A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect

Gabriela Ilie and William Forde Thompson
University of Toronto

Using a three-dimensional model of affect, we compared the affective consequences of manipulating intensity, rate, and pitch height in music and speech. Participants rated 64 music and 64 speech excerpts on valence (pleasant-unpleasant), energy arousal (awake-tired), and tension arousal (tense-relaxed). For music and speech, loud excerpts were judged as more pleasant, energetic, and tense than soft excerpts. Manipulations of rate had overlapping effects on music and speech. Fast music and speech were judged as having greater energy than slow music and speech. However, whereas fast speech was judged as less pleasant than slow speech, fast music was judged as having greater tension than slow music. Pitch height had opposite consequences for music and speech, with high-pitched speech but low-pitched music associated with higher ratings of valence (more pleasant). Interactive effects on judgments were also observed. We discuss similarities and differences between vocal and musical communication of affect, and the need to distinguish between two types of arousal: energy and tension.

Received August 24, 2004, accepted June 5, 2005

In the past several decades there has been considerable research on the perception of emotional meaning in two nonverbal domains: music and vocal expression (i.e., prosody). Music and prosody are acoustic stimuli that convey emotional meaning through variation in pitch height, pitch contour, intensity, rate (tempo), timbre, and rhythmic grouping. Music and speech prosody are involved in affiliative interactions between mothers and infants (Dissanayake, 2000), and sensitivity to music and speech prosody may have similar developmental underpinnings (McMullen & Saffran, 2004). Training in music leads to enhanced sensitivity to emotional meaning conveyed by speech prosody (Thompson, Schellenberg & Husain, 2004). Such associations suggest that musical behavior and vocal communication share common ancestry (Brown, 2000; Dissanayake, 2000; Joseph, 1988; Pinker, 1995) and are associated with overlapping neural resources (Deutsch, Henthorn, & Dolson, 2004; Patel, Peretz, Tramo, & Labrecque, 1998).

There is widespread evidence supporting a strong link between music and emotion (for a review, see Juslin & Sloboda, 2001). Hevner (1935a, 1935b, 1936, 1937) reported that listeners associate qualities of music such as tempo and pitch height with specific emotions. Listeners can decode the emotional meaning of music by attending to these qualities, even when they are listening to unfamiliar music from another culture (Balkwill & Thompson, 1999). Sensitivity to emotional meaning in music emerges early in development and improves with age (Cunningham & Sterling, 1988; Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Kratus, 1993; Terwogt & van Grinsven, 1988, 1991). Listeners not only perceive emotional meaning in music but also may experience physical sensations in response to the music, such as tears, tingles down the spine ("chills"), or changes in heart rate, blood pressure, and skin conductance levels (Goldstein, 1980; Krumhansl, 1997; Panksepp, 1995; Sloboda, 1991, 1992; Thayer & Levenson, 1983). Indeed, music is often "used" as a stimulus for changing one's mood (Sloboda, 1992), and musical properties such as tempo and mode affect listeners' self-reported mood and arousal levels (Husain, Thompson, & Schellenberg, 2002).

Emotions are also conveyed by vocal qualities of speech, or *speech prosody* (Frick, 1985; Juslin & Laukka, 2001). Prosody refers to the use of intonation (pitch variation) and rhythm (stress and timing) in speech, as distinct from verbal content. Speakers from different cultures use similar prosodic cues to convey emotions (Bolinger, 1978). By attending to such cues, speakers can decode emotional meaning in speech from an unfamiliar language (Thompson & Balkwill, 2006). Prosodic cues are often exaggerated in speech directed toward infants and young children and are associated with higher overall pitch levels, larger pitch excursions, slower rate, and longer pauses (Ferguson, 1964; Fernald & Mazzie, 1991).

Music Perception volume 23, issue 4, pp. 319-329, issn 0730-7829, electronic issn 1533-8312

Research on the communication of emotion in the auditory domain has included many studies of music and speech, but rarely have these domains been compared. Juslin and Laukka (2003) performed a meta-analysis of 104 studies of vocal expression and 41 studies of music. The analysis confirmed that decoding accuracy for broad emotion categories is above chance performance levels in both domains, with some emotions (e.g., sadness) more easily communicated than other emotions (e.g., fear). Both domains involve similar emotion-specific acoustic cues, such as overall F0/pitch level, rate (tempo), and intensity. For example, in both music and speech prosody, happiness is associated with a fast pace and high intensity, whereas sadness is associated with a slow pace and low intensity (Juslin & Sloboda, 2001; Scherer, 1986). Thus, there are similarities between music and speech in the patterns of acoustic cues used to communicate emotions. In view of these similarities, Juslin and Laukka (2003) proposed that expressions of emotion are processed by general-purpose brain mechanisms that respond to acoustic features regardless of whether the stimulus input is music or speech.

