

Decoding speech prosody: Do music lessons help?

William Forde Thompson, E. Glenn Schellenberg, & Gabriela Husain
University of Toronto

Three experiments revealed that music lessons promote sensitivity to emotions conveyed by speech prosody. After hearing semantically neutral utterances spoken with emotional (i.e., happy, sad, fearful, or angry) prosody, or tone sequences that mimicked the utterances' prosody, participants identified the emotion conveyed. In Experiment 1 ($N = 20$), musically trained adults performed better than untrained adults. In Experiment 2 ($N = 56$), musically trained adults outperformed untrained adults at identifying sadness, fear, or neutral emotion. In Experiment 3 ($N = 43$), 6-year-olds were tested after being assigned randomly to a year of keyboard, vocal, drama, or no lessons. The keyboard group performed equivalently to the drama group and better than the no-lessons group at identifying anger or fear.

In the past 10 years, the possibility of links between musical and nonmusical domains has generated excitement among researchers and the popular press. One line of research concerns short-term benefits in nonmusical domains that occur as a consequence of passive listening to music. In two widely cited studies (Rauscher, Shaw, & Ky, 1993, 1995), listening to music composed by Mozart led to temporary improvements in spatial abilities. As an instance of cross-modal priming, this result is remarkable because the priming stimulus (music) is seemingly unrelated to the task being primed (a spatial task). Subsequent research indicates, however, that the so-called "Mozart effect" has nothing to do with Mozart in particular or with music in general (Nantais & Schellenberg, 1999). Rather, the apparent benefits of listening to music result from differences in mood and arousal induced by the testing conditions (Chabris, 1999; Thompson, Schellenberg, & Husain, 2001). Listening to fast-tempo music in a major key induces positive moods and arousal levels that facilitate performance on spatial tasks (Thompson et al., 2001; Husain, Thompson, & Schellenberg, 2002).

A second line of research concerns long-term effects of formal training in music. The cognitive implications of taking music lessons are distinct from the short-term effects of music listening (Schellenberg, 2003). Whereas transient effects of music listening on spatial abilities are

said to be instances of priming (Rauscher et al., 1993, 1995; Shaw, 2000), beneficial effects of music lessons on nonmusical abilities are best classified as positive *transfer* effects. Transfer occurs when previous experience in problem solving makes solving a new problem easier (positive transfer) or more difficult (negative transfer). In our view, the issue of whether music lessons yield positive transfer effects is an open question that is amenable to empirical investigation.

Although transfer effects depend critically on similarity between the training and transfer contexts (Barnett & Ceci, 2002; Postman, 1971), previous studies of transfer between music lessons and nonmusical skills have focused on domains that are not linked closely to music. For example,

This research was supported by the International Foundation for Music Research. We thank the Royal Conservatory of Music for their cooperation in providing the lessons; Laura-Lee Balkwill for recording and providing the Tagalog sentences; Jane Campbell and Will Huggon for recruiting and testing the children; Patrik Juslin, Klaus Scherer, Uli Schimmack, and three anonymous reviewers for helpful comments on an earlier draft; and Raul Dudnic, Doug Gifford, Vlad Kosarev, and Cory Sand for technical assistance. Correspondence and requests for reprints should be sent to Bill Thompson, Department of Psychology, University of Toronto at Mississauga, Mississauga, ON, Canada L5L 1C6 (b.thompson@utoronto.ca).

positive associations with music lessons have been reported for measures of general intelligence (Lynn, Wilson, & Gault, 1989; Schellenberg, 2004), symbolic reasoning (Gromko &

relatively widespread in studies with musical stimuli (e.g., Bunt & Pavlicevic, 2001; Terwogt & van Grinsven, 1988, 1991; Thompson & Robitaille, 1992), but contextual factors play a role. For example, anger is expressed well on an electric guitar but not on a flute (Gabrielsson & Juslin, 1996). More generally, music's emotional connotations are dependent on a combination of factors, including instrumentation, musical structure, and performance expression (Gabrielsson & Lindström, 2001, Juslin, 1997, 2001). As one would expect, the ability to identify emotions expressed by music improves with age (Terwogt & van Grinsven, 1988, 1991), and older children consider more factors when making their judgments. For example, when asked to decide whether a piece sounds happy or sad, 6- to 8-year-olds consider tempo (fast or slow) *and* mode (major or minor), but 5-year-olds are influenced only by tempo (Dalla Bella et al., 2001).

Music also evokes physical responses that accompany emotions, such as tears, tingles down the spine (or "chills"), and changes in heart rate, breathing rate, blood pressure, and skin-conductance levels (Goldstein, 1980; Krumhansl, 1997; Panksepp, 1995; Sloboda, 1991, 1992; Thayer & Levenson, 1983). Adults report that they frequently listen to music in order to change their emotional state (Sloboda, 1992), which implies that these physiological changes are accompanied by changes in phenomenological experience. Responses to paper-and-pencil questionnaires confirm that musical properties such as tempo and mode affect listeners' self-reported mood and arousal levels (Husain et al., 2002). In short, a large body of research confirms that listeners decode the intended emotional content of a musical piece, and that they respond emotionally to music.

Emotion and speech prosody

Speech prosody refers to the *musical* aspects of speech, including its *melody* (intonation) and its *rhythm* (stress and timing). Prosody is often used to convey a speaker's emotions (Frick, 1985; Juslin & Laukka, 2001, 2003), a connection that was noted by Darwin

(1872). Since then, theorists have suggested that prosody emerges from the prelinguistic use of pitch to signal emotion (Bolinger, 1978, 1986). Similar prosodic cues are used across cultures to convey emotions (Bolinger, 1978), such that sensitivity to particular meanings does not depend on verbal comprehension. Distinct patterns of vocal cues signal specific emotions as well as the intensity with which emotions are communicated (Juslin & Laukka, 2001).

Prosodic patterns signaling happiness, sadness, anger, and fear are decoded well above chance levels, with anger and sadness decoded more reliably than happiness and fear (Banse & Scherer, 1996; Johnstone & Scherer, 2000). In some stimulus contexts, however, fear (Apple & Hecht, 1982) and happiness (Johnson, Emde, Scherer, & Klinnert, 1986; Juslin & Laukka, 2001) are decoded as well as anger and sadness. Happiness is associated with rapid tempo, high pitch, large pitch range, and bright voice quality; sadness is associated with slow tempo, low pitch, narrow pitch range, and soft voice quality; anger is associated with fast tempo, high pitch, wide pitch range, and rising pitch contours; and fear is associated with fast tempo, high pitch, wide pitch range, large pitch variability, and varied loudness (Scherer, 1986).

Speakers' use of prosody to convey emotions is particularly obvious in speech directed toward infants and young children, which is consistent with Bolinger's (1978, 1986) suggestion that prosody has its roots in prelinguistic speech. Indeed, prosodic cues are the only way to convey emotion when speaking to young infants. Compared with speech directed toward adults, infant-directed speech has higher pitch, larger pitch excursions, slower rate, shorter utterances, and longer pauses (e.g., Ferguson, 1964; Fernald & Mazzie, 1991). Such modifications are made across cultures and genders (e.g., Fernald & Simon, 1984; Fernald et al., 1989; Grieser & Kuhl, 1988). Infant-directed speech facilitates the language-acquisition process by highlighting important words and linguistic structures (e.g., Fernald & Mazzie, 1991; Kemler Nelson, Hirsh-

Pasek, Jusczyk, & Wright Cassidy, 1989), by hyperarticulating vowels (Burnham, Kitamura, & Vollmer-Conna, 2002), and by eliciting and maintaining young listeners' attention (Fernald, 1991; Werker & McLeod, 1989). It also promotes emotional bonding between the speaker and the listener (Trainor, Austin, & Desjardins, 2000).

Music and speech prosody

Emotions are expressed in music and speech through variations in rate, amplitude, pitch, timbre, and stress (Scherer, 1995; Juslin & Laukka, 2003). Evolutionary theories suggest that musical behavior evolved in conjunction with—or as an adaptation of—vocal communication (Pinker, 1995; Brown, 2000; Dissanayake, 2000; Joseph, 1988). For example, Dissanayake (2000) proposed that vocal interactions between mothers and infants provide the foundation for a system of emotional communication that is used in music and other arts. In her view, music, speech prosody, and facial expression share a common ancestry as temporal-spatial patterns used in affiliative interactions between mothers and infants.

