). Looking at individual Listener Culture × Musical Culture × Expression combinations, the largest differences between in-group and out-group accuracy occurred for Indian listeners and angry, fearful, and sad music; Japanese listeners and solemn music; and Swedish listeners and happy, humorous, longing, and solemn music (see Table 2).

Binomial tests were conducted to test whether the proportion of participants who chose the correct response alternative for each target expression was higher than the proportion expected by chance guessing. Using a traditional chance level based on the number of response options (9.09%; 1 out of 11) these tests showed that most expressions were recognized with accuracy above chance from both familiar and unfamiliar music (see Table 2). The exceptions were solemnity, which was not recognized in cross-cultural conditions, and spirituality, which was not recognized above chance at all. There is much discussion regarding the proper definition of chance, however, and some researchers have argued that basing the chance level on the number of response options is too lenient a criterion (e.g., Russell, 1994). We therefore conducted additional binomial tests, wherein we conservatively set the chance level at 20% (1 out of 5), given the four categories of basic emotions: anger, fear, happiness, and sadness, as well as the "nonbasic" affect category (which included the rest of our intended expressions). This approach rests on the assumption that recognition rates could be inflated by guessing strategies reflecting various similarities between expressions (see Frank & Stennett, 2001; Hertenstein, Keltner, App, Bulleit, & Jaskolka, 2006; Simon-Thomas, Keltner, Sauter, Sinicropi-Yao, & Abramson, 2009). Results from the more stringent binomial tests (chance level = 20%) revealed that all expressions, again excluding spirituality, were accurately judged in at least some conditions, which suggests that a wide variety of emotions and related states can be communicated through music. However, when looking specifically at cross-cultural performance, the list of accurately recognized expressions narrowed down to anger, fear, happiness, humor, peacefulness, and sadness (see Table 2).

On the one hand, authors have pointed out that response biases (i.e., biases in the relative utilization of different response alternatives) may affect accuracy indices to the effect that an overuse of a specific response alternative may lead to an overestimation of the accuracy for this category (e.g., Wagner, 1993). On the other hand, some confusions between response categories are meaningful, and therefore the forced-choice procedure may also lead to an underestimation of the signal value of certain stimulus categories. The most complete way to provide all relevant information about biases and confusion patterns is therefore to reproduce the detailed confusion matrices for all pairings of musical and listener cultures

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

(e.g., Bänziger, Mortillaro, & Scherer, 2012), and these are presented in Table 3. Not surprisingly, confusions were most prevalent between the conceptually similar expression labels happiness and humor. In addition, low arousal expressions (affection, longing, peacefulness, and solemnity) were frequently confused with each other and, especially, with sadness, although sadness expressions were seldom confused with other expressions. In general, listeners tended to make similar errors across cultures, as indicated by positive correlations between error profiles. The error profiles consist of the confusion matrices from Table 3 with the diagonal entries removed (see Elfenbein, Mandal, Ambady, Harizuka, & Kumar, 2002; Scherer et al., 2001), and the correlations of the error profiles across combinations of musical and listener cultures are shown in Table S1 of the online supplemental material. The average correlation for error profiles calculated within each musical culture (mean Fisher $z = 0.89$) was higher than the average correlation for error profiles calculated across musical cultures (mean Fisher $z = 0.45$), and this difference suggests that cultural differences in how musical expressions are performed may be partially responsible for variations in error patterns.

## Acoustic Correlates of Musical Affect Expressions: A Lens Model Analysis

Next we investigated how acoustic and musical cues were associated with the musicians' intended expressions and the listeners' judgments. Drawing on the logic of Brunswik's (1956) lens model, we first calculated the point-biserial correlations between the acoustic cues and the performers' intended expression (dichotomously dummy coded as 1 or 0), that is, *cue-validity correlations*. The cue-validity correlations provide a measure of the degree to which various cues were related to the different expressions; a high cue-validity correlation suggests that the musicians used a cue in a consistent fashion to convey a specific expression. Second, we calculated the correlations between the acoustic cues and the listeners' mean affect judgments—that is, *cue-utilization correlations*. A high cue-utilization correlation suggests that decoders, on average, used a cue in a consistent way to make inferences about the conveyed expression. Table 4 presents cue-validity and cue-utilization correlations for selected acoustic cues as a function of expression across musical cultures and listener groups.

Inspection of Table 4 revealed that a wide variety of cues were associated with both the performers expressive intentions (cue validity) and the listeners' judgments (cue utilization) across cultures (see Table 1 for an explanation of cue abbreviations). Focusing on cues that are both valid and utilized, we observed that anger was associated with high sound level (positive correlations with RMS *M* and negative correlations with Low energy), whereas affection and peacefulness were associated with low sound level (negative correlation with RMS *M*). Regarding cues related to rhythm, affection, neutral, and peacefulness were associated with a slow (soft) attack (positive correlations with Attack time *M*), whereas anger and fear were instead associated with a fast (percussive) attack. Humor was associated with a fast tempo (positive correlation with Tempo *M*). Anger, happiness, and humor were further associated with a clear pulse (positive correlations with Pulse clarity), and longing, peacefulness, and spirituality with a more ambiguous pulse (negative correlations with Pulse clarity). Peacefulness and sadness were also associated with low event density (negative correlations with mean number of detected onsets per second), whereas anger and fear were associated with high event density. The associations with timbre cues revealed that anger—and, to some extent, happiness and humor—were associated with a rough timbre that contains much high-frequency energy (positive correlations with Spectral entropy, Roughness, Spectral flux, and negative correlations with MFCC2), whereas affection, longing, neutral, peacefulness, and sadness instead showed an opposite pattern of correlations and thus were associated with a smoother, more organized, and less fluctuating spectrum. Regarding cues related to register, only solemnity showed significant correlations and was associated with low pitch (negative correlation with Salient pitch *M*). The associations with tonal cues revealed that happiness and humor were associated with major mode (positive correlations with Mode) and sadness with minor mode. In addition, affection and neutral were associated with large changes in the harmonic content of the pieces (positive correlations with HCDF), whereas anger was associated with small harmonic changes. Finally, regarding structural cues, affection, longing, neutral, and peacefulness were associated with high novelty (positive correlations with novelty cues, i.e., structural features picking up the amount of temporal originality), whereas anger, happiness, and humor were instead associated with low novelty. In other words, anger, happiness, and humor tended to have more constant and stable feature characteristics across time compared with affection, longing, neutral, and peacefulness, which instead exhibited frequent and more substantial time-varying changes.

A central tenet of the dialect theory of emotion is that any in-group affect-recognition advantage should result from culture-specific cue-validity patterns, unfamiliarity of which leads to worse cue utilization across cultures than within cultures. We therefore wished to investigate if matching between valid cues and cue utilization was better for in-group music versus unfamiliar music. First we calculated the cue-validity and cue-utilization correlations for each expression and all pairings of musical and listener cultures (the resulting correlation matrix is too large to be reproduced in the main text of this article, but is shown in Table S2 of the online supplemental material). Inspection of Table S2 revealed that the culture-specific cue correlation patterns were similar to the "universal" correlation patterns shown in Table 4. However, as expected, there were more mismatches between encoding and decoding in the culture-specific data. Focusing on the Listener Culture × Musical Culture × Expression combinations with the largest differences between in-group and out-group accuracy, we note that tonal and rhythmic novelty were valid and utilized cues for expressing anger and fear, respectively, in Hindustani classical music, but only for in-group conditions. Likewise, event density was a valid and utilized cue for expressing sadness in Japanese music for in-group, but not out-group, conditions.