An alternative view is that music is processed by a modularized system that operates independently of processes involved in speech perception (Ayotte, Peretz, & Hyde, 2002; Peretz & Coltheart, 2003). Reports of selective impairments in music recognition abilities suggest the existence of a music-processing module. For example, some brain-damaged patients with "acquired amusia" are unable to recognize melodies that were familiar to them prior to injury, but have normal ability to recognize familiar song lyrics, voices, and environmental sounds (animal cries, traffic noise). It is important, however, not to draw a false opposition between claims of shared neural resources for decoding emotional meaning from acoustic information in music and speech, and the above evidence for modularity of melodic recognition. More generally, music processing is not a single "module" but is composed of smaller processing subsystems, some that may be specific to music and others that may handle auditory input regardless of whether that input arises from music, vocal utterances, or environmental sounds.

The analysis of Juslin and Laukka illustrates that music and speech convey emotions in similar ways, suggesting the existence of a domain-general mechanism for decoding emotional meaning from acoustic information. However, a full understanding of the degree of overlap between affective qualities in music and speech requires a direct comparison of these domains. To this end, we compared the effects of manipulating three affective cues in music and speech: intensity, pitch height, and rate (tempo). Our manipulations always resulted in speech or music stimuli that could occur naturally. For each quality, manipulations of music excerpts were identical to manipulations of speech excerpts, allowing us to determine whether a particular shift in intensity, pitch level, or rate has the same affective consequence in the two domains. We also examined potential interactions between these qualities to address the general lack of data on such interactions. The affective qualities attributed to music or speech may depend on interactions between qualities such as rate, intensity, and pitch height.

We assessed the affective consequences of our stimulus manipulations on three affective dimensions: valence (pleasant-unpleasant), energy arousal (awake-tired), and tension arousal (tense-relaxed). The dependent measures were based on the three-dimensional model of affect described by Schimmack and Grob (2000)—a model that integrates Thayer's (1978a, 1978b, 1986) multidimensional model of activation with Russell's (1980) circumplex model of valence and arousal. In Thayer's multidimensional model, arousal is a construct derived from two factors: energetic arousal (vigorous, lively, full of pep, active, happy, and activated) and tension arousal (anxious, jittery, clutched up, fearful, intense, and stirred up). In the circumplex model, emotions are described as points on a two-dimensional space composed of arousal (low or high) and valence (unpleasantness or pleasantness). For example, happiness is described as a feeling of excitement (high arousal) combined with positive affect (high valence).

With only one dimension of arousal, however, the circumplex model has difficulty differentiating between emotions such as sadness and fear, which are sometimes comparable in valence and energy but usually differ in the degree of tension experienced. Similarly, the emotions of sadness, fear, and anger cannot be differentiated with only the dimensions of arousal and valence because all three emotions have a negative valence. Indeed, although there have been extensive discussions of arousal and related constructs in music research and theory (Lerdahl & Jackendoff, 1983; Krumhansl, 1996, 1997; Meyer, 1956; Schubert, 2001; Thompson, Schellenberg, & Husain, 2001), this work has not clearly differentiated energy and tension as relatively independent aspects of arousal that are experienced differently and associated with distinct neural systems (Gold, MacLeod, Deary, & Frier, 1995; Schimmack, 1999; Schimmack & Grob, 2000; Schimmack & Reisenzein, 2002; Thayer, 1989; Tucker & Williamson, 1984).

Based on the meta-analysis reported by Juslin and Laukka (2003), we predicted that our manipulations

would influence judgments of all three dimensions of affect and that similar effects would be observed for music and speech. Because no previous studies have directly compared the effects of such manipulations for music and speech, we were unsure of whether identical effects would be observed in the two domains.

Method

Participants

Twenty-seven undergraduate students at the University of Toronto ranging in age from 18 to 27 years participated in the study (20 females and 7 males). Students were recruited from introductory psychology classes and received partial course credit for their participation. They had an average of 3.4 years of formal music lessons (*SD* = 3.94 years; range: 0 to 20 years).

Materials and Pilot Tests

The stimuli consisted of eight excerpts of music and eight excerpts of speech. Music stimuli consisted of complete musical phrases excerpted from eight pieces of classical music composed in the key of D major and performed by a symphony orchestra (see Table 1). The eight pieces were copied from CD recordings and edited using Sound Edit software. Phrases were excerpted from the middle sections rather than the beginning of the pieces. All music excerpts were between 5 and 6 s in length (mean = 5.70 s). The mean amplitude of music stimuli was 74 dB based on audiometric measurements taken at the headphones, and the average pitch across music samples was 175 Hz (*SD* = 83 Hz). The tempi of music samples ranged from 110 bpm to 120 bpm (mean = 117 bpm). Using ProTools (version 5.0.1), we created eight versions of each music excerpt by manipulating intensity (loud = 80% normalization, mean dB value for loud excerpts: 84.42 linear dB SPL; soft = 5% normalization, mean dB value for soft excerpts: 61.12 linear dB SPL), tempo (fast and slow versions were 1.12 and 0.89 times the rate of the original samples, respectively, or 131 bpm and 104 bpm), and pitch height (high = two semitones up from the original recordings, 191.28 Hz mean pitch for the high music samples; low = two semitones down from the original recordings, 156.77 Hz mean pitch for the low music samples). The manipulations yielded 64 presentations (2 pitches × 2 tempi × 2 intensities × 8 musical excerpts).