Pitch contour and rhythmic grouping are critical dimensions in both music and prosody (Patel, Peretz, Tramo, & Labrecque, 1998). In music, pitch and temporal relations define musical tunes, which retain their identities across transformations in pitch level and tempo. In speech, pitch variation provides an important source of semantic and emotional information, and temporal properties help listeners determine boundaries between words and phrases. Descending pitch contours and syllables or notes of long duration typically mark ends of phrases in speech (Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991) and in music (Narmour, 1990). Even young infants parse speech (Hirsh-Pasek et al., 1987) and music (Jusczyk & Krumhansl, 1993) using this information. Musical pitch and speech intonation are also processed preferentially by the right hemisphere, whereas rhythms in music and speech are less clearly lateralized (e.g., McKinnon & Schellenberg, 1997; Peretz, 2001; Snow, 2000; Van Lancker & Sidtis, 1992). Moreover, music and speech share neural resources for combining

their basic elements (i.e., musical tones and words, respectively) into rule-governed sequences (Patel, 2003).

The present investigation

Parallels between music and speech prosody raise the possibility that skills acquired through training in music lead to enhanced sensitivity to emotions conveyed by prosody. In other words, music lessons might nurture a basic skill of emotional intelligence by engaging, developing, and refining processes used for perceiving emotions expressed musically, which, in turn, could have consequences for perceiving emotions expressed in speech. Such effects would represent instances of transfer between training in music and a domain that is similar on several dimensions.

Although effects of musical training on decoding emotions in music have been inconsistent (Gabrielsson & Juslin, 2003), two studies provided preliminary evidence consistent with our hypothesis that music lessons facilitate the ability to decode emotions conveyed by speech prosody. Nilsson and Sundberg (1985) presented music and law students with tone sequences consisting of the fundamental frequencies of voice samples (i.e., no semantic cues) recorded from depressed and nondepressed individuals. Music students were superior at identifying the emotional state of the speakers. The authors suggested that, "The mastery of the expression of emotional information in music, which is a prerequisite for a competent musician, would then correspond to an enhanced ability to decode emotional information in speech" (p. 515).

Thompson, Schellenberg, and Husain (2003) reported findings consistent with this perspective. In one experiment, musically trained and untrained participants heard an utterance spoken in a "happy" manner followed by a sequence of tones that either matched or mismatched the prosody (pitch and duration) of the utterances. The musically trained participants were better at judging whether the utterances and tone sequences matched. In a second experiment, listeners heard "happy" or "sad" sounding

utterances spoken in a foreign language (Spanish). Each utterance was followed by a low-pass filtered version of an utterance spoken with the same emotional tone. Again, musically trained listeners outperformed their untrained counterparts at judging whether the filtered version was derived from the preceding utterance.

In the present investigation, we conducted three experiments that examined whether training in music is predictive of increased sensitivity to emotions conveyed by prosody. In Experiments 1 and 2, we attempted to replicate and extend the findings of Nilsson and Sundberg (1985) and Thompson et al. (2003). Adults were tested on their ability to decode the emotional meaning of spoken utterances or tone sequences that mimicked the prosody of those utterances. Some of the adults had extensive training in music that began in childhood; others had no music lessons. We predicted that adults who received musical training as children would be better than untrained adults at identifying emotions conveyed by the utterances and tone sequences.

In Experiment 3, 6-year-olds were assigned to one year of keyboard, singing, drama, or no lessons, and tested subsequently on their sensitivity to the emotions conveyed by spoken utterances and by tone sequences. There were two experimental groups (i.e., keyboard and singing) and two control groups (i.e., drama and no lessons). The ability of children in the no-lessons group to decode prosody should represent that of the average child. By contrast, vocal expression of emotion is central to drama training. Thus, children who received a year of drama lessons should be better than average at identifying emotions conveyed through prosodic cues. We also expected that training in music would lead to above-average abilities at decoding prosody in speech. It was unclear whether the music groups would perform as well as the drama group, or whether the singing group would perform as well as the keyboard group. On the one hand, singing lessons emphasize the use of the voice and might facilitate the ability to decode the emotional content of vocal utterances in general, and more so

than keyboard lessons. On the other hand, singing lessons emphasize controlled use of the voice to produce a sequence of discrete pitches. This non-prosodic use of the voice could interfere with decoding prosodic expressions of emotion.

Experiment 1

Musically trained and untrained adults were assessed on their ability to decode the emotions conveyed by tone sequences that mimicked the pitch and temporal structure of spoken phrases. Unlike the “matching” judgments used by Thompson et al. (2003), our task required listeners to identify the corresponding emotion. We focused on pitch and temporal structures because they are the most musically relevant dimensions of prosody. To illustrate, a familiar song such as *Happy Birthday* can be identified regardless of timbre (e.g., sung, performed on the piano) or amplitude (e.g., soft, loud), provided the pitch and temporal relations among tones conform to those that define the tune.

Our review of the literature motivated two predictions. The primary prediction was that musically trained participants would outperform their untrained counterparts. We also expected that identification accuracy would differ across the four emotions. As noted, sadness and anger are typically easier to decode than happiness and fear for spoken stimuli, whereas sadness and happiness are easier to decode than fear and anger for musical stimuli. Because our tone sequences combined prosodic and musical features, we expected that identification accuracy would be relatively high for sad sequences but relatively low for fearful sequences.

Method

Participants. Twenty undergraduates (12 men, 8 women) participated in the study. They were recruited from introductory psychology classes and received course credit for participating. The musically trained group consisted of four women and five men who had at least 8 years of formal music lessons ($M = 13.3$ years, $SD = 5.5$ years). All of them began taking music lessons during childhood. On average, they

had 1.7 years of college ($SD = 0.6$ years) and a mean grade point average of 3.0 ($SD = 0.6$). The untrained group consisted of four women and seven men. Ten had never taken music lessons; one had a year of lessons. The average participant in the untrained group had 2.3 years of college ($SD = 0.7$ years) and a grade point average of 2.6 ($SD = 0.7$). The two groups did not differ in age, years of education, or grade point average ($ps \geq .2$).

Apparatus. Stimuli were presented to participants under computer control (Macintosh G4). A customized software program created with Hypercard presented stimuli and recorded responses. Participants listened to the stimuli over Sennheiser HD 480 headphones at a comfortable volume while sitting in a sound-attenuating booth. The stimuli were presented with a flute timbre from a Roland 110 Instrument Set.

Stimuli. Tone sequences were melodic analogues of spoken sentences. Specifically, they were constructed to mimic the prosody of spoken sentences included in the *Name Emotional Prosody* test of the *Florida Affect Battery* (Bowers, Blonder, & Heilman, 1991). The test consists of four semantically neutral sentences (e.g., "The chairs are made of wood") uttered by a female speaker in four different renditions, with each rendition conveying one of four emotions: happiness, sadness, fear, or anger.

Each of the 16 sentences (4 sentences X 4 emotions) was transformed into an analogous tone

sequence in a manner similar to that of Patel et al. (1998). Tone "syllables" were created by calculating the modal pitch and duration of each spoken syllable. The modal pitch was established by locating the longest duration of pitch stability in the spoken syllable, ignoring variations up to 10 Hz (cycles/second) in frequency. When such variation was present, the mode was taken as the median of the frequencies. The selected pitch was verified by a musically trained assistant who compared the tone syllable to the spoken syllable presented in isolation. Tone syllables were combined to create a tone-sequence counterpart for each of the 16 sentences.