In order to quantify the degree of match between musicians' and listeners' uses of cues, we calculated the correlations between the cue-validity and cue-utilization patterns, and these correlations are shown in Table 5 as a function of intended expression, musical culture, and listener culture (i.e., *validity-utilization correlations*). These correlations were calculated across all musical/acoustic cues, and a high validity-utilization correlation suggests a good match between the musicians' cue-validity patterns and the listeners' cue-utilization patterns. Across all expressions, the validity-

Table 3
*Confusion Matrices For All Pairings of Judge and Musical Cultures*

| Judgment | Culture | Intended expression: Affection | Anger | Fear | Happiness | Humor | Longing | Neutral | Peacefulness | Sadness | Solemnity | Spirituality |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Western classical music | | | | | | | | | | | | |
| Affection | Indian | **23** | 3 | 0 | 16 | 6 | 12 | 7 | 4 | 9 | 18 | 14 |
| | Japanese | **28** | 0 | 1 | 10 | 2 | 17 | 1 | 17 | 6 | 12 | 2 |
| | Swedish | **24** | 0 | 1 | 2 | 0 | 17 | 4 | 16 | 11 | 11 | 6 |
| Anger | Indian | 0 | **51** | 8 | 4 | 0 | 0 | 1 | 0 | 0 | 6 | 2 |
| | Japanese | 1 | **46** | 20 | 5 | 0 | 5 | 1 | 1 | 0 | 5 | 0 |
| | Swedish | 0 | **57** | 20 | 4 | 0 | 3 | 3 | 0 | 0 | 3 | 0 |
| Fear | Indian | 1 | 8 | **39** | 2 | 1 | 9 | 16 | 12 | 1 | 7 | 3 |
| | Japanese | 2 | 25 | **40** | 2 | 0 | 19 | 16 | 6 | 11 | 1 | 4 |
| | Swedish | 1 | 26 | **58** | 12 | 0 | 17 | 22 | 3 | 3 | 3 | 0 |
| Happiness | Indian | 1 | 11 | 3 | **41** | 49* | 0 | 0 | 0 | 0 | 3 | 3 |
| | Japanese | 0 | 0 | 0 | **36** | 48* | 1 | 0 | 0 | 0 | 5 | 2 |
| | Swedish | 3 | 1 | 0 | **32** | 37* | 2 | 1 | 0 | 0 | 4 | 3 |
| Humor | Indian | 0 | 19 | 6 | 24 | **42** | 3 | 7 | 0 | 0 | 1 | 0 |
| | Japanese | 0 | 17 | 2 | 15 | **40** | 2 | 1 | 0 | 0 | 1 | 1 |
| | Swedish | 0 | 10 | 1 | 30* | **57** | 0 | 0 | 0 | 0 | 2 | 0 |
| Longing | Indian | 9 | 2 | 9 | 2 | 1 | **24** | 6 | 6 | 10 | 10 | 12 |
| | Japanese | 14 | 2 | 6 | 5 | 6 | **7** | 9 | 7 | 9 | 17 | 16 |
| | Swedish | 16 | 3 | 8 | 2 | 3 | **23** | 10 | 8 | 21 | 13 | 17 |
| Neutral | Indian | 10 | 3 | 26 | 6 | 0 | 10 | **29** | 19 | 8 | 13 | 13 |
| | Japanese | 10 | 5 | 6 | 9 | 1 | 9 | **31** | 5 | 2 | 6 | 12 |
| | Swedish | 6 | 0 | 6 | 8 | 3 | 8 | **28** | 16 | 2 | 8 | 10 |
| Peacefulness | Indian | 18 | 0 | 1 | 0 | 0 | 6 | 9 | **24** | 8 | 14 | 16 |
| | Japanese | 20 | 0 | 1 | 10 | 2 | 6 | 12 | **17** | 7 | 9 | 7 |
| | Swedish | 19 | 0 | 0 | 0 | 0 | 7 | 13 | **22** | 13 | 9 | 13 |
| Sadness | Indian | 20 | 0 | 4 | 0 | 0 | 26 | 16 | 20 | **42** | 8 | 19 |
| | Japanese | 12 | 0 | 6 | 1 | 0 | 22 | 10 | 26 | **52** | 19 | 36* |
| | Swedish | 9 | 2 | 7 | 1 | 0 | 14 | 14 | 24 | **31** | 11 | 26 |
| Solemnity | Indian | 11 | 2 | 3 | 3 | 1 | 6 | 4 | 10 | 17 | **12** | 11 |
| | Japanese | 6 | 5 | 9 | 6 | 0 | 5 | 2 | 7 | 4 | **17** | 12 |
| | Swedish | 16 | 1 | 0 | 7 | 0 | 2 | 1 | 4 | 8 | **29** | 17 |
| Spirituality | Indian | 7 | 0 | 1 | 1 | 0 | 4 | 7 | 4 | 6 | 8 | **6** |
| | Japanese | 6 | 0 | 9 | 1 | 0 | 6 | 16 | 12 | 9 | 7 | **6** |
| | Swedish | 7 | 0 | 0 | 1 | 0 | 7 | 2 | 7 | 10 | 6 | **9** |
| Hindustani classical music | | | | | | | | | | | | |
| Affection | Indian | **19** | 4 | 1 | 34* | 23 | 8 | 16 | 7 | 4 | 7 | 27 |
| | Japanese | **15** | 1 | 2 | 19 | 11 | 7 | 30* | 7 | 5 | 2 | 12 |
| | Swedish | **8** | 0 | 4 | 20 | 11 | 6 | 18 | 11 | 3 | 4 | 13 |
| Anger | Indian | 0 | **62** | 3 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| | Japanese | 2 | **47** | 11 | 12 | 1 | 4 | 0 | 2 | 5 | 6 | 5 |
| | Swedish | 0 | **46** | 6 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 3 |
| Fear | Indian | 1 | 11 | **40** | 3 | 0 | 6 | 0 | 0 | 4 | 3 | 1 |
| | Japanese | 7 | 22 | **30** | 7 | 0 | 11 | 4 | 7 | 15 | 22 | 11 |
| | Swedish | 10 | 38* | **10** | 10 | 2 | 8 | 2 | 4 | 6 | 18 | 12 |
| Happiness | Indian | 0 | 12 | 0 | **12** | 39* | 0 | 6 | 0 | 0 | 7 | 9 |
| | Japanese | 1 | 2 | 0 | **15** | 35* | 0 | 7 | 0 | 1 | 2 | 4 |
| | Swedish | 4 | 3 | 0 | **10** | 34* | 0 | 10 | 0 | 0 | 1 | 10 |
| Humor | Indian | 0 | 0 | 0 | 0 | **22** | 0 | 3 | 0 | 0 | 0 | 0 |
| | Japanese | 1 | 11 | 0 | 5 | **30** | 1 | 5 | 0 | 1 | 2 | 2 |
| | Swedish | 1 | 9 | 0 | 7 | **24** | 0 | 0 | 0 | 0 | 1 | 2 |
| Longing | Indian | 8 | 3 | 9 | 11 | 6 | **18** | 11 | 12 | 19 | 13 | 16 |
| | Japanese | 6 | 7 | 5 | 6 | 7 | **16** | 10 | 4 | 12 | 9 | 16 |
| | Swedish | 20 | 1 | 16 | 11 | 6 | **24** | 11 | 16 | 20 | 22 | 12 |
| Neutral | Indian | 6 | 2 | 4 | 2 | 1 | 3 | **22** | 1 | 1 | 2 | 2 |
| | Japanese | 15 | 5 | 6 | 6 | 2 | 12 | **6** | 11 | 10 | 6 | 7 |
| | Swedish | 2 | 1 | 1 | 9 | 10 | 4 | **9** | 8 | 6 | 8 | 9 |
| Peacefulness | Indian | 20 | 0 | 0 | 9 | 7 | 9 | 9 | **18** | 3 | 16 | 8 |
| | Japanese | 17 | 0 | 1 | 16 | 10 | 5 | 17 | **14** | 9 | 4 | 14 |
| | Swedish | 17 | 0 | 3 | 8 | 3 | 10 | 20 | **28** | 19 | 6 | 3 |
| Sadness | Indian | 28* | 2 | 33* | 7 | 0 | 42* | 9 | 37* | **62** | 31* | 18 |
| | Japanese | 14 | 0 | 27 | 1 | 0 | 16 | 1 | 25 | **15** | 14 | 9 |
| | Swedish | 18 | 1 | 50* | 8 | 1 | 18 | 10 | 20 | **23** | 28* | 12 |