TABLE 1. Musical pieces used to create the music presentations.

Composer	Musical piece
Vivaldi	Allegro Vivace in D Major, from RV 109
Stradella	Sonata in D Major
Mozart	Serenade in D Major K 320, Andantino
Haydn	Sonata nr. 9 in D, Menuetto
Handel	Water Music Suite 2 in D; Menuet
Handel	Water Music Suite 2 in D; Allegro
Handel	Water Music Suite 2 in D; Menuet Alla Hornpipe
Alberti	Sonata in D Major

Speech stimuli consisted of a text spoken by each of eight students (4 females, 4 males) majoring in Theatre and Drama at the University of Toronto. The actors read sections from a descriptive text about sea turtles and were recorded using a Tascam 244 mixer, a compressor limiter dB × 163 and a unidirectional dynamic microphone ATM 63 (25 ohms). Phrases of 5 to 7 s in length were extracted from the recordings using SoundEdit (mean = 6.23 s). The mean amplitude of speech stimuli was 70 dB based on audiometric measurements taken at the headphones, and the average pitch of speech samples was 125 Hz for male speakers (*SD* = 34 Hz) and 202 Hz for female speakers (*SD* = 46 Hz). The average speaking rate was 149 words per minute (about 2.5 words per second, 3.41 syllables per second). ProTools was used to create eight versions of each spoken phrase by manipulating the intensity (loud = 80% normalization, mean dB value for loud excerpts: 80.13 linear dB SPL; soft = 5% normalization, mean dB value for soft excerpts: 56.03 dB linear SPL), speaking rate (fast and slow versions were 1.12 and 0.89 times the rate of the original samples, respectively, or 167 and 133 words per minute), and pitch height (high = two semitones up from the original recordings, 180.79 Hz mean pitch for the high speech samples; low = two semitones down from the original recordings, 144.37 Hz mean pitch for the low speech samples). The manipulations yielded 64 speech presentations.

Because uniform manipulations of intensity, rate, and pitch may not occur in normal music and speech, we conducted pilot testing to ensure that music and speech stimuli sounded natural. In one pilot study, eight listeners heard speech samples at various intensity levels and judged their intelligibility as well as their degree of comfort. Intensity manipulations were selected based on these judgments with the lower limit corresponding to the lowest intensity value that was still judged to be intelligible by all listeners, and an upper limit that corresponded to the highest intensity that was still judged to be comfortable by all listeners.

In another pilot study, the same listeners judged the naturalness of speech stimuli after transposing them two and three semitones up or down in pitch. All participants judged all of the two-semitone shifted samples

as natural sounding and within the expected range for speech. Most manipulations of three semitones were also judged as sounding natural, but there were a few exceptions. From these data we decided to use pitch manipulations of two semitones for our investigation.

Rate (tempo) manipulations were based on pitch manipulations. Although ProTools allowed us to manipulate pitch height and rate independently, the chosen tempi were based on presentation rates that would result in pitch shifts of two semitones higher or lower than the original recordings if accomplished by analogue manipulation (as in speeding up or slowing down a tape recorder). Selecting these values meant that pitch height and rate manipulations were of comparable magnitude. As mentioned, manipulations of music stimuli were identical to manipulations of speech stimuli, and all such manipulations resulted in intensity, pitch, and tempo values that occur often in music. In short, all of our manipulations resulted in stimuli that sounded natural, were intelligible, and were presented at comfortable listening levels.

Dependent Measures

Ratings were obtained for each pole of the three dimensions of affect: valence (pleasant and unpleasant), energy arousal (energetic and boring), and tension arousal (tense and calm). To allow for the possibility that opposing qualities for each dimension of affect are somewhat independent of each other (e.g., a stimulus might be rated as both very energetic and very boring), the response format was a unipolar intensity scale (0 = not at all, to 4 = extremely). For example, a rating of 0 on the pleasant scale indicated that the excerpt did not sound pleasant, and a rating of 4 indicated that the excerpt sounded very pleasant. Preliminary analyses suggested, however, that the six rating scales could be reduced to three bipolar scales. Although use of unipolar scales is often prudent, the three-dimensional model of affect assumes approximate bipolarity of valence, energy arousal, and tension arousal (Schimmack & Grob, 2000; Schimmack & Reisenzein, 2002). To obtain bipolar indicators of each dimension, ratings on the lower-pole items were subtracted from ratings on the higher-pole items and then transposed to a scale from 1 to 9, with ratings of 5 indicating equivalent ratings on the lower and higher poles, and ratings above or below 5 indicating greater weighting on the high or low poles, respectively.

