Japanese Psychological Research
2004, Volume 46, No. 4, 337–349
Special Issue: Cognition and emotion in music

Recognition of emotion in Japanese, Western, and Hindustani music by Japanese listeners[1]

LAURA-LEE BALKWILL[2]
Queen's University, Kingston, Ontario K7L 3N6, Canada

WILLIAM FORDE THOMPSON
University of Toronto, Mississauga, Ontario M5S 1A1, Canada

RIE MATSUNAGA
Department of Psychology, Graduate School of Letters, Hokkaido University, Kita-ku, Sapporo 060-0810, Japan

Abstract: Japanese listeners rated the expression of joy, anger and sadness in Japanese, Western, and Hindustani music. Excerpts were also rated for tempo, loudness, and complexity. Listeners were sensitive to the intended emotion in music from all three cultures, and judgments of emotion were related to judgments of acoustic cues. High ratings of joy were associated with music judged to be fast in tempo and melodically simple. High ratings of sadness were associated with music judged to be slow in tempo and melodically complex. High ratings of anger were associated with music judged to be louder and more complex. The findings suggest that listeners are sensitive to emotion in familiar and unfamiliar music, and this sensitivity is associated with the perception of acoustic cues that transcend cultural boundaries.

Key words: recognition of emotion, music, cross-culture.

Music is strongly associated with emotions. Evocative music is used in advertising, television, movies, and the music industry, and the effects are powerful. Listeners readily interpret the emotional meaning of music by attending to specific properties of the music (Hevner, 1935; Rigg, 1940). For example, joyful music is typically fast in tempo, major in mode, wide in pitch range, high in loudness, regular in rhythm, and low in complexity (Behrens & Green, 1993; Deva & Virmani, 1975; Dolgin & Adelson, 1990; Gabrielsson & Juslin, 1996; Gerardi & Gerken, 1995; Hoshino, 1996; Kratus, 1993; Robazza, Macaluso, & D'Urso, 1994; Thompson & Robitaille, 1992; Terwogt & Van Grinsven, 1991). In some cases, listening to music may give rise to changes in mood and arousal (Husain, Thompson, & Schellenberg, 2002; Thompson, Schellenberg, & Husain, 2001).

Certain properties of music, such as tempo and loudness, might provide universal cues to emotional meaning. It is difficult to assess this hypothesis, however, because most research on the topic has involved asking listeners to judge the emotional meaning of music from their own culture (cf. Gregory & Varney, 1996), and most of these studies only involved Western music (cf. Deva & Virmani, 1975; cf. Hoshino, 1996).

[1] This research was supported by grants to the first and second authors from the Natural Sciences and Engineering Research Council. We thank Doug Gifford and Kevin Cooke for technical assistance.

[2] Correspondence concerning this article should be sent to: Prof. Laura-Lee Balkwill, Queens University, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada. (Email: lbalk@psyc.queensu.ca)

In one exception, Deva and Virmani presented Indian listeners with excerpts from Hindustani ragas and asked them to judge the mood, color, season, and time of day that best characterized the ragas. Judgments of ragas were often consistent with conventional (intended) associations, suggesting that listeners were sensitive to properties in the ragas that connote moods, colors and temporal associations.

As another example, Gregory and Varney (1996) asked Western and Indian listeners to assess the emotional content of commercially recorded excerpts of Hindustani ragas, Western classical music, and Western new-age music. Listeners indicated their assessment of the emotional content in the pieces by selecting adjectives from a list of mood terms taken from Hevner (1936). Western and Indian listeners were sensitive to intended emotions in Western music, but not Hindustani ragas. Methodological details may account for the lack of sensitivity to intended emotion in ragas, however. Because the experimenters used commercially recorded music, they could not confirm the mood intended by performers. Performers have considerable latitude in their expression of emotion, and textbook descriptions of ragas do not always reflect the mood intended in a given performance. This complication is particularly serious for commercial recordings, which contain production elements that enhance the popularity of music but may obscure or conflict with the intended mood specified by the raga-rasa system.

Balkwill and Thompson (1999) reported that acoustic cues in Hindustani music influenced the judgment of emotion in that music. Acoustic cues are operationally defined as properties of the acoustic signal that can be perceived and interpreted without explicit or implicit knowledge of music conventions. By definition, they are interpretable in music across cultures, unlike culture-specific cues, which rely on knowledge of music conventions. Acoustic cues include qualities such as loudness, pace (tempo), and timbre.

During a field research trip to India, the first author recorded Hindustani ragas that were performed with the intention to evoke specific moods. For each excerpt, Western listeners judged the emotional meaning and the presence of specific acoustic cues. Excerpts intended to convey joy/*hasya* were assigned high ratings of joy; excerpts intended to convey sadness/*karuna* were assigned high ratings of sadness; and excerpts intended to convey anger/*raudra* were assigned high ratings of anger. High ratings of joy were associated with high ratings of tempo and low ratings of melodic complexity. High ratings of sadness were associated with low ratings of tempo and high ratings of melodic complexity. Anger ratings were not significantly associated with ratings of tempo or complexity, but ragas performed on stringed instruments were rated much higher in anger than those performed on the flute.