(*table continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3 (*continued*)

| Judgment | Culture | Intended expression | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Affection | Anger | Fear | Happiness | Humor | Longing | Neutral | Peacefulness | Sadness | Solemnity | Spirituality |
| Solemnity | Indian | 8 | 1 | 6 | 4 | 0 | 7 | 11 | 8 | 2 | **10** | 4 |
| | Japanese | 11 | 2 | 10 | 5 | 0 | 6 | 12 | 15 | 10 | **14** | 7 |
| | Swedish | 4 | 1 | 6 | 10 | 3 | 4 | 14 | 2 | 6 | **3** | 9 |
| Spirituality | Indian | 11 | 1 | 3 | 17 | 2 | 7 | 13 | 17 | 3 | 10 | **16** |
| | Japanese | 10 | 1 | 7 | 7 | 4 | 21 | 7 | 15 | 17 | 19 | **12** |
| | Swedish | 16 | 0 | 4 | 7 | 4 | 26 | 4 | 11 | 17 | 9 | **13** |
| Japanese traditional music | | | | | | | | | | | | |
| Affection | Indian | **2** | 7 | 1 | 8 | 11 | 4 | 1 | 6 | 4 | 7 | 2 |
| | Japanese | **9** | 0 | 1 | 6 | 10 | 15 | 6 | 7 | 7 | 5 | 6 |
| | Swedish | **8** | 3 | 2 | 3 | 9 | 7 | 7 | 10 | 2 | 8 | 4 |
| Anger | Indian | 2 | **26** | 18 | 6 | 4 | 2 | 0 | 1 | 4 | 3 | 2 |
| | Japanese | 0 | **22** | 6 | 1 | 23 | 0 | 1 | 0 | 1 | 5 | 6 |
| | Swedish | 1 | **21** | 10 | 12 | 19 | 0 | 2 | 4 | 3 | 1 | 1 |
| Fear | Indian | 2 | 14 | **32** | 16 | 12 | 13 | 13 | 7 | 20 | 12 | 14 |
| | Japanese | 10 | 5 | **32** | 1 | 5 | 5 | 2 | 1 | 17 | 5 | 11 |
| | Swedish | 1 | 11 | **57** | 10 | 10 | 3 | 3 | 1 | 16 | 8 | 3 |
| Happiness | Indian | 0 | 10 | 11 | **7** | 8 | 2 | 0 | 4 | 2 | 1 | 0 |
| | Japanese | 1 | 7 | 9 | **14** | 25 | 6 | 1 | 11 | 2 | 4 | 0 |
| | Swedish | 0 | 8 | 4 | **11** | 14 | 2 | 2 | 3 | 3 | 1 | 0 |
| Humor | Indian | 0 | 2 | 3 | 2 | **14** | 0 | 0 | 2 | 4 | 0 | 0 |
| | Japanese | 1 | 11 | 10 | 26 | **16** | 2 | 1 | 7 | 9 | 1 | 0 |
| | Swedish | 0 | 13 | 8 | 10 | **13** | 0 | 1 | 3 | 6 | 1 | 1 |
| Longing | Indian | 12 | 13 | 8 | 14 | 10 | **11** | 22 | 13 | 9 | 7 | 10 |
| | Japanese | 4 | 21 | 10 | 10 | 9 | **15** | 14 | 5 | 11 | 9 | 6 |
| | Swedish | 18 | 7 | 3 | 9 | 8 | **24** | 14 | 17 | 24 | 12 | 14 |
| Neutral | Indian | 12 | 10 | 6 | 17 | 12 | 7 | **7** | 21 | 6 | 12 | 13 |
| | Japanese | 15 | 5 | 5 | 12 | 2 | 5 | **15** | 10 | 4 | 11 | 6 |
| | Swedish | 16 | 8 | 1 | 7 | 9 | 7 | **11** | 4 | 1 | 12 | 13 |
| Peacefulness | Indian | 4 | 0 | 0 | 2 | 3 | 4 | 3 | **11** | 3 | 12 | 9 |
| | Japanese | 14 | 4 | 2 | 7 | 5 | 12 | 15 | **20** | 9 | 9 | 2 |
| | Swedish | 13 | 1 | 0 | 1 | 1 | 4 | 7 | **21** | 11 | 12 | 16 |
| Sadness | Indian | 48* | 6 | 13 | 12 | 9 | 38* | 26 | 12 | **39** | 17 | 33* |
| | Japanese | 23 | 0 | 17 | 5 | 2 | 19 | 14 | 14 | **22** | 16 | 31* |
| | Swedish | 31* | 8 | 6 | 4 | 3 | 36* | 32* | 13 | **26** | 21 | 27 |
| Solemnity | Indian | 9 | 8 | 4 | 7 | 8 | 11 | 11 | 8 | 4 | **10** | 7 |
| | Japanese | 10 | 16 | 5 | 9 | 1 | 12 | 17 | 14 | 11 | **23** | 19 |
| | Swedish | 6 | 18 | 8 | 18 | 8 | 9 | 8 | 16 | 3 | **13** | 7 |
| Spirituality | Indian | 8 | 4 | 3 | 10 | 8 | 7 | 17 | 14 | 3 | 19 | **9** |
| | Japanese | 14 | 9 | 2 | 9 | 1 | 9 | 14 | 11 | 6 | 12 | **12** |
| | Swedish | 7 | 2 | 1 | 14 | 6 | 8 | 12 | 7 | 4 | 10 | **13** |
| Swedish folk music | | | | | | | | | | | | |
| Affection | Indian | **23** | 10 | 12 | 3 | 8 | 22 | 8 | 17 | 17 | 11 | 11 |
| | Japanese | **20** | 1 | 1 | 6 | 6 | 5 | 28* | 15 | 1 | 11 | 9 |
| | Swedish | **19** | 1 | 3 | 0 | 1 | 7 | 12 | 13 | 7 | 10 | 13 |
| Anger | Indian | 1 | **11** | 9 | 1 | 1 | 2 | 0 | 1 | 4 | 3 | 1 |
| | Japanese | 0 | **25** | 11 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| | Swedish | 0 | **28** | 4 | 1 | 1 | 1 | 0 | 0 | 4 | 2 | 1 |
| Fear | Indian | 2 | 6 | **18** | 1 | 2 | 0 | 0 | 0 | 10 | 3 | 6 |
| | Japanese | 2 | 10 | **25** | 0 | 0 | 7 | 0 | 1 | 16 | 6 | 9 |
| | Swedish | 1 | 14 | **30** | 1 | 1 | 1 | 2 | 2 | 2 | 0 | 1 |
| Happiness | Indian | 22 | 20 | 21 | **58** | 37* | 6 | 6 | 8 | 3 | 31* | 2 |
| | Japanese | 20 | 9 | 7 | **54** | 28* | 4 | 6 | 10 | 1 | 21 | 0 |
| | Swedish | 23 | 17 | 21 | **53** | 36* | 2 | 12 | 3 | 0 | 14 | 1 |
| Humor | Indian | 6 | 16 | 19 | 23 | **39** | 2 | 2 | 0 | 2 | 7 | 4 |
| | Japanese | 4 | 9 | 16 | 32* | **47** | 1 | 1 | 2 | 0 | 9 | 0 |
| | Swedish | 7 | 13 | 17 | 40* | **57** | 3 | 2 | 0 | 1 | 14 | 0 |
| Longing | Indian | 9 | 11 | 9 | 3 | 4 | **10** | 6 | 8 | 12 | 7 | 8 |
| | Japanese | 5 | 16 | 10 | 2 | 5 | **10** | 6 | 12 | 7 | 6 | 16 |
| | Swedish | 10 | 9 | 10 | 1 | 1 | **32** | 11 | 16 | 26 | 10 | 24 |
| Neutral | Indian | 8 | 16 | 6 | 2 | 3 | 16 | **20** | 8 | 13 | 12 | 12 |
| | Japanese | 7 | 5 | 7 | 0 | 1 | 10 | **22** | 12 | 7 | 4 | 5 |
| | Swedish | 4 | 2 | 8 | 1 | 1 | 10 | **14** | 8 | 4 | 9 | 4 |
| Peacefulness | Indian | 9 | 1 | 0 | 1 | 0 | 7 | 28* | **16** | 7 | 9 | 12 |
| | Japanese | 15 | 2 | 0 | 2 | 9 | 9 | 31* | **21** | 6 | 19 | 7 |
| | Swedish | 6 | 1 | 0 | 0 | 1 | 1 | 18 | **19** | 6 | 6 | 7 |

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3 (*continued*)

| Judgment | Culture | Intended expression | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Affection | Anger | Fear | Happiness | Humor | Longing | Neutral | Peacefulness | Sadness | Solemnity | Spirituality |
| Sadness | Indian | 4 | 2 | 3 | 0 | 0 | 19 | 4 | 18 | **13** | 7 | 26 |
| | Japanese | 12 | 4 | 6 | 0 | 0 | 33* | 2 | 21 | **28** | 10 | 42* |
| | Swedish | 9 | 1 | 2 | 1 | 0 | 21 | 3 | 22 | **29** | 6 | 33* |
| Solemnity | Indian | 11 | 4 | 3 | 7 | 3 | 8 | 14 | 16 | 13 | **7** | 10 |
| | Japanese | 7 | 15 | 10 | 2 | 1 | 9 | 1 | 5 | 11 | **10** | 5 |
| | Swedish | 18 | 12 | 3 | 1 | 1 | 16 | 19 | 16 | 16 | **26** | 7 |
| Spirituality | Indian | 4 | 3 | 0 | 0 | 2 | 9 | 12 | 10 | 4 | 3 | **8** |
| | Japanese | 7 | 5 | 6 | 0 | 2 | 10 | 1 | 0 | 19 | 5 | **7** |
| | Swedish | 3 | 1 | 1 | 0 | 0 | 6 | 6 | 1 | 6 | 3 | **8** |

*Note.* $N = 90$ observations/cell for Indian and Swedish listeners, and $N = 81$ observations/cell for Japanese listeners. Bold typeface indicates the values in the diagonal cells, which represent the hit rate (percentage accuracy) for which the expression portrayed is the same as the expression judged. Asterisks indicate misattributions with frequency higher than chance guessing (binomial tests, chance level = 20%, $p < .05$).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

utilization correlations were significantly higher in in-group conditions (mean Fisher $z = 1.00$) than in out-group conditions (mean Fisher $z = 0.68$, $t_{10} = 3.63$, $p = .0046$, $d = .631$), suggesting a better match between encoding and decoding when musicians and listeners share the same musical culture, as hypothesized.