Procedure

Participants were seated in a sound-attenuated booth and given a short demonstration of the rating task. They were instructed to use the full range of the rating scale. Six ratings were made after each of the 128 presentations. The order in which music and speech stimuli were presented was fully randomized. Presentation was controlled by the software package *Experiment Creator*, which was designed by the second author and is freely available at http://ccit.utm.utoronto.ca/billt. Listeners were tested individually in a sound-attenuated booth and used a mouse to initiate presentations and input their ratings. All excerpts were presented through Sennheiser HD headphones. After completion of the ratings for all musical and speech excerpts, participants completed a demographics questionnaire and a few questions about the experiment. They were then debriefed.

Results

We conducted an ANOVA with repeated measures on domain (music or speech), intensity (loud or soft), rate (fast or slow), and pitch height (high or low). ANOVA was conducted for each affective rating: valence, energy arousal, and tension arousal. For valence ratings there was a main effect of domain, $F(1, 26) = 60.03$, $p < .001$, with higher ratings of valence associated with music ($M = 6.69$, $SE = 0.22$) than with speech ($M = 4.63$, $SE = 0.11$). There was also a main effect of domain for energy arousal, $F(1, 26) = 60.47$, $p < .001$, with higher ratings of energy associated with music ($M = 5.76$, $SE = 0.24$) than with speech ($M = 3.69$, $SE = 0.20$). For tension arousal, mean ratings of music ($M = 4.36$, $SE = 0.15$) were not reliably different than mean ratings of speech ($M = 4.44$, $SE = 0.13$).

We observed a number of significant interactions with domain, which motivated separate analyses for music and speech stimuli. For valence ratings, there were significant interactions between domain and pitch height, $F(1, 26) = 11.22$, $p < .01$, domain and rate, $F(1, 26) = 14.16$, $p < .01$, and domain, intensity, pitch height, and rate, $F(1, 26) = 5.08$, $p < .05$. For energy arousal ratings, there were significant interactions between domain and intensity, $F(1, 26) = 11.97$, $p < .01$, and domain and rate, $F(1, 26) = 12.59$, $p < .01$. For tension arousal ratings there were significant interactions between domain and intensity, $F(1, 26) = 22.25$, $p < .001$, and domain and rate, $F(1, 26) = 10.04$, $p < .01$. These interactions were explored by conducting repeated-measures ANOVAs for music and speech stimuli separately, with intensity, rate, and pitch height as independent variables in each analysis. For each domain, a separate ANOVA was performed for ratings of valence, energy arousal, and tension arousal. We also examined partial

eta-squared (the proportion of the effect + error variance that is attributable to the effect) for each significant effect. Table 2 displays mean ratings of music and speech for all combinations of the three dimensions of affect. Table 3 summarizes the significant main effects and indicates the direction of these effects.

Valence Ratings

Music stimuli. There were main effects of intensity, $F(1, 26) = 9.01$, $p < .05$, partial eta-squared = .26, and pitch height, $F(1, 26) = 9.04$, $p < .01$, partial eta-squared = .26. Across conditions, soft excerpts were judged as more pleasant than loud excerpts. Rate manipulations did not significantly influence valence ratings for music stimuli. A significant interaction between pitch height and intensity suggested that the effects of intensity were greater when pitch height was high than when it was low, $F(1, 26) = 10.19$, $p < .01$, partial eta-squared = .28, with the lowest ratings of valence assigned to loud high-pitched music. There was also a significant interaction between pitch height and rate, $F(1, 26) = 4.83$, $p < .05$, partial eta-squared = .16. When the tempo was fast, music excerpts were judged as more pleasant when the pitch height was low than when it was high. When the tempo was slow, there was no significant effect of pitch height on ratings of valence.

Speech stimuli. There were reliable main effects of intensity, $F(1, 26) = 5.26$, $p < .05$, partial eta-squared = .17, rate, $F(1, 26) = 15.54$, $p < .01$, partial eta-squared = .37, and pitch height, $F(1, 26) = 4.47$, $p < .05$, partial eta-squared = .15. Across conditions, soft excerpts were judged as more pleasant than loud excerpts, slow excerpts were judged as more pleasant than fast excerpts, and high-pitched excerpts were judged as more pleasant than low-pitched excerpts. There was a significant interaction between pitch height and intensity, $F(1, 26) = 11.72$, $p < .01$, partial eta-squared = .31. As with music stimuli, effects of intensity were greater when pitch height was high than when it was low. The interaction between pitch height and intensity suggests that soft high-pitched voices were perceived as more pleasant than any other pitch height and intensity combination. Unlike valence ratings for music, the interaction between pitch height and rate was not significant. Across conditions, there was a main effect of speaker sex, $F(1, 26) = 20.99$, $p < .01$, partial eta-squared = .45, with more positive valence attributed to female speakers ($M = 4.89$, $SE = 0.13$) than to male speakers ($M = 4.37$, $SE = 0.11$).

Energy Arousal

Music stimuli. There were main effects of intensity, $F(1, 26) = 75.79$, $p < .001$, partial eta-squared = .75, and rate, $F(1, 26) = 45.40$, $p < .001$, partial eta-squared = .15. Listeners perceived loud and fast music as more energetic than soft and slow music, respectively.