Balkwill and Thompson (1999) interpreted these results within the framework of their cue redundancy model for the representation and recognition of emotion in music (see Figure 1). A central tenet of the cue redundancy model is that listeners can appreciate affective qualities of unfamiliar music by attending to acoustic cues. Consciously or intuitively, composers and performers draw upon acoustic cues, as well as culturally determined conventions, in order to express emotion in music. Listeners, in turn, attend to either or both of these sources of emotional meaning. Because emotion is conveyed through both culture-specific and acoustic cues, there is redundancy in how a piece reinforces a specific emotion. The greater the number of cues in the music (both culture-specific and acoustic) the stronger the expression of emotion. Listeners' ability to decode musically expressed emotion is affected both by their familiarity with the conventions of the tonal system and by their sensitivity to acoustic cues. The expression of emotion in the music of a given culture is most salient to listeners of that same culture, because such listeners have acquired the ability to decode conventional cues of emotion. When cultural-specific cues are absent, listeners may still attend to acoustic cues such as tempo and loudness. These cues provide listeners with a general understanding of the intended emotion.

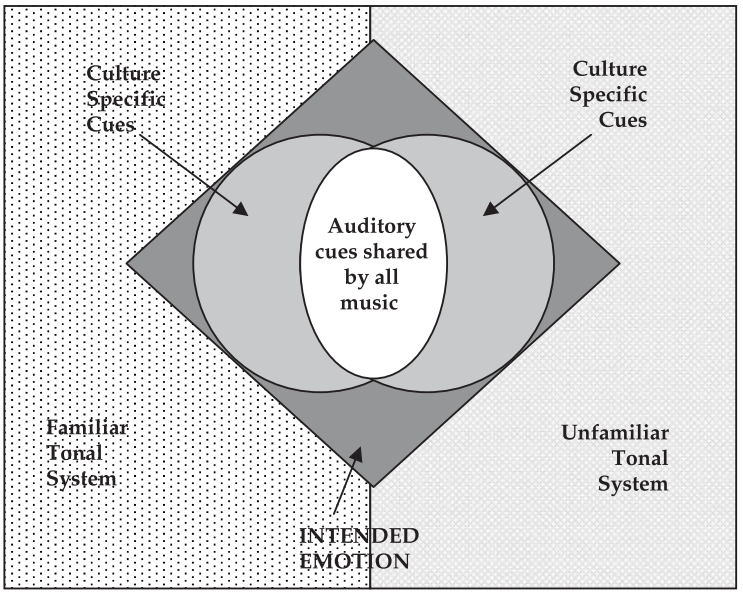

Figure 1. The cue redundancy model (Balkwill & Thompson, 1999).

The present study was designed to provide additional evidence for the cue redundancy model. In two experiments, Japanese listeners rated the degree of anger, joy, sadness, complexity, loudness, and tempo in Western, Japanese and Hindustani music. Japanese listeners represent a particularly interesting population because they are familiar with Western music but are also enculturated to Japanese music through children's songs, teaching songs, traditional festival songs, and music accompanying government-sponsored cultural events (Oku, 1995). Familiarity with two musical systems (Japanese and Western) might influence sensitivity to emotional meaning in music. Japanese listeners are not typically exposed to Hindustani music, however, which allowed us to assess their sensitivity to emotion in unfamiliar music.

Japanese, Western, and Hindustani music were examined because they are distinctive musical systems that are emotionally evocative. Western music relies mainly on three scales; the major, harmonic minor, and melodic minor. These scales are based on equal-tempered tuning, and any combination of scale notes creates a pitch interval that is a multiple of a semitone (the smallest pitch interval used in Western tonal music). Japanese music is based on two pentatonic scales, one that uses semitones (the "In" scale) and one that does not (the "Yo" scale). The size of the semitone also varies as a function of the song being played, the instrument, the sex of the player, and other factors. Finally, Hindustani music has more than 400 scales, one for each raga, each with differing sizes of semitones (Daniélou, 1980)

Skilled performers in Japan, Canada, and India performed the music used in the study. For each music set (Japanese, Western, Hindustani), half of the recordings were performed using a stringed instrument and half were performed using a flute instrument. Performers were asked to select music that they would normally choose to play for the purpose of communicating joy, sadness, and anger. Selections by North Indian performers were typically the *alap* (or introductory) section of mood-specific *ragas*. Selections by Western musicians were typically improvised variations of familiar material, and selections by Japanese musicians were typically standard repertoire pieces from *gagaku*,

shintou, or *minyo* genres appropriate to the emotive context. Recordings were then edited into shorter segments (mean = 30 s). Each music set consisted of 10 excerpts that were intended to evoke anger, 10 that were intended to evoke joy, and 10 that were intended to evoke sadness.

Experiment 1: Judgment of emotion within and across cultures

Japanese listeners were asked to judge the intensity of one emotion (either joy, anger, or sadness) in excerpts from Japanese, Western, and Hindustani music. We expected that Japanese listeners would be able to distinguish between anger, joy, and sadness in music from all cultures. They are familiar with traditional Japanese music from childhood and Western music is popular in Japan. Thus, both acoustic and culture-specific cues to emotion should be available in Japanese and Western music, making it easy for listeners to decode emotional meaning. Hindustani music is unfamiliar to Japanese listeners. Thus, culture-specific cues to emotion should be unavailable in this music. Nonetheless, listeners should be able to decode emotional meaning in Hindustani music by attending to acoustic cues.