Unlike natural speech, each tone syllable had equal amplitude. Moreover, spoken language has pitch glides (i.e., continuous transitions), whereas the tone sequences had discrete steps although these did not conform to any musical scale. Discrete pitches were used in order to convey the syllabic segments that give spoken utterances their essential rhythmic character (i.e., segmentation in speech is conveyed poorly by continuous pitch changes when consonants are absent). In short, the translation from spoken sentences to tone sequences isolated pitch and timing dimensions of prosody, which are known to be important cues to the emotions conveyed by speakers (e.g., Juslin & Laukka, 2001, 2003). Table 1 (upper) provides duration and pitch information for tone sequences used in each of the four emotion categories. The table confirms that

Table 1. Descriptive statistics for the tone sequences used as stimuli in Experiments 1 (English) and 2 (English and Tagalog). The values are averaged across the four tone sequences used in each emotion category. Duration values are related inversely to the speed (tempo) of the sequences. Frequency values correspond to perceived pitch (higher frequency = higher perceived pitch).

	<i>Happy</i>	<i>Sad</i>	<i>Fearful</i>	<i>Angry</i>
<i>English</i>				
Duration (seconds)	1.48	1.79	1.51	1.76
Highest Frequency (Hz)	462.14	267.19	290.98	244.40
Lowest Frequency (Hz)	178.73	184.99	240.18	130.62
<i>Tagalog</i>				
Duration (seconds)	1.55	1.85	1.70	1.31
Highest Frequency (Hz)	405.75	203.82	296.96	332.13
Lowest Frequency (Hz)	263.36	162.02	219.06	221.46

these cues were typical of the intended emotions. For example, "happy" sequences were relatively quick (short duration) with a wide pitch (frequency) range, whereas "sad" sequences were slow with low pitch and a narrow pitch range.

Procedure. Participants were tested individually. They were told that they would hear a total of 16 tone sequences and that for each they should choose the emotion conveyed from a set of four alternatives. They were also told that each tone sequence mirrored the pitch and temporal information of a spoken phrase, and that the original phrases were spoken in a way that conveyed a happy, sad, fearful, or angry emotion. Before the test session began, practice trials were allowed until participants understood the task. The practice trials were drawn at random from the same set of stimuli as the test trials. Participants typically completed two or three practice trials before initiating the 16 test trials.

Results and Discussion

Each participant had four scores that represented the percentage of correct responses in each of the four conditions. Each condition corresponded to one of the four target emotions. The data are illustrated in Figure 1. The main analysis consisted of a 2 X 4 mixed-design analysis of variance

(ANOVA) with musical training (trained or untrained) as the between-subjects factor and emotion (happiness, sadness, fear, or anger) as the within-subjects factor. The main effect of training was reliable, $F(1, 18) = 8.26, p = 0.010$. In line with our predictions, the musically trained group ($M = 45\%$ correct) performed better than the untrained group ($M = 29\%$). Average levels of performance exceeded chance levels (25% correct) for the musically trained participants, $t(8) = 4.72, p = .001$, but not for their untrained counterparts.

The main effect of emotion was also reliable, $F(3, 54) = 2.82, p = .048$. As expected, performance was best in the sad condition ($M = 49\%$ correct), worst in the fearful condition ($M = 26\%$), and intermediate in the happy and angry conditions (M s = 42% and 28%, respectively). Performance in the sad condition exceeded performance in the other three conditions, $F(1, 54) = 4.75, p = .034$, whereas performance in the fearful condition was marginally worse than performance in the other three conditions, $F(1, 54) = 2.99, p = .090$. The lack of a two-way interaction between musical training and emotion indicates that differences among emotions were similar for the two groups of participants (see Figure 1).

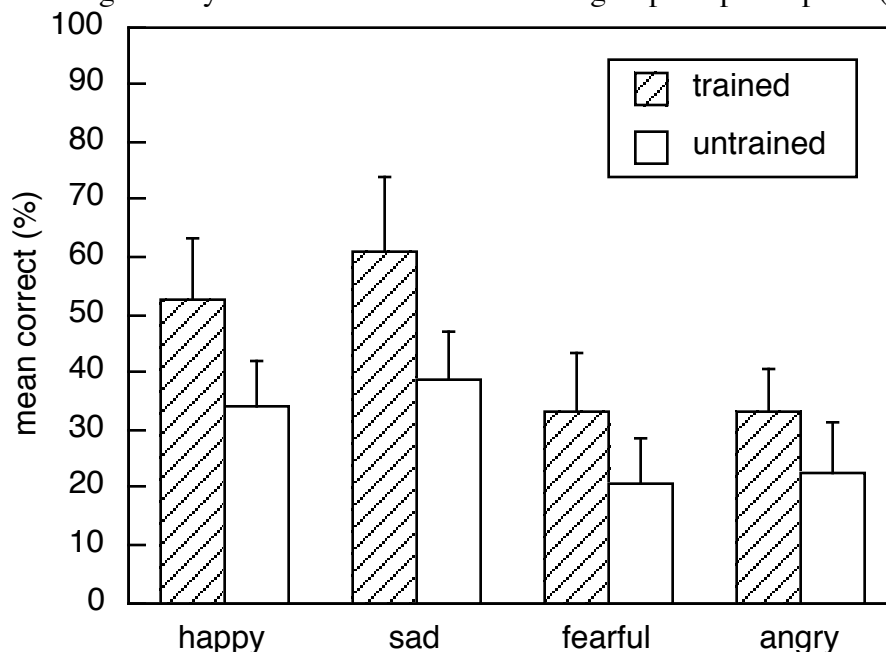


Figure 1. Mean levels of performance in Experiment 1 (adults) as a function of musical training and emotion. Error bars represent standard errors.

In sum, the findings are consistent with the hypothesis that music lessons are associated positively with decoding speech prosody. Previous findings indicate that musicians exhibit advantages in detecting whether a prosodic pattern came from a depressed person (Nilsson & Sundberg, 1985), and in extracting the pitch and duration patterns from happy and sad sounding speech (Thompson et al., 2003). The present findings reveal an advantage for musically trained adults in identifying emotions conveyed by prosodic cues. The results also corroborate previous indications that sadness is expressed with relatively distinctive and salient cues whether it is conveyed musically or prosodically (Bunt & Pavlicevic, 2001; Sloboda & Juslin, 2001; Terwogt & van Grinsven, 1988, 1991).

An alternative interpretation of these data is that trained listeners performed better than untrained listeners because the stimuli were tone sequences. Although the sequences were not tonal melodies (i.e., in a recognizable key) and did not sound like music, musicians are known to be better than nonmusicians at processing unconventional musical sequences (e.g., Lynch, Eilers, Oller, Urbano, &

fearful sounding. For tone sequences (which have prosodic and musical properties), we predicted that decoding sadness would be particularly accurate whereas decoding fear would be particularly inaccurate (as in Experiment 1).

Method

Participants. Fifty-six adults from a university community (18 men, 38 women) participated in the study. Some participants were recruited from introductory psychology classes and received course credit for participating. Others were recruited from the Faculty of Music and received token remuneration. The musically trained group consisted of 24 women and 4 men who had at least 8 years of music lessons ($M = 12.5$ years, $SD = 3.5$ years). All of them began taking music lessons during childhood. The untrained group consisted of 14 women and 14 men. None had ever taken private music lessons.

Apparatus. PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993) installed on a Macintosh computer (iMac) was used to create a customized program that controlled presentation of stimuli and collection of responses. The auditory stimuli were presented through Telephonics TDH-39P headphones at a comfortable listening level. As in Experiment 1, participants were tested in a sound-attenuating booth.

Stimuli. The stimuli consisted of 20 sentences uttered in English, 20 sentences uttered in Tagalog, and 32 tone-sequences derived from the utterances. English sentences were taken from the speech prosody section of the Florida Affect Battery. Tagalog sentences were a subset of sentences used by Balkwill, Thompson, and Schubert (2002). In both languages, four sentences with semantically neutral content were uttered by a native female speaker in five different ways, corresponding to the four target emotions (happy, sad, fearful, and angry) plus one with *neutral* emotion. Utterances with neutral emotion were included to avoid ceiling levels of performance on the emotion-identification task.