Speech stimuli. There were main effects of intensity, $F(1, 26) = 22.74$, $p < .001$, partial eta-squared = .47, pitch height, $F(1, 26) = 12.59$, $p < .01$, partial eta-squared = .33, and rate, $F(1, 26) = 46.23$, $p < .001$, partial eta-squared = .64. There was also a significant three-way interaction between rate, pitch height, and intensity, $F(1, 26) = 4.83$, $p < .05$, partial eta-squared = .16. The highest ratings of energy were associated with loud, fast, and high-pitched speech samples, and the lowest ratings of energy were associated with

TABLE 2. Bipolar ratings of music and speech for the three dimensions of affect.

		Valence				Energy arousal				Tension arousal			
		Music		Speech		Music		Speech		Music		Speech	
		Loud	Soft	Loud	Soft	Loud	Soft	Loud	Soft	Loud	Soft	Loud	Soft
High pitch	Fast	6.06 (0.30)	6.98 (0.23)	4.51 (0.16)	5.24 (0.19)	6.64 (0.28)	5.43 (0.27)	4.74 (0.26)	3.78 (0.23)	5.32 (0.24)	4.00 (0.20)	4.91 (0.20)	4.18 (0.16)
	Slow	6.33 (0.26)	7.01 (0.21)	4.22 (0.16)	4.90 (0.14)	6.04 (0.28)	5.16 (0.25)	3.73 (0.23)	3.12 (0.21)	5.02 (0.21)	3.48 (0.17)	4.65 (0.18)	4.14 (0.18)
Low pitch	Fast	6.71 (0.26)	7.02 (0.24)	4.66 (0.16)	4.94 (0.18)	6.52 (0.22)	5.42 (0.25)	4.33 (0.27)	3.73 (0.21)	5.22 (0.20)	3.73 (0.20)	4.75 (0.17)	4.13 (0.16)
	Slow	6.49 (0.27)	6.94 (0.27)	4.39 (0.18)	4.18 (0.18)	6.06 (0.26)	4.85 (0.27)	3.41 (0.26)	2.67 (0.21)	4.69 (0.21)	3.41 (0.17)	4.65 (0.19)	4.09 (0.22)

Note. Standard errors are shown in parentheses.

TABLE 3. Significant main effects of intensity, rate, pitch height on valence, energy arousal, and tension arousal for music and speech.

Main finding by emotion dimension					
Stimulus properties			Valence	Energy arousal	Tension arousal
Intensity	*Loud*	Music	−	+	+
	Soft		+	−	−
	Loud	Speech	−	+	+
	Soft		+	−	−
Rate	*Fast*	Music	/	+	+
	Slow		/	−	−
	Fast	Speech	−	+	/
	Slow		+	−	/
Pitch height	*High*	Music	−	/	+
	Low		+	/	−
	High	Speech	+	+	/
	Low		−	−	/

Note. "+" indicates significantly higher ratings on the affect dimension represented, "−" indicates lower ratings on the affect dimension represented, "/" indicates no significant difference.

soft, slow, and low-pitched speech samples. Across conditions, there was a main effect of speaker sex, $F(1, 26) = 11.72$, $p < .01$, partial eta-squared = .31, with greater energy attributed to female speakers ($M = 3.79$, $SE = 0.21$) than to male speakers ($M = 3.59$, $SE = 0.20$).

Tension Arousal

Music stimuli. There were main effects of pitch height, $F(1, 26) = 10.60$, $p < .01$, partial eta-squared = .29, rate, $F(1, 26) = 21.37$, $p < .001$, partial eta-squared = .45, and intensity, $F(1, 26) = 51.10$, $p < .001$, partial eta-squared = .66. Listeners judged loud, fast, and high-pitched music as more tense than soft, slow, and low-pitched music, respectively.

Speech stimuli. There was a main effect of intensity, $F(1, 26) = 15.66$, $p < .01$, partial eta-squared = .38, with significantly higher ratings of tension assigned to loud voices than to soft voices. Across conditions, there was a main effect of speaker sex, $F(1, 26) = 11.72$, $p < .01$, partial eta-squared = .311, with greater tension attributed to female speakers ($M = 4.28$, $SE = 0.14$) than to male speakers ($M = 4.60$, $SE = 0.14$).

Discussion

Manipulations of intensity, rate, and pitch height had affective consequences in music and speech, influencing judgments of valence, energetic arousal, and tension arousal. To our knowledge, this is the first study involving a direct comparison of the affective consequences of manipulating acoustic features in music and speech. The results support the view that the capacity to process certain features of speech also may subserve the perception of music (Deutsch, Henthorn, & Dolson, 2004; McMullen & Saffran, 2004; Patel & Daniele, 2003; Thompson, Schellenberg, & Husain, 2004) and that certain acoustic events may have an affective meaning that is appraised by circuitry that does not differentiate between the type of stimuli being processed. The well-established link between music and emotion may arise, in part, from a general mechanism that connects all acoustic attributes such as intensity, rate, and pitch height with affective connotations, regardless of whether those acoustic attributes are associated with music or speech.