Method

Listeners. There were 147 Japanese listeners (76 women, 71 men, mean age = 23.7) who participated on a volunteer basis without financial or academic compensation. They consisted of students (undergraduate and graduate) and staff members at the Department of Psychology at Hokkaido University, as well as associates of these participants (friends and colleagues). All participants were born in Japan and cited Japanese as their first language. Of the participants, 25.32% had no musical training, 23.32% had up to 5 years of training, and 51.46% had >5 years of training. Participants were recruited with an aim of achieving a relatively equal balance between men and women.

Thirty-six listeners rated Japanese music (18 women and 18 men, mean age = 23.77 years): 12 provided anger ratings (six women, six men), 12 provided joy ratings (six women, six men), and 12 provided sadness ratings (six women, six men). Fifty-six listeners rated Western music (30 women and 26 men, mean age = 23.21 years): 18 provided anger ratings (10 women, eight men), 20 provided joy ratings (11 women, nine men), and 18 provided sadness ratings (nine women, nine men). Fifty-five listeners rated Hindustani music (28 women and 27 men, mean age = 24.12 years): 19 provided anger ratings (10 women, nine men), 19 provided joy ratings (nine women, 10 men), and 17 provided sadness ratings (nine women, eight men).

Materials

Each musician was asked to play music they would normally play in order to evoke each of the three intended emotions (indicated by Japanese, Western, and Hindustani musicians as: (i) *ikari*/anger/*raudra*, (ii) *tanoshii*/joy/*hasya*, and (iii) *kanashii*/sadness/*karuna*). The term "intended emotion" refers to the musicians' stated intention to evoke a specific emotion in their performance. That is, intended emotion is defined by the performer's own indication of the emotion that was expressed, a procedure known as target scoring (Kruglanski, 1989; Roberts, Zeidner, & Matthews, 2001). All of the participating musicians were professional performers with more than 20 years of performance experience. Each musician was recorded separately with the assistance of translators when necessary.

Six Japanese musicians, three who played a flute instrument (shakuhachi, ryuteki, or shinobue), and three who played a stringed instrument (shamisen or koto), performed the Japanese excerpts (range = 24.00–37.25 s). The recording sessions took place during a field research trip to Sapporo, Hokkaido (2000–2002). Two Canadian musicians, one who played a flute, and the other who played a guitar, performed the Western excerpts (range = 22.0–37.5 s). Two Hindustani musicians, one who played the Bansuri flute, and the other who played stringed instruments (sitar, surbahar,

Table 1. Objective measures of selected musical stimuli

Music set	Duration	Event density	*M*-ff	*SD*-ff	Diff-ff	*M*-dB	*SD*-dB	Diff-dB
Top-rated angry excerpts								
Western	20.66	4.55	264.39	148.65	558.74	65.47	2.37	25.30
Japanese	28.42	1.86	503.99	86.79	427.70	78.23	11.86	70.20
Hindustani	31.16	1.67	300.68	129.15	533.95	64.07	6.56	34.38
Means	26.74	2.69	356.35	121.53	506.79	69.25	6.93	43.29
Top-rated joyful excerpts								
Western	31.42	2.42	201.13	87.68	527.98	64.34	7.20	51.31
Japanese	33.45	2.96	301.72	166.86	559.74	72.67	11.99	68.69
Hindustani	32.67	3.12	248.57	147.19	549.41	62.65	5.84	30.42
Means	32.51	2.83	250.47	133.91	545.71	66.55	8.34	50.14
Top-rated sad excerpts								
Western	33.98	1.03	245.27	136.68	549.62	59.92	10.26	45.07
Japanese	33.10	1.18	438.62	90.64	527.69	67.09	10.27	48.58
Hindustani	30.77	0.84	272.05	138.57	527.60	60.15	8.80	38.53
Means	32.61	1.01	318.64	121.96	534.83	62.38	9.77	44.06

Diff-ff, difference between minimum and maximum fundamental frequency; Diff-dB, difference between minimum and maximum intensity; *M*-dB, mean intensity; *M*-ff, mean fundamental frequency; *SD*-dB, standard deviation of intensity; *SD*-ff, standard deviation of fundamental frequency.

and dilruba), performed the Hindustani excerpts (range = 24.5–38.0 s). The latter recording sessions took place during a field research trip to Pune, India (1996–1997). Each musician was asked to play three pieces appropriate for each emotion. These recordings resulted in the following three music sets corresponding to the Japanese, Western, and Hindustani performances.

Japanese and Western performances were recorded onto digital video tape using a Sony DCR-TRV900 (Tokyo, Japan) mini digital video camera with an external ECM-MS907 stereo shotgun microphone, and downloaded as digital audio using the Adobe Premier video editing program. Hindustani performances were recorded using an AIWA (Tokyo, Japan) portable cassette recorder and an external electret stereo microphone, and converted to digital audio using SoundEdit software (MacroMedia, San Francisco, USA). For each music set, 10 excerpts (those with the best sound quality) for each intended emotion were selected for use in the experiment. These excerpts were edited in SoundEdit for length and converted to QuickTime movie files for use in the experiment.