Tone sequences were derived from utterances spoken in a happy, sad, fearful, or

angry manner using the procedure described in Experiment 1. Tone sequences were not derived from emotionally neutral utterances because their inclusion could have made the emotion-identification task formidable, particularly for untrained listeners. The 16 tone sequences derived from English sentences were identical to those used in Experiment 1 (Table 1, upper). Sixteen additional tone sequences were derived from Tagalog sentences (Table 1, lower). In sum, there were 40 spoken utterances (2 languages X 4 sentences X 5 emotions) and 32 tone sequences (2 languages X 4 sentences X 4 emotions).

Procedure. Participants were tested individually in the sound-attenuating booth. They were told that they would hear 40 spoken utterances and 32 tone sequences and that for each they should select the emotion conveyed from the set provided. They were also told that each tone sequence mirrored the pitch and temporal information of a spoken phrase, and that the original phrases were spoken in a way that conveyed a happy, sad, fearful, angry, or neutral emotion.

In the first task of the test phase, listeners heard the 40 spoken utterances presented in random order. On each trial, they decided whether the speaker sounded happy, sad, angry, fearful, or emotionally neutral by clicking one of the options displayed on the computer screen. In the second task, listeners heard the 32 tone sequences derived from happy, sad, fearful, and angry sounding spoken utterances. Listeners were told that tone sequences were derived from spoken utterances presented in the first task and they were encouraged to imagine that each was a sentence spoken by the computer. On each trial, they decided whether the tone sequence was derived from happy, sad, fearful, or angry speech.

Before the first (spoken utterances) and second (tone sequences) tasks began, participants had two and four practice trials, respectively, drawn randomly from the same sets of stimuli used in the actual tasks. Following the test phase, participants were given a maximum of 20 minutes to complete the short form of the Raven's

Table 2. Mean percent correct (and standard deviation) in Experiment 2.

	<i>Musically Trained</i>		<i>Musically Untrained</i>	
	<i>English</i>	<i>Tagalog</i>	<i>English</i>	<i>Tagalog</i>
<i>Spoken Utterances</i>				
<i>Happy</i>	96.4 (13.1)	58.9 (20.7)	97.3 (7.9)	63.4 (24.1)
<i>Sad</i>	87.5 (21.0)	77.7 (26.7)	83.9 (24.7)	70.5 (23.6)
<i>Fearful</i>	73.2 (21.5)	70.5 (19.3)	67.9 (23.4)	47.3 (29.9)
<i>Angry</i>	82.2 (17.8)	96.4 (8.9)	90.2 (19.7)	100.0 (0.0)
<i>Neutral</i>	85.7 (18.6)	92.0 (16.8)	81.3 (20.0)	76.8 (22.5)
<i>Tone Sequences</i>				
<i>Happy</i>	51.8 (31.1)	39.3 (25.9)	60.7 (27.6)	50.0 (24.5)
<i>Sad</i>	53.6 (23.3)	75.0 (23.6)	44.7 (23.9)	60.7 (26.7)
<i>Fearful</i>	34.8 (22.9)	39.3 (24.0)	33.0 (25.5)	33.0 (24.6)
<i>Angry</i>	32.2 (27.1)	37.5 (24.1)	27.7 (25.8)	40.2 (25.8)

Advanced Progressive Matrices (Bors & Stokes, 1998). On average, participants took approximately 10 minutes to complete the Raven's test and 20 minutes to complete the entire procedure.

Results and Discussion

Each listener had 5 scores representing the percentage of correct responses for the five conditions with spoken stimuli, and 4 scores for the four conditions with tone sequences. Descriptive statistics are provided in Table 2. Performance levels for both trained and untrained participants were reliably higher than chance levels in all conditions ($ps < .001$). The primary analysis was a 2 X 2 X 2 X 4 mixed-design ANOVA with one between-subjects factor (musical training) and three within-subjects factors: language (English or

9707

for Tagalog compared to English spoken conditions, noted above) was not evident for the angry sounding stimuli. In other words, the speaker who uttered the Tagalog sentences in an angry manner was relatively successful at conveying her intended emotion to our sample of English-speaking Canadians, and her unintelligible words did not interfere in these instances.

Effects of training were less straightforward than those observed in Experiment 1 but generally consistent with our hypothesis. The main effect of musical training was not significant but there was a significant interaction between emotion and musical training, $F(3, 162) = 5.77$, $p = .001$. Musical training did not interact with any other variable or combination of variables. Follow-up tests of the interaction between emotion and musical training showed enhanced decoding accuracy among trained adults for some emotions but not others. Specifically, we examined each of the four emotion categories separately with a 3-way

(musical training X language X modality) mixed-design ANOVA.

Figure 2 displays mean scores (percent correct) of trained and untrained listeners for each of the four emotions as well as for the neutral emotion condition. A significant performance advantage for the musically trained group was evident for the sad sounding stimuli, $F(1, 54) = 5.05$, $p = .029$, and for the fearful sounding stimuli, $F(1, 54) = 5.94$, $p = .018$. The groups performed similarly in the angry conditions, $F < 1$. Interestingly, the untrained group performed marginally better with the happy sounding stimuli, $F(1, 54) = 3.88$, $p = .054$. Although we do not have data that address this issue (data were recorded as correct or incorrect), one possibility is that untrained listeners had a bias to respond “happy” that would have inflated performance with happy-sounding stimuli (see also Juslin & Luakka, 2001).

Response patterns for spoken utterances with neutral emotion were analyzed separately with a 2 X 2 (musical training X language)

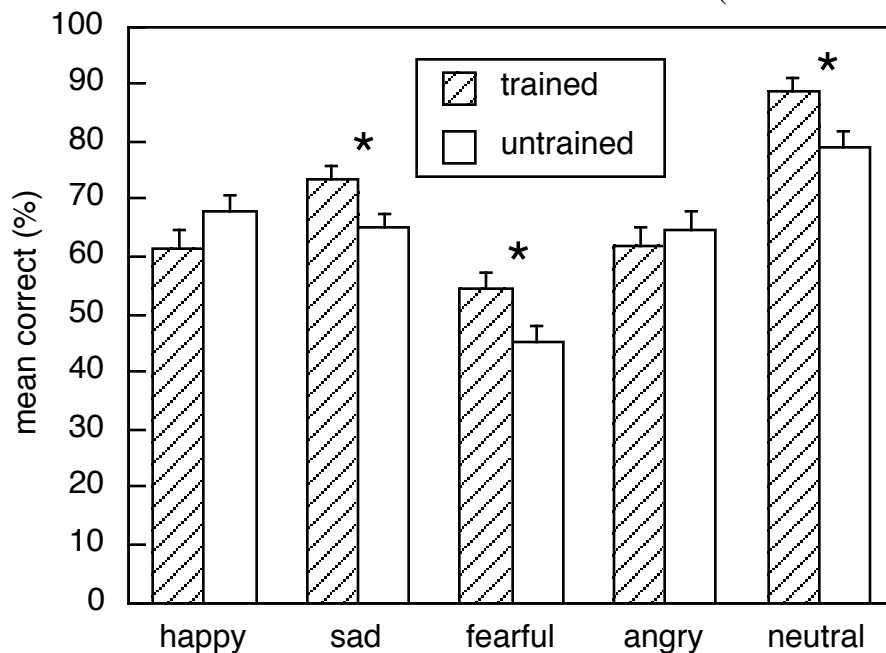


Figure 2. Mean levels of performance in Experiment 2 (adults) as a function of musical training and emotion. Error bars represent standard errors. Asterisks indicate that the effect of music training is significant at $p < 0.05$. Means for happy, sad, fearful and angry emotions are averaged across English and Tagalog spoken utterances and tone sequences. Means for the neutral emotion are averaged across English and Tagalog spoken utterances.

mixed-design ANOVA. As with the sad and fearful sounding stimuli, the neutral sounding utterances were identified better by musically trained ($M = 89\%$ correct) than by untrained ($M = 79\%$) listeners, $F(1, 54) = 5.64, p = .021$. There was no main effect of language and language did not interact with musical training.