Finally, we investigated the associations between (a) cultural familiarity with a specific musical genre (i.e., the last row in Table 2), (b) mean affect recognition accuracy (i.e., the second last row in Table 2), and (c) the degree of match between musicians' and listeners' uses of cues (mean validity-utilization correlations; i.e., the last row in Table 5) across expressions ($N = 12$ for all analyses). Familiarity was positively correlated with both recognition accuracy ($r = .66$, $p = .020$) and the pattern of validity-utilization correlations ($r = .67$, $p = .018$), suggesting that familiarity facilitates recognition accuracy and knowledge about culture-specific performance strategies. As expected, recognition accuracy was also positively correlated with the validity-utilization correlations ($r = .91$, $p < .0001$), suggesting that a good match between musician and listener cue patterns enhances affect recognition accuracy.

## Discussion

To summarize the main findings, the results first showed that significantly more listeners than would be expected by a stringent chance threshold accurately identified a wide range of emotions and related states from culturally familiar music. In addition to replication of previous findings for affection, anger, fear, happiness, neutral, sadness, and solemnity, our participants also performed above chance for humor and longing. However, with the exception of humor, basic emotions (anger, fear, happiness, and sadness) were better recognized than the other "nonbasic" affective states. We also found that listeners were able to decode the musicians' intended expressions from culturally unfamiliar music with accuracy above chance, which suggests that affect expression may be a partly universal aspect of musical meaning. However, only anger, fear, happiness, humor, peacefulness, and sadness were recognized with accuracy above chance in cross-cultural conditions. Importantly, listeners performed better when judging affect from culturally familiar versus unfamiliar music. This is the first demonstration of an in-group advantage in musical affect recognition using a balanced design—and, in turn, suggests a culture-specific component to expression of affect. We next proceeded to investigate the mechanism that may give rise to the observed in-group advantage, using a lens-model analysis to illustrate how musicians express emotions and related states by varying acoustic and musical cues, which are then utilized by listeners to judge the expressed affect. Here we observed how a wide variety of cues were correlated with both the musicians' expressive intentions and the listeners' affect judgments. The acoustic correlates for basic emotions were generally in accord with previous studies on music and speech (Gabrielsson & Lindström, 2010; Juslin & Laukka, 2003), but we also reported data for many novel cues and "non-basic" affective states. The cue correlation patterns were in general similar across listener and musical cultures, but—in line with the dialect theory of emotion (Elfenbein et al., 2007)—we also demonstrated that musicians' and listeners' uses of cues were better matched in within-cultural versus cross-cultural conditions.

The results overall supported our hypotheses, which were based on the observed similarities between affective expression in speech and music. Taken together, the dual findings of cross-cultural invariance and relativity suggest that affective expression in music may depend on a combination of universal and culture-specific factors. In the following paragraphs, discussion will focus on the questions about which emotions and related states, and which musical features, may be part of the universal characteristics of music.

### Sources of Cultural Invariance and Relativity in Musical Expression

Traditionally, findings of universality in facial or vocal expression are interpreted as evidence for the proposition that emotions and their expressions are grounded in evolved, biologically driven mechanisms (e.g., Ekman, 1992). We argue that the universal component of musical affect expression also may be grounded in evolved mechanisms, but assuming that musical expressions are partly conveyed using a code originating from vocal expressions, the biological component probably originates from vocal expression rather than from music itself. Indeed, recent studies on the neural mechanisms involved in emotion decoding from vocal (Leitman et al., 2010; Schirmer & Kotz, 2006) and musical (Escoffier, Zhong, Schirmer, & Qiu, in

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 4
*A Lens-Model Analysis of the Associations Between Selected Acoustic and Musical Cues and (A) the Musicians' Intended Expression, and (B) the Listeners' Perceived Expression, Across Musical Cultures and Listener Groups*