Table 1 provides objective measures of selected acoustic characteristics for one excerpt from each music set and emotion. The excerpts selected for this analysis were those that most successfully communicated the intended emotion to listeners, that is, excerpts that were assigned the highest ratings of the intended emotion. Table 1 lists measures of duration and event density, as well as the mean, standard deviation, and range of the intensity and fundamental frequency of the music. An examination of acoustic properties of all stimuli is beyond the scope of this study, and the table provides descriptions of only our most emotive stimuli.

With the exception of event density, each measure was extracted using the software program Praat (Boersma & Weenink, 1999). Event density was calculated using a visual and auditory count of changes in pitch, including peaks and valleys of trills. The total number of events was divided by the duration of the piece to arrive at an estimate of event density. This may be considered a rough measure of how fast events unfolded in the music (i.e., tempo). As these analyses were not conducted on all musical stimuli in the present study, no firm

conclusions can be drawn. However, data for these samples illustrate that acoustic measures differed depending on the intended emotion. For example, among these exemplars, estimates of event density and loudness were lower for sad music than they were for angry or joyful music.

Procedure. Joy, anger, and sadness ratings were obtained for each music excerpt. Each listener provided one emotion rating for one music set. For example, a listener might be asked to rate joy for each excerpt in the Western music set. Ratings were assigned on a scale from 0 to 9, on which a rating of zero indicated that the emotional quality was not communicated in the music and a rating near 9 indicated that the emotion was clearly communicated in the music. Responses were made using a G3 Macintosh powerbook computer (Apple, Cupertino, CA, USA) in a sound-attenuated booth. Listeners heard the music samples through a pair of Sony dynamic stereo headphones. The volume of the excerpts varied between 55 and 80 dB. Listeners completed a demographic questionnaire, which included items about age, sex, musical training, and familiarity with Western, Japanese, and Hindustani music.

Results

Analyses of the questionnaire data revealed that less than 6% of participants were familiar with Hindustani ragas, 41% were familiar with traditional Japanese music, and 36% were familiar with Western improvisation. We also asked participants about their musical preferences: 36.98% indicated a preference for pop music, 21.23% indicated a preference for classical music, 18.49 indicated a preference for Japanese pop music, and 23.3% preferred rhythm and blues, ska, hip hop, or dance music. Less than 1% indicated a preference for Japanese traditional music.

We next analyzed ratings of anger, joy and sadness. For each rating scale, we conducted a 3×3 mixed-design analysis of variance (ANOVA) with music set (Japanese, Western, Hindustani) as the between-subjects factor and intended emotion (anger, joy, sadness) as the within-subjects factor. Our primary prediction for each rating scale was a significant main effect of intended emotion; ratings of anger should be higher for music intended to be angry; ratings of joy should be higher for music intended to be joyful; and ratings of sadness should be higher for music intended to be sad.

Mean ratings for each music set and intended emotion are displayed in Figure 2 (anger ratings), Figure 3 (joy ratings), and Figure 4 (sadness ratings). As seen in Figures 1–3, the highest overall ratings were always assigned to the intended emotion, and there was a main

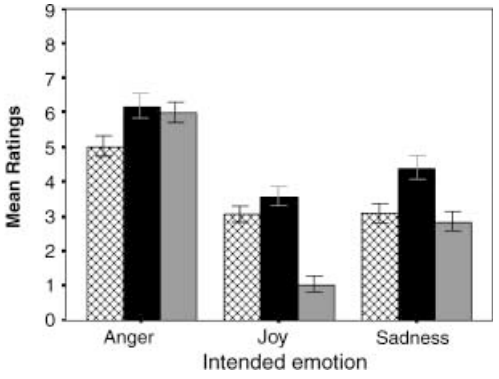

Figure 2. Mean ratings of anger for (▨) Hindustani, (■) Japanese and (▩) Western music intended to be angry, joyful or sad.

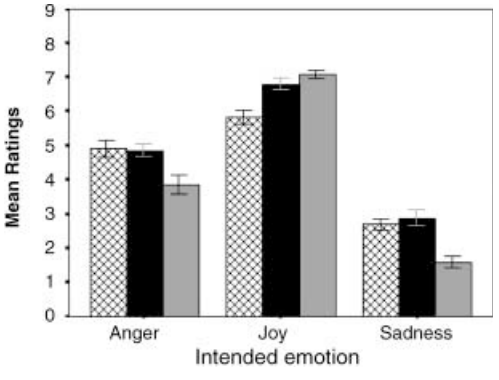

Figure 3. Mean ratings of joy for (▨) Hindustani, (■) Japanese and (▩) Western music intended to be angry, joyful or sad.