Although benefits of musical training were observed for some emotions but not for others, it is remarkable that enhanced performance levels were evident not only for tone sequences and for English speech, but also for utterances spoken in a language that our participants did not understand (Tagalog). Indeed, a separate analysis of responses to Tagalog spoken utterances revealed a significant main effect of training, $F(1, 54) = 7.39, p = .009$, with musically trained adults ($M = 79\%$ correct) outperforming untrained adults ($M = 72\%$). To the best of our knowledge, this finding is the first to suggest that musical training is associated with enhanced sensitivity to emotions conveyed by prosody in a foreign language.

A final set of analyses examined scores on the short version of Raven's Advanced Progressive Matrices. Although the musically trained group had higher scores, $t(54) = 2.88, p = .006$, as one might expect (Schellenberg, 2004), these scores were *not* correlated with identification accuracy for English utterances ($p > .8$), Tagalog utterances ($p > .2$), English tone sequences ($p > .9$), or Tagalog tone sequences ($p > .5$). In fact, the association was negative (but non-significant) in one case. In short, it is highly unlikely that differences in fluid intelligence between musically trained and untrained adults were the source of differential responding on our emotion-identification tasks.

The results of Experiments 1 and 2 suggest that a personal history of music lessons in the childhood and teenage years may enhance one's ability to decode prosody in adulthood. As with any quasi-experiment, however, it is impossible to determine the direction of causation. Nonetheless, the results cannot be explained by differences in years of education,

grade-point average, or IQ. It is possible, however, that individuals with a naturally keen sensitivity to emotions expressed by prosody gravitate toward music lessons in their younger years, perhaps because they find the lessons more rewarding compared to their peers. In Experiment 3, we clarified this issue by adopting an experimental design. Specifically, we examined whether 6-year-olds assigned randomly to a year of music lessons would show similar advantages in decoding speech prosody.

Experiment 3

We assigned 6-year-olds randomly to one of four conditions and tested their prosody-decoding abilities a year later when they were 7 years of age. In three of the conditions, children took weekly arts lessons in the intervening year, during which they studied keyboard, singing, or drama in small groups. The fourth condition received no lessons. Because using the voice to convey emphasis, surprise, and emotion was central to the drama lessons, we expected that these children would be better than the no-lessons group at decoding prosody. We contrasted the abilities of the music (keyboard and singing) groups with the no-lessons group and with the drama group.

As in Experiment 2, the children were asked to make emotional judgments about the prosody of spoken utterances (English and Tagalog) in addition to making judgments about tone sequences. We predicted better performance with English compared to Tagalog stimuli, and with spoken utterances compared to tone sequences. We also predicted that the music groups would have better decoding skills than the no-lessons group. We had no predictions about whether the music groups would match the drama group in decoding accuracy, or whether the two music groups (keyboard and singing) would show equivalent performance levels. As noted earlier, performance could be enhanced or impaired for the singing group.

The particular age group (6 years when the lessons began) was chosen based on several

factors. The children needed to be old enough so that the lessons could be relatively rigorous, yet young enough so that the experience would have maximum impact on development. In music conservatories, 6-year-olds are considered mature enough to begin serious instruction in music. Six-year-olds are also more sensitive than younger children at decoding the emotions conveyed by tunes that conform to the rules of their musical culture (Dalla Bella et al., 2001). Evidence of reduced plasticity for children slightly older is provided by studies of *absolute pitch*. Children who take music lessons before the age of 7 are more likely than other children to have this rare ability, which implies a critical period for its acquisition (Takeuchi & Hulse, 1993).

Compared to Experiments 1 and 2, the task was simplified for children by reducing the number of alternatives in the forced-choice response to two (from four or five). On some trials, children heard either a happy or a sad sounding utterance or tone sequence. On other trials, the stimuli were fearful or angry sounding. In general, we predicted superior performance in the happy/sad conditions compared to the fearful/angry conditions.

Method

Participants. Forty-three 7-year-olds participated in the study (11 boys, 32 girls). Thirty had recently completed one-year of formal training in keyboard ($n = 10$), singing ($n = 11$), or drama ($n = 9$). A fourth, no-lessons group had no training in music or drama ($n = 13$). The children were assigned at random to one of the four conditions as they entered first grade (at 6 years of age). The lessons were provided free of charge. Children in the no-lessons group received training the next year.

The sample was recruited from a larger group of 144 families who volunteered to participate in a large-scale study designed to investigate whether arts lessons are predictive of intellectual development (Schellenberg, 2004). The children came from families in the local area who responded to a newspaper advertisement for

"free arts lessons." All of the families had a keyboard with full-sized keys but none of the children had prior music or drama lessons. The original study consisted of a pre-test [including the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991), the Kaufman Test of Educational Achievement (Kaufman & Kaufman, 1985), and the Parent Rating Scale of the Behavioral Assessment System for Children (Reynolds & Kamphaus, 1992)], the arts lessons (except for children in the no-lessons group), and a post-test (all tests re-administered). Because the lessons were taught outside the laboratory (at the conservatory), all four groups spent an equivalent amount of time in the laboratory prior to the present study. Table 3 provides descriptive statistics for scores on the pretest measures. In each case, there were no differences among groups. After the post-test, families were invited to participate in the present study. Children in the present sample were tested in the summer months between first and second grade. Each child received a gift certificate for participating.

Training. The children received weekly lessons at the Mississauga location of the Royal Conservatory of Music (Toronto). The lessons were approximately 45 minutes in length, taught to groups of six children. The instructors (two each for keyboard, singing, and drama) were professional female teachers who were Conservatory affiliates or graduates.

The keyboard lessons were designed by teachers at the Conservatory and consisted of traditional piano-training approaches using electronic keyboards. The children studied sight-reading, music notation, fingering patterns, playing from memory, and clapping rhythms. Children in the singing group received training in the Kodály method (Choksy, 1999), an intensive musical program that involves improvisation as well as singing, playing, and dancing to simple tunes and folk melodies. Children in the drama group studied simple scripts for plays, intonation, memorization of lines, staging, and acting. In all groups, children

Table 3. Means (and standard deviations) and group comparisons on the pretest measures in Experiment 3. VIQ, PIQ, FSIQ, and K-TEA are Z scores ($M = 100$, $SD = 15$). BSI and Adaptive Skills are T scores ($M = 50$, $SD = 10$).

Group	<i>VIQ</i>	<i>PIQ</i>	<i>FSIQ</i>	<i>K-TEA</i>	<i>BSI</i>	<i>Adaptive Skills</i>
Keyboard	106 (12)	109 (11)	108 (9)	109 (8)	43 (8)	57 (8)
Singing	107 (11)	117 (10)	112 (11)	104 (10)	45 (8)	60 (10)
Drama	108 (10)	107 (21)	107 (15)	106 (11)	41 (7)	57 (7)
No-Lessons	108 (11)	104 (14)	106 (11)	105 (10)	42 (6)	59 (11)
$F(3, 39)$	< 1	1.72	< 1	< 1	< 1	< 1
p -value	.95	.18	.57	.59	.72	.93

VIQ = WISC-III, Verbal IQ

PIQ = WISC-III, Performance IQ

FSIQ = WISC-III, Full-Scale IQ

K-TEA = Kaufman Test of Educational Achievement, Composite Score

BSI = Behavioral Assessment System for Children, Behavioral Symptoms Index (Maladaptive)

Adaptive Skills = Behavioral Assessment System for Children, Adaptive Skills Composite Score

were expected to practice at home between lessons.

Apparatus. The apparatus was identical to Experiment 2 except that children responded by pressing one of two buttons on a button box connected to the computer.

Stimuli. The stimuli were identical to those of Experiment 2 except the spoken utterances with neutral prosody were excluded. Hence, there were 32 spoken sentences (2 languages X 4 sentences X 4 emotions) and 32 corresponding tone sequences.

Procedure. Children were tested individually by an assistant who was *blind* to group assignment. They were informed that they would participate in a series of short tasks and that each would require eight responses. For each task, they were told to choose between two answers and shown how to respond.

In two tasks, children heard 8 happy and sad sentences presented in random order, once in English and once in Tagalog. They were asked to decide whether the speaker sounded happy or sad by pressing one button directly below a picture of a happy person, or a different button below a sad person. Another two tasks were identical to the first two, except that children heard the fearful and angry sentences (8 in English, 8 in Tagalog) and judged whether the speaker sounded scared or

angry by pressing one of two buttons below corresponding images.