| Cue | Affection | Anger | Fear | Happiness | Humor | Longing | Neutral | Peacefulness | Sadness | Solemnity | Spirituality |
|---|---|---|---|---|---|---|---|---|---|---|---|
| RMS *M* | | | | | | | | | | | |
| Validity | **−0.19**$^{*}$ | **0.58**$^{***}$ | −0.04 | 0.15 | −0.02 | 0.07 | −0.12 | **−0.25**$^{**}$ | −0.16 | 0.13 | −0.16 |
| Utilization | **−0.20**$^{*}$ | **0.51**$^{***}$ | 0.06 | 0.17$^{*}$ | 0.22$^{*}$ | −0.05 | −0.08 | **−0.44**$^{***}$ | −0.37$^{***}$ | 0.01 | −0.17 |
| Low energy | | | | | | | | | | | |
| Validity | −0.02 | **−0.20**$^{*}$ | 0.02 | −0.08 | 0.17$^{*}$ | 0.03 | −0.17 | 0.10 | 0.09 | 0.07 | −0.02 |
| Utilization | 0.00 | **−0.19**$^{*}$ | −0.04 | 0.07 | 0.08 | −0.05 | −0.22$^{*}$ | 0.06 | 0.12 | −0.06 | 0.07 |
| Attack time *M* | | | | | | | | | | | |
| Validity | **0.22**$^{*}$ | **−0.38**$^{***}$ | **−0.31**$^{***}$ | −0.15 | −0.17 | 0.12 | **0.18**$^{*}$ | **0.18**$^{*}$ | 0.10 | 0.06 | 0.17 |
| Utilization | **0.26**$^{**}$ | **−0.45**$^{***}$ | **−0.39**$^{***}$ | −0.31$^{***}$ | −0.37$^{***}$ | 0.36$^{***}$ | **0.24**$^{**}$ | **0.47**$^{***}$ | 0.42$^{***}$ | 0.31$^{***}$ | 0.44$^{***}$ |
| Tempo *M* | | | | | | | | | | | |
| Validity | −0.08 | 0.03 | 0.10 | 0.08 | **0.19**$^{*}$ | 0.03 | −0.13 | 0.01 | −0.06 | −0.03 | −0.15 |
| Utilization | −0.23$^{**}$ | 0.12 | 0.03 | 0.25$^{**}$ | **0.23**$^{**}$ | −0.23$^{**}$ | −0.01 | −0.25$^{**}$ | −0.10 | −0.17 | −0.15 |
| Pulse clarity | | | | | | | | | | | |
| Validity | −0.11 | **0.47**$^{***}$ | 0.07 | **0.18**$^{*}$ | **0.21**$^{*}$ | **−0.23**$^{**}$ | −0.04 | **−0.19**$^{*}$ | −0.05 | −0.08 | **−0.24**$^{**}$ |
| Utilization | −0.27$^{**}$ | **0.46**$^{***}$ | 0.18$^{*}$ | **0.35**$^{***}$ | **0.47**$^{***}$ | **−0.40**$^{***}$ | −0.09 | **−0.38**$^{***}$ | −0.47$^{***}$ | −0.29$^{***}$ | **−0.43**$^{***}$ |
| Event density | | | | | | | | | | | |
| Validity | −0.05 | **0.32**$^{***}$ | **0.25**$^{**}$ | 0.13 | 0.15 | −0.05 | −0.08 | **−0.20**$^{*}$ | **−0.26**$^{**}$ | −0.13 | −0.08 |
| Utilization | −0.13 | **0.33**$^{***}$ | **0.20**$^{*}$ | 0.27$^{**}$ | 0.34$^{***}$ | −0.21$^{*}$ | −0.09 | **−0.39**$^{***}$ | **−0.40**$^{***}$ | −0.20$^{*}$ | −0.32$^{***}$ |
| Spectral entropy | | | | | | | | | | | |
| Validity | −0.04 | **0.44**$^{***}$ | 0.12 | 0.11 | **0.22**$^{*}$ | **−0.24**$^{**}$ | **−0.25**$^{**}$ | −0.06 | **−0.17**$^{*}$ | 0.02 | −0.15 |
| Utilization | −0.14 | **0.47**$^{***}$ | 0.13 | 0.33$^{***}$ | **0.35**$^{***}$ | **−0.45**$^{***}$ | **−0.22**$^{*}$ | −0.27$^{***}$ | **−0.48**$^{***}$ | −0.11 | −0.35$^{***}$ |
| Roughness | | | | | | | | | | | |
| Validity | −0.17 | **0.64**$^{***}$ | −0.04 | 0.02 | 0.07 | −0.07 | **−0.18**$^{*}$ | **−0.19**$^{*}$ | −0.11 | 0.20$^{*}$ | −0.17 |
| Utilization | −0.24$^{**}$ | **0.59**$^{***}$ | 0.09 | 0.13 | 0.28$^{***}$ | −0.21$^{*}$ | **−0.18**$^{*}$ | **−0.36**$^{***}$ | −0.35$^{***}$ | 0.00 | −0.27$^{**}$ |
| Spectral flux | | | | | | | | | | | |
| Validity | **−0.20**$^{*}$ | **0.62**$^{***}$ | 0.09 | **0.20**$^{*}$ | 0.14 | −0.01 | **−0.22**$^{*}$ | **−0.26**$^{**}$ | −0.13 | −0.01 | **−0.22**$^{*}$ |
| Utilization | **−0.28**$^{***}$ | **0.64**$^{***}$ | 0.22$^{*}$ | **0.31**$^{***}$ | 0.41$^{***}$ | −0.24$^{**}$ | **−0.22**$^{*}$ | **−0.56**$^{***}$ | −0.51$^{***}$ | −0.19$^{*}$ | **−0.38**$^{***}$ |
| MFCC2 | | | | | | | | | | | |
| Validity | **0.17**$^{*}$ | **−0.21**$^{*}$ | −0.12 | 0.00 | **−0.22**$^{*}$ | −0.10 | −0.03 | 0.10 | 0.05 | **0.21**$^{*}$ | 0.14 |
| Utilization | **0.19**$^{*}$ | **−0.22**$^{*}$ | −0.22$^{*}$ | −0.13 | **−0.28**$^{***}$ | 0.07 | 0.01 | 0.38$^{***}$ | 0.19$^{*}$ | **0.29**$^{***}$ | 0.26$^{**}$ |
| Salient pitch *M* | | | | | | | | | | | |
| Validity | −0.10 | −0.07 | 0.05 | −0.02 | 0.07 | 0.09 | 0.03 | −0.04 | 0.16 | **−0.22**$^{*}$ | 0.07 |
| Utilization | −0.08 | 0.03 | 0.20$^{*}$ | −0.05 | 0.02 | 0.15 | 0.02 | −0.17$^{*}$ | 0.05 | **−0.26**$^{**}$ | −0.12 |
| Key clarity | | | | | | | | | | | |
| Validity | 0.00 | −0.06 | −0.13 | 0.08 | 0.06 | −0.02 | 0.07 | 0.17 | −0.16 | 0.06 | −0.06 |
| Utilization | 0.12 | −0.18$^{*}$ | −0.37$^{***}$ | 0.31$^{***}$ | 0.18$^{*}$ | −0.18$^{*}$ | 0.04 | 0.15 | −0.12 | 0.11 | −0.05 |
| Mode | | | | | | | | | | | |
| Validity | 0.08 | −0.14 | 0.00 | **0.18**$^{*}$ | **0.35**$^{***}$ | −0.10 | 0.15 | −0.13 | **−0.19**$^{*}$ | −0.08 | −0.12 |
| Utilization | 0.02 | −0.07 | −0.03 | **0.45**$^{***}$ | **0.41**$^{***}$ | −0.42$^{***}$ | −0.03 | −0.12 | **−0.31**$^{***}$ | −0.20$^{*}$ | −0.31$^{***}$ |
| HCDF | | | | | | | | | | | |
| Validity | **0.23**$^{**}$ | **−0.24**$^{**}$ | −0.06 | −0.09 | −0.13 | 0.08 | **0.23**$^{**}$ | 0.08 | −0.01 | −0.15 | 0.05 |
| Utilization | **0.30**$^{***}$ | **−0.39**$^{***}$ | −0.13 | −0.17$^{*}$ | −0.25$^{**}$ | 0.37$^{***}$ | **0.22**$^{*}$ | 0.24$^{**}$ | 0.18$^{*}$ | 0.11 | 0.17$^{*}$ |
| Spectral novelty | | | | | | | | | | | |
| Validity | **0.20**$^{*}$ | **−0.31**$^{***}$ | 0.02 | **−0.20**$^{*}$ | **−0.33**$^{***}$ | 0.05 | **0.22**$^{*}$ | **0.24**$^{**}$ | 0.06 | 0.02 | 0.03 |
| Utilization | **0.19**$^{*}$ | **−0.33**$^{***}$ | −0.04 | **−0.45**$^{***}$ | **−0.52**$^{***}$ | 0.32$^{***}$ | **0.21**$^{*}$ | **0.41**$^{***}$ | 0.45$^{***}$ | 0.20$^{*}$ | 0.42$^{***}$ |
| Rhythm novelty | | | | | | | | | | | |
| Validity | **0.20**$^{*}$ | **−0.31**$^{***}$ | −0.04 | **−0.21**$^{*}$ | **−0.29**$^{***}$ | 0.01 | **0.25**$^{**}$ | **0.21**$^{*}$ | 0.04 | 0.06 | 0.09 |
| Utilization | **0.18**$^{*}$ | **−0.33**$^{***}$ | −0.11 | **−0.43**$^{***}$ | **−0.49**$^{***}$ | 0.29$^{***}$ | **0.23**$^{**}$ | **0.43**$^{***}$ | 0.43$^{***}$ | 0.21$^{*}$ | 0.43$^{***}$ |
| Tonal novelty | | | | | | | | | | | |
| Validity | **0.19**$^{*}$ | **−0.34**$^{***}$ | 0.00 | **−0.21**$^{*}$ | **−0.36**$^{***}$ | **0.19**$^{*}$ | 0.17 | 0.15 | 0.14 | 0.00 | 0.07 |
| Utilization | **0.19**$^{*}$ | **−0.36**$^{***}$ | −0.01 | **−0.50**$^{***}$ | **−0.54**$^{***}$ | **0.44**$^{***}$ | 0.17$^{*}$ | 0.36$^{***}$ | 0.52$^{***}$ | 0.17 | 0.39$^{***}$ |
| Register novelty | | | | | | | | | | | |
| Validity | 0.17 | **−0.38**$^{***}$ | 0.00 | **−0.25**$^{**}$ | **−0.34**$^{***}$ | **0.21**$^{*}$ | **0.19**$^{*}$ | **0.20**$^{*}$ | 0.11 | 0.03 | 0.07 |
| Utilization | 0.21$^{*}$ | **−0.39**$^{***}$ | −0.04 | **−0.51**$^{***}$ | **−0.56**$^{***}$ | **0.45**$^{***}$ | **0.20**$^{*}$ | **0.39**$^{***}$ | 0.52$^{***}$ | 0.23$^{**}$ | 0.43$^{***}$ |

*Note.* $N = 132$. A significant cue-validity correlation (i.e., the correlation between a cue and the musicians' intended expression) suggests that the cue is used in a consistent fashion by the musicians to convey a certain expression. A significant cue-utilization correlation (i.e., the correlation between a cue and the listeners' mean recognition accuracy) suggests that the cue is used in a consistent fashion to make inferences about the conveyed expression. Bold typeface indicates which cues were both valid and utilized for each expression. See Table 1 for an explanation of cue abbreviations.
$^{*}$ $p < .05$. $^{**}$ $p < .01$. $^{***}$ $p < .001$.

press) expressions suggest that partly overlapping temporofrontal circuits support emotion inference from both domains. Although music and speech share many important attributes, there are nevertheless some critical differences between the domains (Zatorre & Baum, 2012). Notably, pitch variations in music are mostly discrete (e.g., melodic intervals in melodies), whereas pitch variations in speech are continuous, and, as a consequence, music requires a more fine-grained representation of