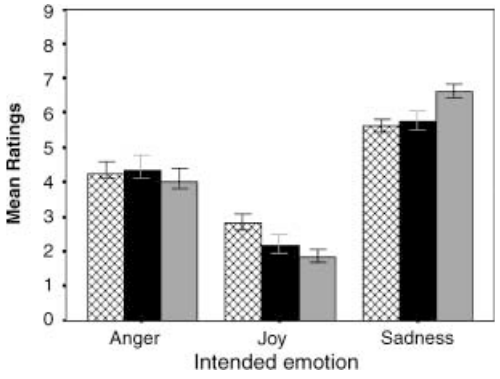

Figure 4. Mean ratings of sadness for (⊠) Hindustani, (■) Japanese and (□) Western music intended to be angry, joyful or sad.

effect of intended emotion for all three emotion ratings: anger, $F(2,92) = 182.511$, $p < 0.0001$, $eta^2 = 0.799$; joy, $F(2,96) = 389.916$, $p < 0.0001$, $eta^2 = 0.890$; and sadness, $F(2,88) = 214.655$, $p < 0.0001$, $eta^2 = 0.830$. These findings suggest that Japanese listeners were able to decode emotional meaning in familiar as well as unfamiliar music.

There were significant interactions between intended emotion and music set for ratings of anger, $F(4,92) = 16.611$, $p < 0.0001$, $eta^2 = 0.419$; joy, $F(4,96) = 17.404$, $p < 0.0001$, $eta^2 = 0.420$; and sadness, $F(4,88) = 6.392$, $p < 0.0001$, $eta^2 = 0.225$. These interactions indicate that sensitivity to emotional meaning in music depended on the culture with which the music was associated and suggest that sensitivity to emotional meaning is partially determined by experience. They also raise the possibility that listeners were unable to decode some emotions in one or more of the music sets.

To verify that listeners could reliably decode all intended emotions in all three music sets, we conducted a series of analyses for each music set separately (Japanese, Western, and Hindustani). For each rating scale and music set, we conducted a one-way ANOVA with repeated measures on intended emotion. As summarized in Table 2, there was a significant main effect of intended emotion for each rating scale and music set ($p < 0.01$ in all cases). In all nine analyses, ratings of the intended emotion were higher than ratings of unintended emotions. Thus, Japanese listeners could decode anger, joy, and sadness in Japanese, Western, and Hindustani music.

The ability to decode emotional meaning in Hindustani music is striking because this music was unfamiliar. According to the cue redundancy model, sensitivity to emotional meaning in unfamiliar music is possible because listeners can attend to acoustic cues that transcend cultural boundaries. Experiment 2 was conducted to examine this interpretation by asking listeners to rate the presence of acoustic cues in music.

Experiment 2: Sensitivity to acoustic cues within and across cultures

In Experiment 2, Japanese listeners judged tempo, loudness, and complexity in the same sets of Japanese, Western, and Hindustani

Table 2. Analyses of ratings of emotion in Japanese, Western and Hindustani music

	Japanese		Western		Hindustani	
Source	*DF*	*F*	*DF*	*F*	*DF*	*F*
Anger	2	32.328**	2	169.393**	2	32.643**
	22	(0.746)	34	(0.909)	36	(0.645)
Joy	2	226.022**	2	174.794**	2	124.245**
	22	(0.954)	38	(0.902)	36	(0.873)
Sadness	2	45.488**	2	130.353**	2	56.497**
	22	(0.805)	34	(0.885)	32	(0.779)

** $p < 0.01$. Values enclosed in parentheses represent eta^2.

music used in Experiment 1. We expected that excerpts intended to be angry, joyful, and sad would be differentiated by these acoustic cues. If so, then such cues could be used by listeners to decode emotional meaning. Other qualities such as modality, contour, consonance, and pitch register might also affect judgments of emotion in music, but we focused on tempo, loudness, and complexity because these cues are thought to be powerful signals of emotional meaning (Balkwill & Thompson, 1999; Juslin & Sloboda, 2001).

Method

Listeners. The same listeners who participated in Experiment 1 participated in Experiment 2; however, our testing procedures ensured that no listener was presented the same music in the two experiments. That is, for each music set, ratings of joy, sadness, anger, loudness, tempo, and complexity were always provided by different listeners, avoiding carry-over effects.

Thirty-nine listeners rated Japanese music (20 women and 19 men, mean age = 23.62 years): 12 provided complexity ratings (six women, six men), 12 provided loudness ratings (six women, six men), and 15 provided tempo ratings (eight women, seven men). Fifty-four listeners rated Western music (27 women and 27 men, mean age = 22.92 years): 18 provided complexity ratings (nine women, nine men), 18 provided loudness ratings (10 women, eight men), and 18 provided tempo ratings (eight women, 10 men). Fifty-six listeners rated Hindustani music (29 women and 27 men, mean age = 24.28 years): 17 provided complexity ratings (nine women, eight men), 18 provided loudness ratings (nine women, nine men), and 21 provided tempo ratings (11 women, 10 men).

Materials

The music materials were the same as those used in Experiment 1.

Procedure

Complexity, loudness, and tempo ratings were obtained for each music excerpt. As in Experiment 1, each listener was presented only one music set and rated only one acoustic cue. For example, a listener might be asked to rate tempo for the Hindustani music set. Ratings were assigned on a scale from 0 to 9, on which a rating near zero indicated a lower level of the acoustic cue in the music and a rating near 9 indicated a higher level of the acoustic cue. Other aspects of the procedure were identical to those used in Experiment 1.