The tone sequences were substituted for the spoken utterances to create another four tasks. The children were asked to imagine that each was a sentence spoken by the computer, and to decide whether the computer was happy or sad in two tasks, and scared or angry in the other two. The tasks were otherwise identical to the four with spoken sentences.

Each child was first tested in all four happy-sad conditions (e.g., spoken English, spoken Tagalog, English tone sequences, Tagalog tone sequences) or in all four fearful-angry conditions. In both cases, judgments of the spoken sentences preceded judgments of the tone sequences (as in Experiment 2). Four testing orders were formed, with emotion (happy-sad then fearful-angry, or vice versa) crossed with language (English then Tagalog or vice versa). Children were assigned at random to one of the four orders. They were allowed to take a break between tasks. The entire testing session took approximately 25 minutes.

Results and Discussion

Each child had eight scores corresponding to the percentage of correct responses in each testing condition. Descriptive statistics are

provided in Table 4. As shown in the table, performance was good and reliably better than chance (50% correct) for all groups in all happy/sad conditions. Performance was more variable in the fearful/angry conditions, being better than chance levels with spoken utterances in all instances, but at chance levels with tone sequences in some cases. Preliminary analyses revealed no differences among the four testing orders, and no interactions between testing order and any other independent variable or combination of variables. Testing order was not considered further.

Response patterns were analyzed with a 4 X 2 X 2 X 2 mixed-design ANOVA, with lessons (keyboard, singing, drama, or no lessons) as a between-subjects factor, and language (English or Tagalog), modality (spoken utterances or tone sequences), and emotion (happy-sad or fearful-angry) as within-subjects factors. As predicted, responses were more accurate for spoken utterances ($M = 92\%$ correct) than for tone sequences ($M = 75\%$), $F(1, 39) = 107.11, p < .001$, for English stimuli ($M = 88\%$) than for Tagalog stimuli ($M = 79\%$), $F(1, 39) = 26.24, p < .001$, and for happy/sad comparisons ($M = 91\%$) than for fearful/angry comparisons ($M = 76\%$), $F(1, 39) = 73.17, p < .001$. These main effects were qualified by three interactions: a two-way interaction

between emotion and modality, $F(1, 39) = 75.37, p < .001$, a two-way interaction between emotion and group, $F(3, 39) = 2.97, p = .043$, and a three-way interaction among emotion, language, and modality, $F(1, 39) = 10.89, p = .002$.

Because all three interactions involved emotion, follow-up analyses were conducted separately for the happy/sad and fearful/angry conditions. In both cases, we used a three-way (4 groups X 2 languages X 2 modalities) mixed-design ANOVA. For happy/sad conditions, high overall performance levels meant that differences between groups were not reliable. Hence, interactions in the original analysis appear to be a consequence of these near-ceiling effects. Lower performance levels in the fearful/angry conditions allowed for differences among groups to emerge. The main effect of group was reliable (see Figure 3), $F(3, 39) = 4.19, p = .011$, and did not interact with language or modality. The drama group outperformed the no-lessons group, $F(1, 39) = 4.35, p = .044$, as expected. More importantly, children in the music groups also outperformed the no-lessons group, $F(1, 39) = 4.79, p = .035$, and were no different from the drama group. Performance differed, however, between the keyboard and singing groups, $F(1, 39) = 6.71, p = .013$. The keyboard group was significantly better than the no-lessons group, $F(1, 39) = 10.13,$

Table 4. Mean percent correct (and standard deviation) in Experiment 3. Conditions in which performance exceeded chance levels (50% correct) are indicated with an asterisk ($p < .05$, one-tailed).

	Keyboard	Singing	Drama	No Lessons
<i>Spoken Utterances—English</i>				
Happy/Sad	100.0* (0.0)	97.7* (7.5)	97.2* (8.3)	97.2* (5.3)
Fearful/Angry	95.2* (6.2)	91.0* (12.7)	94.6* (9.0)	90.5* (15.4)
<i>Spoken Utterances—Tagalog</i>				
Happy/Sad	90.1* (16.4)	87.7* (18.4)	77.8* (29.8)	90.5* (17.6)
Fearful/Angry	95.1* (8.6)	84.3* (21.6)	94.6* (9.0)	85.8* (11.4)
<i>Tone Sequences—English</i>				
Happy/Sad	97.6* (5.1)	91.0* (12.7)	90.2* (16.5)	88.5* (14.1)
Fearful/Angry	78.6* (15.8)	65.9 (33.3)	65.3* (21.5)	67.3* (23.7)
<i>Tone Sequences—Tagalog</i>				
Happy/Sad	90.2* (9.8)	89.9* (12.3)	78.0* (18.6)	91.4* (14.0)
Fearful/Angry	58.9 (18.6)	51.3 (12.9)	59.7* (8.3)	42.4 (18.8)

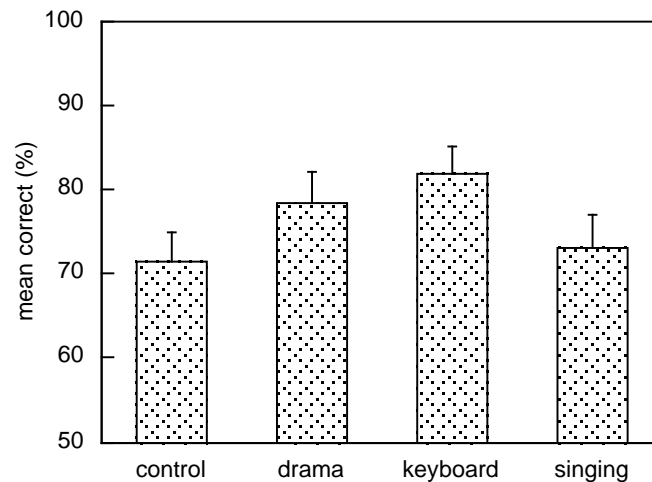


Figure 3. Mean levels of performance in Experiment 3 (children) in the fear/anger conditions as a function of training group. Error bars represent standard errors.

$p = .003$, and no different from the drama group. By contrast, the results were ambiguous for the singing group, which had intermediate levels of performance that were not significantly different from either the no-lessons group or the drama group.

In sum, the results were consistent with predictions. Happy/sad judgments were easier than fearful/angry judgments, spoken utterances were easier than tone sequences, and English stimuli were easier than Tagalog stimuli. Performance on our 2-alternative forced-choice task in the happy/sad conditions was very good across groups, precluding detection of differences among groups. Lower levels of performance in the fearful/angry conditions allowed differences among groups to emerge. Such differences were consistent with predictions and indicative of benefits in decoding prosody as a consequence of music lessons.

General Discussion

We tested whether music lessons enhance the ability to decode emotions conveyed by speech prosody. In Experiment 1, adults who took music lessons in childhood were better than untrained adults at identifying the emotions conveyed by tone sequences that mimicked the prosody of spoken utterances. In Experiment 2,

musically trained adults were better than untrained adults at identifying sadness and fear conveyed by utterances spoken in both familiar (English) and unfamiliar (Tagalog) languages, and by prosody-mimicking tone sequences. Musically trained adults were also better at identifying spoken utterances with emotionally neutral prosody. In Experiment 3, 7-year-olds were asked to identify the emotions conveyed in speech or in tone sequences that mimicked speech. For fearful/angry comparisons, children who took keyboard lessons during the previous year performed better than children with no arts lessons and equivalently to children with drama lessons. The equivalence between keyboard and drama lessons is particularly noteworthy because the drama lessons focused specifically on training the speaking voice and the use of prosody.

In Experiment 3, performance was better for children who received keyboard training than for children who received training in singing. As noted earlier, benefits of music lessons may have been attenuated for the singing group because singing lessons emphasize non-prosodic uses of the voice. Children in this group were trained to attend to voice-pitch mappings that do not coincide (and often conflict) with normal prosodic uses of the voice. Moreover, keyboards

produce precise pitches and pitch intervals, but the singing voice, particularly the voice of young singers, typically approximates exact pitch levels and intervals within a rather large margin of error (Dowling, 1999). The process of extracting meaning from pitch could become more rapidly refined among children taking keyboard lessons, who have repeated exposure to exact pitch levels and pitch intervals.