Effects on energetic arousal and tension arousal were not identical, confirming the need to distinguish between these two types of arousal. The distinction has important implications for research and theories on music that attach significance to arousal and related constructs (Juslin & Sloboda, 2001, and references therein; Lerdahl & Jackendoff, 1983; Meyer, 1956; Thompson, Schellenberg, & Husain, 2001). Conceiving emotion in terms of a small number of core dimensions addresses a concern that research on music and emotion often fails to discriminate between the essential dimensions of affect (Scherer & Zentner, 2001, p. 382). The three-dimensional model of affect provides a parsimonious means of organizing emotional associations with music or speech without blurring important psychological distinctions (Sloboda & Juslin, 2001).

Across conditions, music stimuli were assigned higher ratings of valence and energetic arousal than speech stimuli, suggesting broad differences in the affective consequences of listening to these two types of stimuli. The finding is consistent with the observation that people often listen to music for pleasure and to modify energetic states (DeNora, 2001), whereas people rarely listen to speech for this purpose. Nonetheless, manipulations of acoustic attributes influenced affective appraisals of both music and speech, with all three dimensions of affect influenced by at least one acoustic attribute. The finding suggests that intensity, rate, and pitch height provide listeners with valuable perceptual information that allows them to decode emotional meaning in either music or speech prosody.

Table 3 illustrates that manipulating stimulus properties often had similar affective consequences in music and speech. Across other conditions, manipulations of

intensity had identical effect in music and speech for all three judgments: valence, energy arousal, and tension arousal. Soft speech and soft music were perceived as more pleasant, less energetic, and less tense than their loud counterparts, consistent with previous research on either music or speech (Gundlach, 1935; Juslin & Laukka, 2003; Watson, 1942; Wedin, 1972). This convergence suggests that intensity manipulations have domain-general connotations, possibly because intensity is a basic biological signal with deep evolutionary roots. As such, the appraisal of intensity may occur automatically and unconsciously at lower levels of the CNS (Scherer & Zentner, 2001).

Rate manipulations had the same effects on music and speech for judgments of energy arousal. Fast music and speech were judged to have higher levels of energy than slow music and speech, consistent with perceptual studies of music and speech (see Juslin & Laukka, 2003) and experiential studies of music (Husain, Thompson, & Schellenberg, 2002). The effects of rate manipulations on valence and tension differed for music and speech, however. Manipulations of rate influenced valence judgments for speech, but not music. Consistent with this finding, Husain et al. (2002) found that tempo manipulations in music had no effect on mood but greatly affected levels of energetic arousal. Manipulations of rate influenced tension judgments for music, but not speech.

Manipulations of pitch height had opposite effects on valence for music and speech. High-pitched speech was assigned higher ratings of valence (more positive) than low-pitched speech. In contrast, low-pitched music was assigned higher ratings of valence than high-pitched music. Although these findings may be surprising on first glance, they are compatible with previous results. High or rising vocal pitch is generally associated with politeness and deference, whereas low or falling vocal pitch is associated with authority, threat, and aggression (Bolinger, 1978; Morton, 1994; Ohala, 1984). In contrast, low-pitched music is associated with pleasantness whereas high-pitched music is associated with surprise, anger, and activity (Scherer & Oshinsky, 1977). Pitch height also influenced judgments of energy and tension, but in different ways for music and speech. High-pitched stimuli were judged to be significantly more energetic for spoken utterances and more tense for musical excerpts than low-pitched stimuli. That is, raising the pitch level of speech had the greatest impact on its perceived energy, whereas raising the pitch level of music had the greatest impact on its perceived tension.

All three acoustic attributes influenced perceived tension in music, whereas only intensity influenced perceived tension in speech. The finding suggests that music may have greater potential than speech to communicate tension in a variety of ways. It is possible that more extreme manipulations of rate and pitch height in speech would result in reliable effects on all three dimensions of affect, but our goal was to examine the effects of changes that are within the range of what naturally occurs. Within this modest range, pitch manipulations for music and speech had opposite effects on perceived valence, and somewhat different effects for perceived energy and tension.

Manipulations sometimes had interactive effects on judgments. For both music and speech stimuli, intensity manipulations had greater effects on valence ratings when the pitch height was raised than when the pitch height was lowered. However, the nature of the interactions was somewhat different in the two domains. For music, valence ratings were comparatively low when stimuli were presented loudly and at a raised pitch level, and comparatively high for all soft music. For speech, valence ratings were comparatively high when stimuli were presented softly and at a raised pitch level.