Results

For each rating scale, we conducted a 3×3 mixed-design ANOVA with music set (Japanese, Western, Hindustani) as the between-subjects factor and intended emotion (anger, joy, sadness) as the within-subjects factor. Our primary prediction for each rating scale was a significant main effect of intended emotion. That is, ratings of complexity, loudness, and tempo should be significantly different for music intended to be angry, joyful, or sad.

Mean ratings for each intended emotion and music set are displayed in Figure 5 (complexity ratings), Figure 6 (loudness ratings), and Figure 7 (tempo ratings). Figures 5–7 illustrate that the mean ratings for all three acoustic cues varied depending on the intended emotion. Listeners assigned relatively low ratings of complexity, loudness, and tempo to music intended to be sad, a pattern that occurred for all three music sets. Listeners also assigned

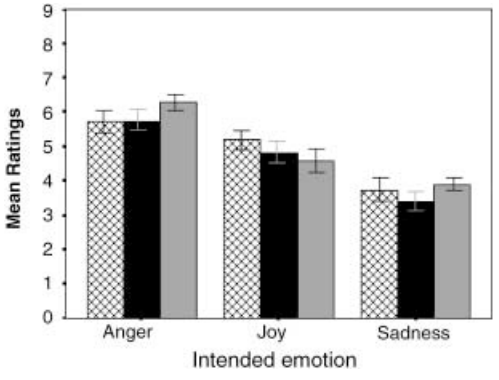

Figure 5. Mean ratings of complexity for (▨) Hindustani, (■) Japanese and (▩) Western music intended to be angry, joyful or sad.

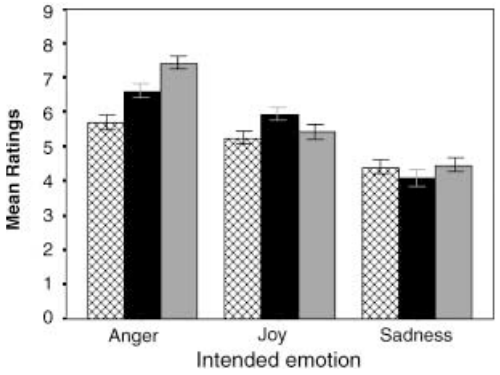

Figure 6. Mean ratings of loudness for (▨) Hindustani, (■) Japanese and (□) Western music intended to be angry, joyful or sad.

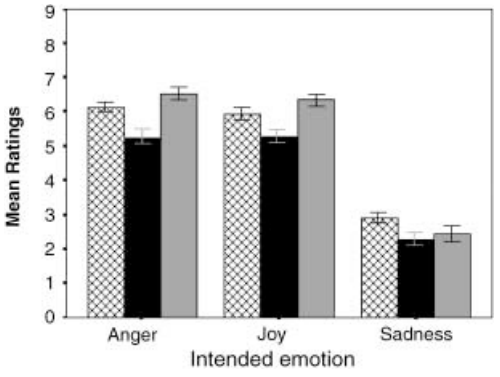

Figure 7. Mean ratings of tempo for (▨) Hindustani, (■) Japanese and (□) Western music intended to be angry, joyful or sad.

relatively high ratings of complexity, loudness, and tempo to music intended to be angry, a pattern that also occurred for all three music sets.

The reliability of these patterns was confirmed by a main effect of intended emotion for all three acoustic cues: complexity, $F(2,90) = 60.651$, $p < 0.0001$, $eta^2 = 0.580$, loudness, $F(2,90) = 216.666$, $p < 0.0001$, $eta^2 = 0.828$, and tempo, $F(2,100) = 523.374$, $p < 0.0001$, $eta^2 = 0.913$. The effects indicate that Japanese listeners are sensitive to acoustic cues in both familiar and unfamiliar music, and that their perceptions of these acoustic cues vary as a function of the intended emotion. In other words, listeners should be able to differentiate joyful, sad, and angry music by attending to differences in acoustic cues.

There were significant interactions between intended emotion and music set for ratings of loudness, $F(4,90) = 17.024$, $p < 0.0001$, $eta^2 = 0.431$, and tempo, $F(4,100) = 4.579$, $p < 0.002$, $eta^2 = 0.155$. These interactions indicate that sensitivity to loudness and tempo in music depended on the culture with which the music was associated, suggesting that perceptions of acoustic properties depend on familiarity with the musical style. They also raise the possibility that listeners were insensitive to some acoustic cues in one or more of the music sets.

To confirm that listeners were sensitive to acoustic cues in all three music sets, we conducted a one-way ANOVA with repeated measures on intended emotion for each music set and rating scale separately. As shown in Table 3, there was a significant main effect of intended emotion for all nine analyses ($p < 0.01$ in all cases), confirming that ratings of complexity, loudness and tempo always

Table 3. Analyses of ratings of acoustic cues in Japanese, Western and Hindustani music

	Japanese		Western		Hindustani	
Source	*DF*	*F*	*DF*	*F*	*DF*	*F*
Complexity	2	22.350**	2	24.981**	2	19.602**
	22	(0.670)	34	(0.595)	36	(0.551)
Loudness	2	79.602**	2	112.861**	2	40.856**
	22	(0.879)	38	(0.869)	36	(0.706)
Tempo	2	184.150**	2	162.917**	2	224.819**
	22	(0.929)	34	(0.906)	32	(0.922)