Another possibility is that the difference between the keyboard and singing lessons stemmed from pedagogical differences that distinguish Kodály training from standard conservatory approaches. Schellenberg (2004) found, however, that increases in full-scale IQ from pre- to post-lessons were larger among both music groups compared to the control (drama and no lessons) groups, with no difference between music groups. These effects were evident across the various dimensions (i.e., the four factor scores) and subtests of the WISC-III. The two music groups were also similar on measures of social adjustment and academic achievement. In sum, it would be a remarkable coincidence if the single difference between groups was in a domain related to the voice yet it was independent of vocal training.

A third possibility is that children in the keyboard group spent more time each week playing music than did children in the singing group. For example, keyboard practice may have more entertaining for children than practicing singing, because a keyboard could be construed as a toy. Unfortunately, the precise amount of time spent playing music each week is unknown, so we are unable to evaluate this possibility. Moreover, parents reported a minimal level of home practice across conditions (e.g., 10-15 minutes per week).

In general, decoding accuracy in Experiment 3 was better for English than for Tagalog stimuli, for spoken than for tonal stimuli, and for happy and sad sounding stimuli than for fearful or angry sounding stimuli. The difficulty with Tagalog relative to English stimuli likely reflects children's confusion when

asked to make judgments about speech in a foreign language, even though the English words did not indicate the speaker's emotion. Similar interference effects were evident among the adults tested in Experiment 2. Nonetheless, the robust advantage for spoken utterances over tone sequences in both languages confirms that acoustic variables *other* than pitch and duration, such as loudness and timbre, are important in conveying emotions through prosody (Juslin & Laukka, 2001, 2003).

The advantage for happy and sad stimuli among the children tested in Experiment 3 suggests that prosodic cues distinguishing happy from sad speech are more salient than those distinguishing fearful from angry speech. For example, it may be easier to distinguish a positive emotion from a negative emotion than to distinguish two negative emotions. Alternatively, the advantage may reflect contributions of familiarity. Assuming that expressions of happiness and sadness are more common than expressions of fear or anger in speech to children, performance advantages for *familiar* stimuli were evident in Experiments 2 and 3.

The present findings point to a complex association between musical training and the ability to decode speech prosody, while supporting the hypothesis that training in music enhances sensitivity to emotions conveyed by prosody. One explanation of this transfer effect is that similar or overlapping processes are involved in decoding emotional meaning in music and in speech prosody. Training in music might engage, refine, and develop skills used in associating pitch and temporal patterns with emotions. If the same skills are used for perceiving emotions conveyed musically or in speech, then music training would yield benefits for decoding emotions expressed in speech prosody. Indeed, it seems unlikely that human evolution led to duplicate mechanisms for associating pitch and temporal cues with emotions.

In Mayer et al.'s (1999, 2001) view, the ability to perceive emotions is the most basic

component of emotional intelligence. Our results suggest that development of this basic component can be facilitated through training in music. Although the status of emotional intelligence as a *distinct* intelligence is contentious (Izard, 2001; Kaufman & Kaufman, 2001; Roberts, Zeidner, & Matthews, 2001; Schaie, 2001), many theorists include emotion perception as a component of social intelligence (Gardner, 1993; Thorndike, 1920), empathic ability (Guilford, 1959), or crystallized intelligence (Cattell, 1971; Horn, 1988). Age-related improvements in emotion perception are thought to occur as a result of cognitive maturity, social interaction, and other relevant experiences (Roberts et al., 2001, p. 198). Taking music lessons appears to be one experience that accelerates age-related improvements in perceiving emotion.

Our data do not specify the precise stage in the information-processing system at which transfer occurs between music lessons and decoding prosody in speech. Transfer could occur at relatively high levels, when auditory signals are interpreted for emotional meaning. Another possibility is that transfer occurs at a lower level involving pitch analysis. Training in music is likely to engage and refine processes involved in the analysis of pitch patterns over time (Jakobson et al., 2003; Thompson et al., 2003), which could then be activated when interpreting emotions expressed in speech.

Although the current results suggest that music and speech prosody are processed by shared mechanisms, neuropsychological evidence makes it clear that many aspects of music engage domain-specific processes (Patel & Peretz, 1997, Peretz & Coltheart, 2003). One example is *tonality*—the system of organizing pitch around a reference point (the tonic) with which other tones are compared. Peretz et al. (1994) report a case of a brain-injured patient who is completely insensitive to this aspect of music but nonetheless sensitive to various aspects of language processing, such as phonology, syntax, and semantic meaning. Even

when two domains such as speech and music share neural resources, *apparent* dissociations can arise if the task is more relevant to one of the two domains. For example, in the first documented case of congenital amusia (Peretz et al., 2002), the patient had a severe deficit in fine-grained pitch discrimination. Remarkably, she interpreted intonation in speech correctly (e.g., an utterance that ends with rising pitch is likely to be a question). Her dissociation may highlight the different levels of pitch processing required for appreciating music and decoding prosody. Because prosody typically involves coarse-grained pitch variation, deficits in fine-grained pitch discrimination are more detrimental to music perception than to decoding prosody (Ayotte, Peretz, & Hyde, 2002; Hyde & Peretz, 2003).

In sum, perceiving music and speech are likely to engage a number of interacting processes, some of which are shared by both domains and some of which are domain-specific. Processes that lead to an emotional interpretation of auditory cues are likely to be shared by both domains. Indeed, there may be a general processor for the identification of emotional stimuli that extends to visual stimuli. For example, performance is correlated across tasks when adults are required to make judgments about the emotional content of speech prosody in one task, and the emotions expressed by faces in another task (Borod et al., 2000).

Other studies of possible by-products of music lessons have typically suffered from challenges in design, interpretation, and conceptualization (for a review, see Schellenberg, 2003). For example, associations between musical abilities and non-musical abilities in correlational studies or quasi-experiments could be the consequence of differences in general intelligence, socio-economic status, or other environmental factors. Simple comparisons between individuals taking music lessons and those with no lessons (Bilhartz, Bruhn, & Olson, 2000; Costa-Giomi, 1999; Ho et al., 2003; Rauscher, 2002) make it

impossible to attribute the findings to music lessons per se, rather than to additional lessons of any type, or perhaps to any extra-curricular activity (Schellenberg, 2003). Moreover, theories about possible links between music lessons and "spatial-temporal" reasoning (e.g., Rauscher et al., 1995; Shaw, 2000) posit improbable transfer effects between unrelated domains. By contrast, the present report tested the possibility of links

between music lessons and a domain that is closely related. Our results suggest that some types of music lessons are as effective as drama lessons in their capacity to nurture sensitivity to vocal expressions of emotion, and that such benefits are evident well into adulthood. Indeed, training in music may facilitate a basic skill of emotional intelligence: the ability to decode emotional meaning in speech prosody.

References

- Apple, W., & Hecht, K. (1982). Speaking emotionally: The relation between verbal and vocal communication of affect. *Journal of Personality and Social Psychology*, 42, 864-875.
- 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.
- Balkwill, L. L., & Thompson, W. F., & Schubert, E. (July, 2002). Rhetoric revisited: Cues to affective content in music and speech. In R. Parncutt and S. Larson (Chairs), *Art meets science*. Symposium conducted at the 7th International Conference on Music Perception and Cognition, Sydney, Australia.
- Banase, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70, 614-636.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, 128, 612-637.
- Bilhartz, T. D., Bruhn, R. A., & Olson, J. E. (2000). The effect of early music training on child cognitive development. *Journal of Applied Developmental Psychology*, 20, 615-636.
- Bolinger, D. (1978). Intonation across languages. In J. Greenberg (Ed.). *Universals in human language, vol. II: Phonology* (pp. 472-524). Palo Alto, CA: Stanford University Press.
- Bolinger, D. (1986). *Intonation and its parts*. Palo Alto, CA: Stanford University Press.
- Borod, J., Pick, L. H., Hall, S., Sliwinski, M., Madigan, N., Obler, L. K., Welkowitz, J., Canino, E., Erhan, H., Goral, M., Morrison, C., & Tabert, M. (2000). Relationship among facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition and Emotion*, 14, 193-211.
- Bors, D. A., & Stokes, T. L. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form. *Educational & Psychological Measurement*, 58, 382-398.
- Bowers, D., Blonder, L. X., & Heilman, K. M. (1991). *Florida Affect Battery (FAB)*. Gainesville, FL: University of Florida.
- 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.
- Bunt, L., & Pavlicevic, M. (2001). Music and emotion: Perspectives from music therapy. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 181-201). Oxford, UK: Oxford University Press.
- Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science*, 296, 1435.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and measurement*. Boston: Houghton Mifflin.