Table 5
*Correlations (Pearson R) Between Patterns of Valid and Utilized Cues Are Shown As a Function of Intended Expression, Musical Culture, and Listener Culture*

| | Correlation between cue-validity and cue-utilization patterns | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Western classical music | | | Hindustani classical music | | | Japanese traditional music | | | Swedish folk music | | |
| Intended expression | Ind | Jap | Swe | Ind | Jap | Swe | Ind | Jap | Swe | Ind | Jap | Swe |
| Affection | .68 | .82 | .84 | .69 | .39 | .64 | −.67 | .27 | .50 | .07 | .17 | .16 |
| Anger | .99 | .99 | .99 | 1.00 | .97 | 1.00 | .82 | .74 | .87 | .77 | .92 | .99 |
| Fear | .72 | .88 | .78 | .88 | .15 | −.29 | .84 | .82 | .93 | .76 | .50 | .86 |
| Happiness | .91 | .86 | .86 | −.18 | −.08 | −.10 | .57 | .61 | .63 | .95 | .95 | .93 |
| Humor | .86 | .90 | .93 | .95 | .85 | .87 | .86 | .62 | .64 | .93 | .92 | .93 |
| Longing | .49 | .13 | .42 | .51 | .35 | .68 | .25 | .31 | .39 | .37 | .50 | .68 |
| Neutral | .69 | .90 | .89 | .88 | −.38 | .16 | −.10 | .86 | .65 | .77 | .88 | .93 |
| Peacefulness | .87 | .80 | .89 | .87 | .75 | .90 | .55 | .49 | .52 | .83 | .78 | .91 |
| Sadness | .87 | .80 | .73 | .61 | .32 | .34 | .55 | .58 | .29 | .52 | .59 | .58 |
| Solemnity | .49 | .86 | .81 | −.04 | −.08 | −.38 | .57 | .81 | .04 | −.16 | .49 | .28 |
| Spirituality | .53 | .55 | .64 | .36 | .11 | .09 | −.08 | .22 | .09 | .71 | .18 | .81 |
| Mean correlation | .81 | .85 | .85 | .77 | .44 | .57 | .46 | .62 | .58 | .70 | .72 | .84 |

*Note.* $N = 26$. A high validity-utilization correlation suggests a good match between the listeners' cue-utilization patterns and the musicians' cue-validity patterns. When calculating the mean correlations, the raw $r$ values were transformed to Fisher $z$ values before averaging, and the mean $z$ values were then transformed back to $r$ values (for consistency with the rest of the table). Correlations $r \geq .39$, $p < .05$; $r \geq .50$, $p < .01$; $r \geq .61$, $p < .001$.

pitch than speech. In addition, although musical rhythm is often periodic, speech rhythms are nonperiodic. Therefore, the hypothesis that music expresses affect using a code derived from vocal expression comes with the boundary condition that it does not apply to cues that are unique to music (Juslin & Laukka, 2003). For example, musical cultures exhibit great variability with regard to musical scales, including the number and tuning of scale notes, and the melodic intervals that can be formed based on the notes of the scale (e.g., Patel, 2008, pp. 16–22). The affective connotations of such musical cues may be largely shaped by cultural conventions over music's history, and involve implicit knowledge of a particular musical culture including perception of musical structures and the development of expectations for future musical events (Huron, 2006).

Universal communication of emotion further requires that the very emotion labels that are to be communicated are transferable across cultures (Mesquita & Frijda, 1992). Basic emotions are generally considered largely invariant across cultures (Ekman, 1992), whereas "nonbasic" affective states may show more cultural variability, both regarding how they are conceptualized and expressed (e.g., Matsumoto & Hwang, 2012). Recent studies have further proposed that knowledge of emotion words and concepts may provide an important context in emotion perception, and that conceptual knowledge may participate in the judgment of another person's emotion (Gendron, Lindquist, Barsalou, & Barrett, 2012). Applied to music, this would imply that cultural differences regarding the affective categories that are associated with expression in music may also be a source of cultural relativity, and several examples of how the language used to describe musical expression may differ between cultures have been reported by anthropological music researchers (Becker, 2010; Benamou, 2003).

## Which Emotions and Related States Can Be Universally Conveyed Through Music?

Focusing on recognition rates significantly higher than chance using our stricter chance level criterion (20%), the most prudent interpretation of the pattern of results suggests that decoding of basic emotions was fairly robust from both familiar and unfamiliar music, whereas above-chance recognition of "nonbasic" affective states was more limited and only occurred for some listener groups and/or for some (mostly familiar) musical cultures. These results are in line with the proposition that the universal component of musically expressed affect is largely a consequence of the universality of vocal expressions. Basic emotions are well recognized across cultures from vocal expressions, and are also characterized by relatively distinct patterns of cues that are similar for both speech and music (Juslin & Laukka, 2003). In contrast, for some "nonbasic" affective states, such as solemnity and longing, there is currently no evidence that they can be expressed nonverbally through the voice. If musical expression of "nonbasic" affective states thus depends on the utilization of music- and culture-specific cues, rather than on cues shared with vocalizations, this could help explain the difficulties of communicating them across cultures. Humor also received fairly high cross-cultural recognition rates, which may suggest that also some "nonbasic" affective states can be robustly decoded from unfamiliar music. However, humor was frequently confused with happiness, and it therefore remains a possibility that cross-cultural recognition of humor may have resulted from both conceptual and acoustic similarity with happiness. Spirituality was not accurately recognized in any condition and may not be recognizable from purely auditory information without appropriate contextual cues, such as information from the situation in which "spiritual" music is usually encountered.

In addition, basic emotions are commonly associated with musical expression in both Western (e.g., Gabrielsson & Juslin, 2003) and Eastern cultures (e.g., in the Indian rasa theory; see Balkwill & Thompson, 1999), whereas there may be cultural variation with regard to how applicable "nonbasic" affect labels—albeit important in a Western musical context—are to music in other traditions (e.g., Benamou, 2003). It should be noted that our list of expression labels was based on studies rooted in a Western musical

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

context, and thus the included "nonbasic" affective states may have been more applicable to Western than non-Western music. This could help explain the finding that a wider palette of expressions was recognized from Western classical and Swedish folk music compared with Hindustani classical and Japanese traditional music.

Finally, Brunswik's (1956) concept of vicarious functioning entails that performers and listeners may use many partially interchangeable cues in flexible ways, sometimes shifting from one cue that is unavailable to another that is available (see Juslin, 2000). Such a communication system allows for cultural specificity in both encoding and decoding of musical affect expressions, but also limits the complexity of the information that can be conveyed (because each cue is partly redundant; see Juslin & Laukka, 2003). Thus, cue redundancy could also be part of an explanation of the observed pattern of results, where communication of expressions beyond basic emotions was less robust.

## Which Musical Features Are Universally Associated With Affect?

Our study demonstrated the utility of using a lens-model analysis in cross-cultural research (Scherer et al., 2011), and illustrated how a large selection of acoustic and musical cues were associated with the musician's intentions and the listener's judgments for different emotions and related states within and across musical cultures (see Table 4 and Table S2). In general, it seems that the included cues cannot easily be categorized as either universal or culture specific. Cultural effects rather appear to be more subtle—where each cue is neither completely universal nor completely culture specific—and may manifest in cultural relativity in how a particular cue is used to express a particular affective character. This is consistent with the dialect theory of emotions (Elfenbein et al., 2007), which suggests that the in-group advantage for emotion decoding results from subtle "dialects" in the way that encoders use various cues to express emotions.

Based on the similarities between music and speech, it might be expected that cross-cultural cue utilization would be more consistent for acoustic cues shared between music and speech, compared with music-only features. However, we were not able to find such a pattern in our results. This may be because several cues reflect the shared contribution of the musician's performance strategies and the musical structure of the performed piece, and thus the division of features into acoustic and musical cues is in many cases not clear-cut. For example, a high event density reflects both the tempo of the performance (a high tempo is associated with a high event density) and the number of tones per beat of the melody. Similarly, tonal cues and pulse clarity are related to both musical structure (types of intervals, chords, and periodic patterns) and expressive performance (primarily by tempo and how individual musical events are emphasized by the performer). Some aspects of musical structure, such as consonance and dissonance, and the major/minor distinction—which originate from basic perceptual and psychophysical processes—can also be expected to be largely universal (e.g., Fritz et al., 2009), and share affective connotations with speech (Bowling et al., 2012). One way to sharpen the distinction between universal and culture-specific cues could therefore be to delve into the deeper semantic content of musical cultures. For example, culture-specific cues may be characterized by recognizable chord patterns and melodic motives associated with particular cultural meanings (e.g., solemnity may be associated with the melody of a national anthem).