** $p < 0.01$. Values enclosed in parentheses represent eta^2.

Table 4. Multiple regression summary

Music set	Emotion	Adjusted R^2	Significant predictors
Japanese	Anger	0.718	⇑ Loudness, ⇑ Complexity, ⇓ Tempo
	Joy	0.662	⇑ Tempo, ⇓ Complexity
	Sadness	0.690	⇓ Tempo, ⇑ Complexity
Western	Anger	0.799	⇑ Loudness, ⇑ Complexity, ⇑ Tempo
	Joy	0.794	⇑ Tempo, ⇓ Complexity
	Sadness	0.875	⇓ Tempo, ⇑ Complexity, Flute timbre
Hindustani	Anger	0.525	⇑ Loudness, ⇑ Complexity, ⇓ Tempo
	Joy	0.903	⇑ Tempo, ⇓ Complexity, ⇓ Loudness
	Sadness	0.788	⇓ Tempo, ⇑ Complexity, ⇑ Loudness
Average		0.750	

differed as a function of the intended emotional meaning.

Predicting emotion ratings from ratings of acoustic cues. We next assessed whether emotion ratings could be predicted using ratings of acoustic cues, along with a coded variable for timbre (strings or flute). For each music set, we conducted separate multiple regression analyses for ratings of anger, joy, and sadness (nine analyses). For each emotion rating and music set, the criterion variable consisted of the set of mean emotion ratings assigned to each of the 30 musical excerpts (10 excerpts for each of three intended emotions). The predictors consisted of mean complexity ratings, loudness ratings, and tempo ratings, and a coded variable indicating whether the timbre was created using a flute or a stringed instrument.

Table 4 contains a summary of significant predictors in each analysis (stepwise regression). Details of individual analyses are available from the first author. As seen in Table 4, each analysis resulted in at least two significant predictor variables, and there was remarkable consistency in the associations between ratings of emotions and ratings of acoustic cues. On average, acoustic predictors accounted for 75% of variance in mean ratings of emotions. The following associations were observed in all three music sets: anger ratings were positively associated with ratings of complexity and loudness; joy ratings were positively associated with ratings of tempo and negatively associated with ratings of complexity; and sadness ratings were positively associated with ratings of complexity and negatively associated with ratings of tempo.

It should be emphasized that using mean ratings as the dependent variable removed variability in ratings caused by individual differences. Thus, the explanatory power of acoustic predictors is likely to be lower for many individual listeners. Nonetheless, the results are consistent with the view that listeners attend to acoustic cues when judging emotional meaning in both familiar and unfamiliar music, and the associations observed in our analyses are consistent with associations previously reported in the literature (Gabrielsson & Lindström, 2001).

Discussion

The results of this investigation illustrate that Japanese listeners are sensitive to the musical expression of anger, joy and sadness in Western, Japanese and Hindustani music. Moreover, judgments of emotion in the music of all three cultures were associated with subjective judgments of acoustic qualities of complexity, loudness, tempo, and timbre. Across conditions, mean ratings of acoustic cues accounted for approximately three-quarters of the variance in mean emotion ratings. These findings suggest that emotional meaning in music is conveyed by acoustic cues, and interpretation

of these cues does not require familiarity with culture-specific musical conventions.

Associations between ratings of acoustic cues and ratings of emotions were remarkably consistent across the three music sets. First, increases in perceived complexity were consistently associated with the negative emotions of anger and sadness, whereas decreases in perceived complexity were associated with joy. Second, increases in perceived tempo were always associated with joy, whereas decreases in tempo were always associated with sadness. Third, increases in loudness were associated with anger in all three music sets. Of course, it is impossible to verify that listeners used these cues as the basis for making judgments of emotion, but our analyses illustrate that attending to acoustic cues would be an effective strategy for decoding emotional meaning.

A possible interpretation of these results is that the remaining variance in the equations stems from culture-specific cues that were less influential for Japanese listeners than for other listeners. However, interpretation of these results is premature because we did not evaluate listeners' judgments of all possible acoustic cues involved in the expression of emotion in music.

Although an examination of the associations between subjective and objective measures of acoustic cues is beyond the scope of this investigation, our findings are consistent with a wide range of studies in which objective measures of acoustic cues were related to listeners' judgments of emotion in music from their own culture (for a review, see Gabrielsson & Lindström, 2001). Gabrielsson and Juslin (1996) found that fast tempi were associated with listeners' perceptions of joyfulness and anger, and slow tempi were associated with listeners' perceptions of sadness. Hevner (1935, 1936) found that listeners associated simple harmony with happiness and serenity, and complex harmony with sadness. Rigg (1937) found that listeners linked perceptions of joy with faster tempi and simple harmony, and perceptions of "lamentation" with slow tempi. Finally, Imberty (1976) found that high complexity was related to responses of "a melancholic or sorrow kind," and low complexity was related to responses of "joy and well-being". Our findings extend these studies by illustrating that such associations are significant across cultures.