One explanation for the latter finding is that speech that is soft and higher-pitched may be preferred because of its association with infant-directed speech (IDS) or *motherese* (Drach, Kobashigawa, Pfuderer, & Slobin, 1969). This form of speech may be perceived as pleasant because of its special biological function (Pegg, Werker, & McLeod, 1992). Infant-directed speech has been observed in a variety of cultures and is associated with several perceptual benefits (Ruke-Dravina, 1977; Grieser & Kuhl, 1988; Fernald et al., 1989; Shute & Wheldall, 1989). For example, it attracts infants' attention by making speech more interesting (Snow, 1972); it facilitates language-learning tasks (Fernald & Mazzie, 1991; Kaplan, Jung, Ryther, & Zarlengo-Strouse, 1996); and it makes it easier for listeners to separate speech from background noise (Colombo, Frick, Ryther, Coldren, & Mitchell, 1995). Adults also judge infants who are listening to infant-directed speech as more interesting and charming, suggesting that it functions to maintain strong emotional ties between parents and infants (Werker & McLeod, 1989).

The finding that manipulations of acoustic properties influenced affective judgments of both music and speech is compatible with the evolutionary perspective advanced by Juslin and Laukka (2003). In their view, music and speech convey affective meaning by means of similar acoustic parameters. Our findings suggest that intensity manipulations had similar affective consequences for all dimensions of affect, rate manipulations had similar consequences for judgments of energy, and

pitch manipulations had different affective consequences in music and speech. Similarities between music and speech help to explain why listeners perceive music as expressive of emotion (Juslin & Laukka, 2003; Kivy, 1980) and suggest that emotional processing in music and speech involves shared neural resources (see also Patel & Peretz, 1997).

Differences in the effects of acoustic manipulations in music and speech may reflect distinctive brain processes for processing acoustic qualities in music and speech. That is, the affective consequences of changing intensity, tempo, and pitch may occur through a combination of domain-specific and domain-general processes. The existence of domain-specific processes for decoding emotion would be consistent with other forms of neuropsychological dissociations between music and language (e.g., Peretz & Coltheart, 2003). Our view, however, is that the acoustic manipulations examined do not engage different neural resources for speech and music, but are associated with affect through domain-general mechanisms. Nonetheless, interpretation of the affective connotations of such manipulations does not occur independently of the domain.

According to Scherer (2003), the process of decoding affective meaning in music and speech may be explained using Brunswik's (1956) lens model as the framework. According to this model, speakers, music performers, and listeners communicate emotional messages in predictable yet flexible ways. When one cue is unavailable in a performance (e.g., manipulations of intensity when playing a harpsichord), another cue may be substituted to convey a similar affective message. Listeners, in turn, use flexible decoding strategies that allow for such substitutions. Because stimulus properties (e.g., intensity, rate, and pitch height) may be substituted for one another, they are inherently imperfect predictors of specific emotions. For example, high-pitched speech is associated with both anger and happiness and is therefore not a perfect predictor of either emotion.

From this perspective, differences in the influence of affective cues in music and speech may arise because different attentional strategies are used for the two types of stimuli. Listeners may allocate greater attentional resources to salient aesthetic properties of music, whereas they may attend primarily to acoustic attributes of speech that concern its verbal and prosodic elements. In other words, different levels of attention may be directed to various acoustic qualities in the two domains, giving rise to differences in the affective consequences of stimulus manipulations. Associations that are specific to each domain, such as the association between high-pitched speech and *motherese*, may further allow for differences in the affective consequences of manipulating acoustic attributes in music and speech.

Author Note

This research was supported by the Natural Sciences and Engineering Research Council of Canada through a Canada Graduate Scholarship awarded to the first author and a discovery grant awarded to the second author. We thank Carmella Boscarino, Doug Bors, and Ulrich Schimmack for helpful comments.

Address correspondence to: either author, Department of Psychology, University of Toronto at Mississauga, Mississauga ON, Canada L5L 1C6. E-MAIL *ghusain@psych.utoronto.ca; b.thompson@utoronto. ca*

References

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125*, 238–251.

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.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.

Bolinger, D. (1978). Intonation across languages. In J. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals in human language: Vol. 2. Phonology* (pp. 472–524). Palo Alto, CA: Stanford University Press.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: MIT Press.

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

Colombo, J., Frick, J. E., Ryther, J. S., Coldren, J. T., & Mitchell, D. W. (1995). Infants' detection of analogs of "motherese" in noise. *Merrill-Palmer Quarterly, 41*, 104–113.

Cunningham, J. G., & Sterling, R. S. (1988). Developmental change in the understanding of affective meaning in music. *Motivation and Emotion, 12*, 399–413.

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

DeNora, T. (2001). Aesthetic agency and musical practice: New directions in the sociology of music and emotion. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 161–180). New York: Oxford University Press.

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception, 21*, 339–356.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 389–410). Cambridge, MA: MIT Press.

Drach, K. M., Kobashigawa, C., Pfuderer, C., & Slobin, D. I. (1969). *The structure of linguistic input to children.* University of California at Berkeley: Language Behavior Research Laboratory, Working Paper No. 14.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66*, 103–114.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*, 209–293.