Chabris, C. F. (1999). Prelude or requiem for the "Mozart Effect"? *Nature*, 400, 826.

Choksy, L. (1999). *The Kodály method I: Comprehensive music education* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, 25, 257-271.

Cook, N. & Dibben, N. (2001). Musicological approaches to emotion. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 45-70). Oxford, UK: Oxford University Press.

Cooke, D. (1959). *The language of music*. London: Oxford University Press.

Costa-Giomi, E. (1999). The effects of three years of piano instruction on children's cognitive development. *Journal of Research in Music Education*, 47, 198-212.

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.

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

Davies, S. (2001). Philosophical perspectives on music's expressiveness. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 23-44). Oxford, UK: Oxford University Press.

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 1-24). Norwood, NJ: Ablex.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interactions. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 388-410). Cambridge, MA: MIT Press.

Dowling, W. J. (1999). Development of music perception and cognition. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 603-625). San Diego, CA: Academic Press.

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

Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. *Annals of Child Development*, 8, 43-80.

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

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20, 104-113.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & 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.

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

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). Oxford, UK: Oxford University Press.

Gardiner, M. F., Fox, A., Knowles, F., & Jeffrey, D. (1996). Learning improved by arts training. *Nature*, 381, 284.

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gaver, W. W., & Mandler, G. (1987). Play it again, Sam: On liking music. *Cognition and Emotion*, 1, 259-282.

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.

Gromko, J. E., & Poorman, A. S. (1998). Developmental trends and relationships in children's aural perception and symbol use. *Journal of Research in Music Education*, 46, 16-23.

Guilford, J. P. (1959). *Personality*. New York: McGraw-Hill.

Hetland, L. (2000). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education*, 34, (3/4) 179-238.

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.

Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Cassidy, K. W., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269-286.

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, 17, 439-450.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselrode & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 645-685). New York: Plenum.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, 20, 149-169.

Hyde, K. L., & Peretz, I. (2003). Brains that are out-of-tune but in-time. *Psychological Science*, in press.

Izard, C.E. (2001). Emotional intelligence or adaptive emotions? *Emotion*, 1, 249-257.

Jakobson, L. S., Cuddy, L. L., & Kilgour, A. R. (2003). Time tagging: A key to musicians' superior memory. *Music Perception*, 20, 307-313.

Johnson, W. F., Emde, R. N., Scherer, K. R., & Klinnert, M. D. (1986). Recognition of emotion from vocal cues. *Archives of General Psychiatry*, 43, 280-283.

Johnstone, T. & K.R. Scherer (2000). Vocal Communication of Emotion. In Michael Lewis and Jeannette M. Haviland-Jones (Eds.), *Handbook of Emotion: 2nd Edition*, The Guilford Press, New York, pp. 220-235.

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

Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 627-640.

Juslin, P. N. (1997). Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy. *Musicae Scientiae*, 1, 225-256.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 309-340). Oxford, UK: Oxford University Press.

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*. Oxford, UK: Oxford University Press.

Kaufman, A. S., & Kaufman, N. L. (1985). *Kaufman Test of Educational Achievement*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, J. C. (2001). Emotional intelligence as an aspect of general intelligence: What would David Wechsler say? *Emotion, 1*, 258-264.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*, 55-68.

Kilgour, A. R., Jakobson, L. S., & Cuddy, L. L. (2000). Music training and rate of presentation as mediators of text and song recall. *Memory & Cognition, 28*, 700-710.

Kivy, P. (1980). *The corded shell: Reflections on musical expression*. Princeton: 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. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology, 51*, 336-352.

Lamb, S. J., & Gregory, A. H. (1993). The relationship between music and reading in beginning readers. *Educational Psychology, 13*, 19-27.

Langer, S. K. (1957). *Philosophy in a new key* (3rd ed.). New York: New American Library.

Lynch, M. P., Eilers, R. E., Oller, K. D., Urbano, R. C., & Wilson, P. (1991). Influences of acculturation and musical sophistication on perception of musical interval patterns. *Journal of Experimental Psychology: Human Perception & Performance, 17*, 967-975.

Lynn, R., Wilson, R. G., & Gault, A. (1989). Simple musical tests as measures of Spearman's g . *Personality and Individual Differences, 10*, 25-28.

Mayer, J. D., Caruso, D. R., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence, 27*, 267-298.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2001). Emotional intelligence as a standard intelligence. *Emotion, 1*, 232-242.

McKinnon, M. C., & Schellenberg, E. G. (1997). A left-ear advantage for forced-choice judgements of melodic contour. *Canadian Journal of Experimental Psychology, 51*, 171-175.

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

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. *Psychological Science, 10*, 370-373.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Nilssonne, A., & Sundberg, J. (1985). Differences in ability of musicians and nonmusicians to judge emotional state from the fundamental frequency of voice samples. *Music Perception, 2*, 507-516.

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

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6, 674-681.

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

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

Peretz, I. (2001). Music perception and recognition. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology: What deficits reveal about the human mind* (pp. 519-540). Philadelphia: Psychology Press.

Peretz, I. & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, 6(7), 688-91.

Peretz, I., Ayotte, J., Zatorre, R., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, 33, 185-191.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociation following bilateral lesions of auditory cortex. *Brain*, 117, 1283-1301.

Pinker, S. (1995). *The language instinct: How the mind creates language*. New York: Harper Perennial.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90, 2956-2970.

Postman, L. (1971). Transfer, inference and forgetting. In J. W. King & L. A. Riggs (Eds.), *Woodworth & Scholberg's experimental psychology* (3rd ed., pp. 1019-1132). New York: Holt, Rinehart & Winston.

Rauscher, F. H. (2002). Mozart and the mind: Factual and fictional effects of musical enrichment. In J. Aronson (Ed), *Improving academic achievement: Impact of psychological factors on education* (pp. 267-278). San Diego: Academic Press.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature*, 365, 611.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letters*, 185, 44-47.

Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System For Children*. Circle Pines, MN: American Guidance Service.

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.

Schaie, K. W. (2001). Emotional intelligence: Psychometric status and developmental characteristics — Comment on Roberts, Zeidner, and Matthews. *Emotion*, 1, 243-248.

Schellenberg, E. G. (2003). Does exposure to music have beneficial side effects? In I. Peretz & R.J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 430-448). Oxford, UK: Oxford University Press.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, in press.

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

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, 9, 235-248.

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). Oxford, UK: Oxford University Press.

Shaw, G. L. (2000). *Keeping Mozart in mind*. San Diego: Academic Press.

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). Oxford, UK: Oxford University Press.

Snow, D. (2000). The emotional basis of linguistic and nonlinguistic intonation: Implications for hemispheric specialization. *Developmental Neuropsychology*, 17, 1-28.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, 113, 345-361.

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. M., & van Grinsven, F. (1991). Musical expression of moodstates. *Psychology of Music*, 19, 99-109.

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

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). Arousal, mood, and the Mozart effect. *Psychological Science*, 12, 248-251.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech: Effects of music lessons. *Annals of the New York Academy of Sciences*, in press.

Thorndike, E. L. (1920). Intelligence and its use. *Harper Magazine*, 140, 227-235.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, 11, 188-195.

Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research*, 35, 963-970.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Psychological Corporation.

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.