According to the cue redundancy model of Balkwill and Thompson (1999), when musicians express emotion through music, they make use of conventional cues as well as acoustic cues of emotional meaning. Acoustic cues are particularly powerful signals because their interpretation requires no knowledge of musical conventions. Indeed, acoustic cues such as complexity, loudness, tempo, and timbre have emotional significance in other acoustic signals such as speech and animal calls. Vocal expressions of anger are often marked by fast tempo and greater loudness; vocal expressions of joy are often fast in tempo and have an increased pitch range, and vocal expressions of sadness are usually slow in tempo and low in pitch register (e.g., Bachorowski & Owren, 1995; Frick, 1985; Juslin & Laukka, 2000; Scherer, 1986; Scherer & Oshinsky, 1977). When speech prosody and music evoke the same emotion they share many of the same acoustic cues. These commonalities suggest that emotions are communicated through acoustic cues that not only transcend cultural boundaries, but are also manifested in different types of acoustic signals.

Acoustic cues did not always signify emotional meaning in the same way across the three music sets, suggesting a complex mapping between acoustic cues and emotional meaning. For example, whereas anger in Western music was associated with increases in tempo, anger in Japanese and Hindustani music was associated with decreases in tempo. More generally, acoustic cues often provided ambiguous information when considered on their own. For example, increases in perceived complexity were associated with increases in perceived anger as well as increases in perceived sadness.

Such sources of ambiguity suggest that emotional cues must be interpreted probabilistically (Juslin, 2001). The process of interpreting emotional meaning is probabilistic because it involves the integration of multiple cues, none

of which on their own conveys the emotion unambiguously. Each cue is associated with the expression of different emotions to varying degrees. Thus, listeners must consider combinations of cues in order to evaluate the intended emotion. Different cues may have varying influence (weighting) on the listeners' evaluation, depending on their capacity to signify a specific emotion, attentional factors, and the extent to which the performer manipulated each acoustic cue.

To conclude, we observed support for two predictions based on the cue redundancy model: that listeners can interpret the intended emotion expressed in music of an unfamiliar culture, and that these judgments are associated with emotion-specific acoustic cues. There are also complex influences of enculturation and other potential acoustic influences that cannot be revealed by the current data. Identifying and understanding such influences is a critical next step in achieving a full understanding of the perception of emotion in music across cultures. Additional investigations are therefore needed to evaluate the influence of culture-specific cues and other acoustic cues on the ability to decode emotion in music.

References

Bachorowski, J., & Owren, M. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, **6**, 219–224.

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.

Behrens, G. A., & Green, S. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music*, **21**, 20–33.

Boersma, P., & Weenink, D. (1999). Praat, a system for doing phonetics by computer. *Glot International*, **5**, 341–345.

Daniélou, A. (1980). *The ragas of Northern Indian music*. London: Barrie & Rockliff.

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

Dolgin, K. G., & Adelson, E. H. (1990). Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music*, **18**, 87–98.

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, **97**, 412–429.

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

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotion expression. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 223–248). New York: Oxford University Press.

Gerardi, G. M., & Gerken, L. (1995). The development of affective responses to modality and melodic contour. *Music Perception*, **12**, 279–290.

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

Hevner, K. (1935). Expression in music: a discussion of experimental studies and theories. *Psychological Review*, **42**, 186–204.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology*, **48**, 248–268.

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

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, **20**, 151–172.

Imberty, M. (1976). *Signification and meaning in music (on Debussy's preludes pour le piano)*. Montreal: Groupe de recherches en sémiologie musicale, Faculté de Musique, Université de Montréal.

Juslin, P. N. (2001). Communicating emotion in music performance. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309–337). New York: Oxford University Press.

Juslin, P. N., & Laukka, P. (2000). Improving emotional communication in music performance through cognitive feedback. *Musicae Scientia*, **4**, 151–183.

Juslin, P. N., & Sloboda, J. A. (2001). Music and emotion: Introduction. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 3–20). New York: Oxford University Press.

Kratus, J. (1993). A developmental study of children's interpretation of emotion in music. *Psychology of Music*, **21**, 3–19.

Kruglanski, A. W. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*, **106**, 395–409.

Oku, S. (1995). *Music education in Japan*. Nara, Japan: Kansai Insatsu Ltd.

Rigg, M. G. (1937). An experiment to determine how accurately college students can interpret the meanings of musical compositions. *Journal of Experimental Psychology*, **21**, 223–229.

Rigg, M. G. (1940). Speed as a determiner of musical mood. *Journal of Experimental Psychology*, **27**, 566–571.

Robazza, C., Macaluso, C., & D'Urso, V. (1994). Emotional reactions to music by gender, age, and expertise. *Perceptual and Motor Skills*, **79**, 939–944.

Roberts, R. D., Zeidner, M., & Matthews, G. (2001). Does emotional intelligence meet traditional standards for an intelligence? Some new data and conclusions. *Emotion*, **1**, 196–231.

Scherer, K. R. (1986). Vocal affect expression: a review and a model for future research. *Psychological Bulletin*, **99**, 143–165.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, **1**, 331–346.

Terwogt, M., & Van Grinsven, F. (1991). Musical expression of moodstates. *Psychology of Music*, **19**, 99–109.

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

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Mood, arousal, and the Mozart effect. *Psychological Science*, **12**, 248–251.

(Received March 28, 2003; accepted May 22, 2004)