Fernald, A., Taescher, T., Dunn, J., Papousek, M., Boysson-Bardies, B. D., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477–501.

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

Gold, A. E., MacLeod, K. M., Deary, I. J., & Frier, B. M. (1995). Changes in mood during acute hypoglycemia in healthy participants. *Journal of Personality and Social Psychology, 68*, 498–504.

Goldstein, A. (1980). Thrills in response to music and other stimuli. *Physiological Psychology, 8*, 126–129.

Grieser, D. L., & Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology, 24*, 14–20.

Gundlach, R. H. (1935). Factors determining the characterization of musical phrases. *American Journal of Psychology, 47*, 624–644.

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

Hevner, K. (1935b). The affective character of the major and minor modes in music. *American Journal of Psychology, 47*, 103–118.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology, 48*, 246–268.

Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology, 49*, 621–630.

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

Joseph, R. (1988). The right cerebral hemisphere: Emotion, music, visual-spatial skills, body image, dreams, and awareness. *Journal of Clinical Psychology, 44*, 630–673.

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1*, 381–412.

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

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

Kaplan, P. S., Jung, P. C., Ryther, J. S., & Zarlengo-Strouse, P. (1996). Infant-directed versus adult-directed speech as signals for faces. *Developmental Psychobiology, 32*, 880–891.

Kivy, P. (1980). *The corded shell*. Princeton, NJ: Princeton University Press.

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

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception, 13*, 401–432.

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51*, 336–352.

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

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception, 21*, 289–311.

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

Morton, E. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound symbolism* (pp. 348–365). Cambridge, England: Cambridge University Press.

Ohala, J. (1984). An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica, 41*, 1–16.

Panksepp, J. (1995). The emotional source of "chills" induced by music. *Music Perception, 13*, 171–207.

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

Patel, A. D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In I. Deliege & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 191–216). East Sussex, England: Psychology Press.

Patel, A. D., Peretz, I., Tramo, M., & Labrecque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language, 61*, 123–144.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development, 15*, 325–345.

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

Pinker, S. (1995). *The language instinct*. Harper-Collins Publishers, Inc., NY.

Ruke-Dravina, V. (1977). Modifications of speech addressed to young children in Latvian. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 237–253). Cambridge, England: Cambridge University Press.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.

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

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1–2), 227–256.

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

Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361–392). New York: Oxford University Press.

Schimmack, U. (1999). Strukturmodelle der Stimmungen. Rückblick, Überblick, Ausblick [Structural models of mood: Review, overview, and a look into the future]. *Psychologische Rundschau, 50*, 90–97.

Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality, 14*, 325–345.

Schimmack, U., & Reisenzein, R. (2002). Experiencing activation: Energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion, 2*, 412–417.

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 393–414). New York: Oxford University Press.

Shute, B. H. (1987). Vocal pitch in motherese. *Educational Psychology, 7*, 187–205.

Shute, B., & Wheldall, K. (1989). Pitch alterations in British motherese: Some preliminary acoustic data. *Journal of Child Language, 16*, 503–512.

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

Sloboda, J. A. (1992). Empirical studies of emotional response to music. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 33–46). Washington, DC: American Psychological Association.

Sloboda, J. A., & Juslin, P. (2001). Psychological perspectives on music and emotion. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 71–104). New York: Oxford University Press.

Snow, C. (1972). Mothers' speech to children learning language. *Child Development, 43*, 549–565.

Terwogt, M. M., & van Grinsven, F. (1988). Recognition of emotions in music by children and adults. *Perceptual and Motor Skills, 67*, 697–698.

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

Thayer, R. E. (1978a). Toward a psychological theory of multidimensional activation (arousal). *Motivation and Emotion, 2*, 1–34.

Thayer, R. E. (1978b). Factor analytic and reliability studies on the Activation-Deactivation Adjective Check List. *Psychological Reports, 42*, 747–756.

Thayer, R. E. (1986). Activation (arousal): The shift from a single to a multidimensional perspective. In J. Strelau, F. Farley, & A. Gale (Eds.), *The biological basis of personality and behavior* (Vol. 1, pp. 115–127). London: Hemisphere.

Thayer, R. E. (1989). *The biopsychology of mood and arousal*. New York: Oxford University Press.

Thayer, J. F., & Levenson, R. W. (1983). Effects of music on psychophysiological responses to a stressful film. *Psychomusicology, 3*, 44–52.

Thompson, W. F., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica*, *158*(1-4), 407–424.

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

Thompson, W. F., Schellenberg, E. G., & Husain G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*, 46–61.

Tucker, D. M., & Williamson, P. A. (1984). Asymmetric neural control systems in human self-regulation. *Psychological Review, 91*, 185–215.

Watson, K. B. (1942). The nature and measurement of musical meanings. *Psychological Monographs, 54*, 1–43.

Wedin, L. (1972). Multidimensional study of perceptual-emotional qualities in music. *Scandinavian Journal of Psychology, 13*, 241–257.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology, 43*, 230–246.