## Limitations and Future Directions

The present study is also subject to some limitations that could be addressed by future studies. In particular, the small number of performers from each musical culture makes it hard to draw conclusions about cultural differences in encoding. Future studies should try to include a larger number of encoders from each culture to be able to separate within- and cross-cultural sources of variation in expressive styles. In addition, our results regarding cue utilization are based on correlations, and future research using systematic manipulation of cues (e.g., Bhatara, Tirovolas, Duan, Levy, & Levitin, 2011) is needed to draw conclusions about causal relations between culture-specific performance strategies and listeners' affective judgments across cultures. It would also be worthwhile to investigate additional musical cultures, using different types of instruments and musical material, to increase generality. Further comparative studies of expression in speech and music are also warranted, for example, to determine the limits of the range of emotions and related states that can be vocally (or musically) expressed, and whether these expressions can correspondingly be conveyed through music (or speech).

The recognition rates obtained in our study were rather low compared with those reported in many previous studies on nonverbal emotion expression, especially when compared with studies on facial expressions. Several aspects of our design may have contributed to this—for example, we did not use preselection of stimuli (see Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008), and we also included many expression labels that we expected from the outset would be difficult to convey musically, including several "nonbasic" affective states. In particular, we used an 11-alternative judgment task, whereas most previous studies used much fewer response options. The use of a large number of conceptually close response options leads to a more cognitively demanding task, with more room for disagreement, compared with the use of only a few highly differentiated response options (Bänziger et al., 2012). We observed especially low recognition rates for Japanese traditional music—and it remains a possibility that the failure to find an in-group advantage for Japanese music may have been caused by floor effects introduced by the relative difficulty of the Japanese stimuli compared with stimuli from other cultures. It is thus possible that a wider range of emotions and related states would have been recognizable across cultures had the clarity of the stimuli been higher. Therefore, we would welcome future studies that try both to increase the clarity of the musical stimuli, and to systematically investigate how the in-group advantage is affected by the difficulty and response format of the affect recognition task. Another explanation for the low recognition rates for Japanese traditional music could be that judges were especially unfamiliar with this musical culture. Although Japanese listeners reported being familiar with Japanese traditional music in general, they may not have been acquainted with kokyū music in particular because this instrument is rarely encountered in Japan today (Hughes, 2001).

Researchers have further suggested that an in-group advantage for emotion recognition may result from individuals being less motivated to decode accurately facial expressions from members of visibly different cultural groups (e.g., Kilbride & Yarczower, 1983). Because the cultural origin of the musical excerpts was readily perceivable (e.g., from the tonal material and the instrument timbre), there remains a possibility that motivational factors may have influenced our results. Future studies could directly address this issue by, for example, using systematic cue manipulation to create standardized musical stimuli that use exactly the same cue patterns to convey expressions in different musical styles, thereby erasing cultural differences in expressive style while preserving the cultural influence on musical structure. If an in-group advantage can still be observed for such standardized stimuli, then one could conclude that motivational factors may play a role for cross-cultural affect recognition. Prior studies using this technique with facial expressions have reported that the in-group advantage does not result from motivational factors but from cultural variability in expressive styles (see Elfenbein et al., 2007).

Musical performances are complex and difficult to measure, and although we used state-of-the-art methods, the included cues can only give a crude description of what is going on in the music. For example, most cues were averaged over the duration of the whole performance, whereas listeners' judgments may have been influenced by specific musical events that are not reflected in the averaged measures. Future studies should therefore develop measures that better capture the temporal distribution of emotion-relevant events in the musical signal. Studies would also benefit from developing and including a larger selection of musical cues (including fine-grained analyses of the musical structure) in order to test the idea that such cues are more culturally variable. In particular, development of novel musical cues should focus on features appropriate for capturing aspects of non-Western musical structures (e.g., Gedik & Bozkurt, 2010; Gómez & Herrera, 2008).

Limitations aside, the results suggest that listeners are sensitive to the affective content of unfamiliar music, and, in this narrow sense, music may be a "universal language of emotions." However, like most languages, expressivity in music may also come in different dialects, which hinder listeners' understanding of unfamiliar music, as reflected in our findings of more accurate decoding, and better matching between encoders' and decoders' uses of cues, in in-group versus out-group conditions. The present study thus contributes to the current strand of research that tries to integrate evidence for both universality and cultural relativity in emotion expression, and we hope that it may inspire further research on how expressive aspects of music are transmitted within and across cultures. Especially, the methods used in the present study, including the use of a balanced design, and employing a lens-model analysis to illustrate how encoders and decoders use various cues to express and decode emotions and related states, may profitably be used to study affective communication also in other modalities such as facial, vocal, and bodily expression.

## References


Adachi, M., Trehub, S. E., & Abe, J.-I. (2004). Perceiving emotion in children's songs across age and culture. *Japanese Psychological Research, 46,* 322–336. doi:10.1111/j.1468-5584.2004.00264.x

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. doi:10.2307/40285811

Balkwill, L.-L., Thompson, W. F., & Matsunaga, R. (2004). Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners. *Japanese Psychological Research, 46,* 337–349. doi:10.1111/j.1468-5584.2004.00265.x

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal Expression Corpus for experimental research on emotion perception. *Emotion, 12,* 1161–1179. doi:10.1037/a0025827

Becker, J. (2010). Exploring the habitus of listening: Anthropological perspectives. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion. Theory, research, and applications* (pp. 128–157). New York, NY: Oxford University Press.

Benamou, M. (2003). Comparing musical affect: Java and the West. *The World of Music, 45,* 57–76.

Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 921–934. doi:10.1037/a0021922

Bor, J. (1987). The voice of the sarangi: An illustrated history of bowing in India. *Quarterly Journal of the National Centre for the Performing Arts (India), 15 & 16* , 1–183.

Bowling, D. L., Sundararajan, J., Han, S., & Purves, D. (2012). Expression of emotion in Eastern and Western music mirrors vocalization. *PLoS ONE, 7,* e31942. doi:10.1371/journal.pone.0031942

Brown, S., & Jordania, J. (in press). Universals in the world's musics. *Psychology of Music*. doi:10.1177/0305735611425896

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

Budd, M. (1985). *Music and the emotions: The philosophical theories*. London, UK: Routledge.

Burkholder, J. P., Grout, D. J., & Palisca, C. V. (2009). *A history of western music* (8th ed.). New York, NY: Norton.

Collier, G. L. (2002). Why does music express only some emotions? A test of a philosophical theory. *Empirical Studies of the Arts, 20,* 21–31. doi:10.2190/XXD1-W9R4-H81L-ADP2

Cross, I. (2008). Musicality and the human capacity for culture. *Musicae Scientiae, 12,* 147–167. doi:10.1177/1029864908012001071

Dailey, M. N., Joyce, C., Lyons, M. J., Kamachi, M., Ishi, H., Gyoba, J., & Cottrell, G. W. (2010). Evidence and a computational explanation of cultural differences in facial expression recognition. *Emotion, 10,* 874–893. doi:10.1037/a0020019

Darrow, A.-A., Haack, P., & Kuribayashi, F. (1987). Descriptors and preferences for Eastern and Western musics by Japanese and American nonmusic majors. *Journal of Research in Music Education, 35,* 237–248. doi:10.2307/3345076

Davies, S. (2006). Artistic expression and the hard case of pure music. In M. Kieran (Ed.), *Contemporary debates in aesthetics and the philosophy of art* (pp. 179–191). Oxford, UK: Blackwell. doi:10.1093/acprof:oso/9780199608775.003.0002

Davies, S. (2011). Cross-cultural musical expressiveness: Theory and the empirical programme. In E. Schellekens & P. Goldie (Eds.), *The aesthetic mind: Philosophy and psychology* (pp. 376–388). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199691517.003.0023

Eerola, T. (2011). Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research, 40,* 349–366. doi:10.1080/09298215.2011.602195

Eerola, T., Lartillot, O., & Toiviainen, P. (2009). Prediction of multidimensional emotional ratings in music from audio using multivariate regression models. In K. Hirata, G. Tzanetakis, & K. Yoshii (Eds.), *Proceedings of the 10th International Conference on Music Information*


This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

*Retrieval* (pp. 621–626). Kobe, Japan: International Society for Music Information Retrieval.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200. doi:10.1080/02699939208411068

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128,* 203–235. doi:10.1037/0033-2909.128.2.203

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: Cultural differences in the expression and recognition of posed facial expressions. *Emotion, 7,* 131–146. doi:10.1037/1528-3542.7.1.131

Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., & Kumar, S. (2002). Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques. *Emotion, 2,* 75–84. doi:10.1037/1528-3542.2.1.75

Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (in press). Emotional expression in voice and music: Same code, same effect? *Human Brain Mapping*. doi:10.1002/hbm.22029

Frank, M. G., & Stennett, J. (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology, 80,* 75–85. doi:10.1037/0022-3514.80.1.75

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., . . . Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology, 19,* 573–576. doi:10.1016/j.cub.2009.02.058

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music, 24,* 68–91. doi:10.1177/0305735696241007

Gabrielsson, A., & Juslin, P. N. (2003). Emotional expression in music. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 503–534). New York, NY: Oxford University Press.

Gabrielsson, A., & Lindström, E. (2010). The influence of musical structure on emotional expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, and applications* (pp. 223–248). New York, NY: Oxford University Press.

Gedik, A. C., & Bozkurt, B. (2010). Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing, 90,* 1049–1063. doi:10.1016/j.sigpro.2009.06.017

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion words shape emotion percepts. *Emotion, 12,* 314–325. doi:10.1037/a0026007

Gómez, E., & Herrera, P. (2008). Comparative analysis of music recordings from western and non-western traditions by automatic tonal feature extraction. *Empirical Musicology Review, 3,* 140–156.

Gregory, A. H., & Varney, N. (1996). Cross-cultural comparisons in the affective response to music. *Psychology of Music, 24,* 47–52. doi:10.1177/0305735696241005

Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality, 42,* 1476–1489. doi:10.1016/j.jrp.2008.06.013

Hertenstein, M. J., Keltner, D., App, B., Bulleit, B. A., & Jaskolka, A. R. (2006). Touch communicates distinct emotions. *Emotion, 6,* 528–533. doi:10.1037/1528-3542.6.3.528

Hevner, K. (1936). Experimental studies of the elements of expression in music. *The American Journal of Psychology, 48,* 246–268. doi:10.2307/1415746

Higgins, K. M. (2006). The cognitive and appreciative impact of musical universals. *Revue Internationale de Philosophie, 60,* 487–503.

Higgins, K. M. (2012). Biology and culture in musical emotions. *Emotion Review, 4,* 273–282. doi:10.1177/1754073912439762

Hoshino, E. (1996). The feeling of musical mode and its emotional character in a melody. *Psychology of Music, 24,* 29–46. doi:10.1177/0305735696241004

Hughes, D. W. (2001). Kokyū. In S. Sadie & J. Tyrrell (Eds.), *The new Grove dictionary of music and musicians* (2nd ed.). Retrieved from http://www.oxfordmusiconline.com/subscriber/article/grove/music/48402

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

Jersild, M., & Ramsten, M. (2001). Sweden. II. Traditional music. In S. Sadie & J. Tyrrell (Eds.), *The new Grove dictionary of music and musicians* (2nd ed.). Retrieved from http://www.oxfordmusiconline.com/subscriber/article/grove/music/27199

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,* 1797–1812. doi:10.1037/0096-1523.26.6.1797

Juslin, P. N., & Laukka, P. (2003). Communication of emotion in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814. doi:10.1037/0033-2909.129.5.770

Juslin, P. N., & Laukka, P. (2004). Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research, 33,* 217–238. doi:10.1080/0929821042000317813

Juslin, P. N., & Timmers, R. (2010). Expression and communication of emotion in music performance. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion. Theory, research, and applications* (pp. 453–489). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780199230143.003.0017

Keil, A., & Keil, C. (1966). Musical meaning: A preliminary report (The perception of Indian, Western, and Afro-American musical moods by American students). *Ethnomusicology, 10,* 153–173. doi:10.2307/924696

Kilbride, J. E., & Yarczower, M. (1983). Ethnic bias in the recognition of facial expressions. *Journal of Nonverbal Behavior, 8,* 27–41. doi:10.1007/BF00986328

Kwoun, S. J. (2009). An examination of cue redundancy theory in cross-cultural decoding of emotions in music. *Journal of Music Therapy, 46,* 217–237.

Lartillot, O., & Toiviainen, P. (2007). MIR in Matlab (II): A toolbox for musical feature extraction from audio. In S. Dixon, D. Bainbridge, & R. Typke (Eds.), *Proceedings of the 8th International Conference on Music Information Retrieval* (pp. 237–244). Vienna, Austria: Austrian Computer Society.

Laukka, P. (2004). Instrumental music teachers' views on expressivity: A report from music conservatoires. *Music Education Research, 6,* 45–56. doi:10.1080/1461380032000182821

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., . . . Gur, R. C. (2010). "It's not what you say, but how you say it": A reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience, 4,* 19. doi:10.3389/fnhum.2010.00019

Malm, W. P. (2000). *Traditional Japanese music and musical instruments* (rev. ed.). Tokyo, Japan: Kodansha International.

Matsumoto, D., & Hwang, H. S. (2012). Culture and emotion: The integration of biological and cultural contributions. *Journal of Cross-Cultural Psychology, 43,* 91–118. doi:10.1177/0022022111420147

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology, 18,* 452–463. doi:10.1016/j.conb.2008.09.005

Merriam, A. P. (1964). *The anthropology of music*. Evanston, IL: Northwestern University Press.

Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112,* 179–204. doi:10.1037/0033-2909.112.2.179

Morrison, S. J., & Demorest, S. M. (2009). Cultural constraints on music perception and cognition. *Progress in Brain Research, 178,* 67–77. doi:10.1016/S0079-6123(09)17805-6

Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: Oxford University Press.

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37,* 417–435. doi:10.1016/j.wocn.2009.07.005

Ruckert, G. E. (2004). *Music in North India: Experiencing music, expressing culture*. New York, NY: Oxford University Press.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115,* 102–141. doi:10.1037/0033-2909.115.1.102

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 107,* 2408–2412. doi:10.1073/pnas.0908239106

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143–165. doi:10.1037/0033-2909.99.2.143

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice, 9,* 235–248. doi:10.1016/S0892-1997(05)80231-0

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76–92. doi:10.1177/0022022101032001009

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology, 46,* 401–435. doi:10.1080/00207594.2011.626049

Schirmer, A., & Kotz, S. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24–30. doi:10.1016/j.tics.2005.11.009

Simon-Thomas, E. R., Keltner, A. J., Sauter, D., Sinicropi-Yao, L., & Abramson, A. (2009). The voice conveys specific emotions: Evidence from vocal burst displays. *Emotion, 9,* 838–846. doi:10.1037/a0017810

Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science, 4,* 653–667. doi:10.1111/j.1756-8765.2012.01215.x

Thompson, W. F., & Balkwill, L.-L. (2006). Decoding speech prosody in five languages. *Semiotica, 158,* 407–424. doi:10.1515/SEM.2006.017

Thompson, W. F., & Balkwill, L-L. (2010). Cross-cultural similarities and differences. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion. Theory, research, and applications* (pp. 755–788). New York, NY: Oxford University Press.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107,* 885–913. doi:10.1037/0033-295X.107.4.885

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary, and peaceful musical excerpts for research on emotions. *Cognition and Emotion, 22,* 720–752. doi:10.1080/02699930701503567

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17,* 3–28. doi:10.1007/BF00987006

Walker, R. (1996). Open peer commentary: Can we understand the music of another culture? *Psychology of Music, 24,* 103–114. doi:10.1177/0305735696242001

Wedin, L. (1969). Dimension analysis of emotional expression in music. *Swedish Journal of Musicology, 51,* 119–140.

Wieczorkowska, A. A., Datta, A. K., Sengupta, R., Dey, N., & Mukherjee, B. (2010). On search for emotion in Hindusthani vocal music. In Z. W. Ras & A. A. Wieczorkowska (Eds.), *Advances in music information retrieval* (pp. 285–304). Berlin, Germany: Springer. doi:10.1007/978-3-642-11674-2_13

Wong, P. C. M., Roy, A. K., & Margulis, E. H. (2009). Bimusicalism: The implicit dual enculturation of cognitive and affective systems. *Music Perception, 27,* 81–88. doi:10.1525/mp.2009.27.2.81

Zacharopoulou, K., & Kyriakidou, A. (2009). A cross-cultural comparative study of the role of musical structural features in the perception of emotion in Greek traditional music. *Journal of Interdisciplinary Music Studies, 3,* 1–15.

Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different melody? *PLoS Biology, 10,* e1001372. doi:10.1371/journal.pbio.1001372

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Differentiation, classification, and measurement. *Emotion, 8,* 494–521. doi:10.1037/1528-3542.8.4.494

Received March 28, 2012
Revision received November 28, 2012
Accepted November 29, 2012 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.