

# Speaking to the Trained Ear: Musical Expertise Enhances the Recognition of Emotions in Speech Prosody

César F. Lima and São Luís Castro  
University of Porto

Language and music are closely related in our minds. Does musical expertise enhance the recognition of emotions in speech prosody? Forty highly trained musicians were compared with 40 musically untrained adults (controls) in the recognition of emotional prosody. For purposes of generalization, the participants were from two age groups, young (18–30 years) and middle adulthood (40–60 years). They were presented with short sentences expressing six emotions—anger, disgust, fear, happiness, sadness, surprise—and neutrality, by prosody alone. In each trial, they performed a forced-choice identification of the expressed emotion (reaction times, RTs, were collected) and an intensity judgment. General intelligence, cognitive control, and personality traits were also assessed. A robust effect of expertise was found: musicians were more accurate than controls, similarly across emotions and age groups. This effect cannot be attributed to socioeducational background, general cognitive or personality characteristics, because these did not differ between musicians and controls; perceived intensity and RTs were also similar in both groups. Furthermore, basic acoustic properties of the stimuli like fundamental frequency and duration were predictive of the participants' responses, and musicians and controls were similarly efficient in using them. Musical expertise was thus associated with cross-domain benefits to emotional prosody. These results indicate that emotional processing in music and in language engages shared resources.

*Keywords:* speech prosody, emotion recognition, musical expertise, transfer effects

Language and music share important features. Both are finely structured systems of expression and communication based on perceptually discrete units organized into flowing acoustic sequences; both are universal in the human species and have ancient origins (Mithen, 2005). The comparative analysis of language and music is an active area of study in cognitive neuroscience: to which extent is each of them subserved by domain-specific mechanisms, or do both rely on common resources? Answers to this question will illuminate debates on modularity and evolution (Hauser & McDermott, 2003). Neuropsychological research suggests that musical abilities like tonal encoding of pitch are domain-specific (e.g., Peretz, 2008; Peretz & Coltheart, 2003). Neuroimaging and behavioral evidence indicates that other components of language and music might be domain-general (for a review, Patel, 2008a). This is the case of syntactic operations (Patel, 2003), perception of phrase boundaries (Knösche et al., 2005), pitch processing (Fitch, 2006; Marques, Moreno, Castro, & Besson,

2007; Schön, Magne, & Besson, 2004; Trehub & Hannon, 2006), and conceptual processing (Koelsch et al., 2004; Schön, Ystad, Kronland-Martinet, & Besson, 2010). The expression of emotions is one of the shared features of language and music. From a neurocognitive point of view, the question that arises is whether the processing of emotions in both domains engages common mechanisms. The present study examines this question by determining the effects of musical expertise on the recognition of emotional speech prosody.

Emotions can be communicated through speech prosody, that is, the modulation of suprasegmental properties of spoken language including pitch, loudness, tempo, rhythm, and timbre (Grandjean, Bänziger, & Scherer, 2006; Schirmer & Kotz, 2006). Emotions that are recognizable across cultures (e.g., Sauter, Eisner, Ekman, & Scott, 2010), like anger, happiness, sadness, and fear, have characteristic prosodic profiles (Banse & Scherer, 1996). Listeners can accurately identify a range of different emotions from speech, even in the absence of emotional semantic content, as in sentences that are neutral (the table is set) or composed of pseudowords (e.g., Castro & Lima, 2010), or sentences in an unknown language (Pell, Monetta, Paulmann, & Kotz, 2009). The ability to recognize emotions in prosody is pivotal for skilful communication (Carton, Kessler, & Pape, 1999) and might be compromised in neuropathological conditions such as Parkinson's disease and autism spectrum disorders (Golan, Baron-Cohen, & Hill, 2006; Pell & Leonard, 2003). Music is also a powerful means to express emotions through dynamic acoustic events unfolding in time, and emotions lie at the heart of music appreciation (Juslin & Västfjäll, 2008). The scientific study of emotions in music reveals that different emotional qualities are recognized accurately (e.g., Peretz,

---

César F. Lima and São Luís Castro, Faculty of Psychology and Education, University of Porto, Porto, Portugal.

We thank Maria Luís Almeida for assistance in testing participants. We also thank two anonymous reviewers for their valuable comments on an earlier version of the manuscript. This research was supported by grants from the Portuguese Foundation for Science and Technology (COMPETE and FEDER programs) and from Bial Foundation.

Correspondence concerning this article should be addressed to São Luís Castro, Faculty of Psychology and Education, University of Porto, Rua Doutor Alfredo Allen, 4200-135 Porto, Portugal. E-mail: slcastro@fpe.up.pt

Gagnon, & Bouchard, 1998; Vieillard et al., 2008) and universally (Fritz et al., 2009).

The relationship between emotions in speech and in music has long been debated, but relevant empirical evidence emerged only in the last years (Bowling, Gill, Choi, Prinz, & Purves, 2010; Curtis & Bharucha, 2010; Juslin & Laukka, 2003; Scherer, 1995). In a meta-analytic review of studies on vocal expression and music performance, Juslin and Laukka (2003) showed that the acoustic profiles that correspond to specific emotions in both domains are remarkably similar. This was observed for happiness, sadness, tenderness, anger, and fear. Happiness, for instance, is perceived in fast speech rate or musical tempo, medium-high intensity, and medium high-frequency energy, fundamental frequency (F0) that is high, rising and highly variable, fast voice onset or tone attacks, and little microstructural regularity; sadness is a mirror image of this pattern (slow rate/tempo, low intensity, and little high-frequency energy, low and falling F0 with reduced variability, slow onsets/attacks, and microstructural irregularity). The authors (ibid.) suggested that such similarities explain why we perceive emotions in music: by mimicking voice-like attributes of emotion, music would engage neural circuitries primarily dedicated to speech, thereby acquiring emotional salience.

A heuristic hypothesis is that processing emotions in speech and in music engages common mechanisms (Besson, Magne, & Schön, 2002; Juslin, Liljeström, Västfjäll, & Lundqvist, 2010; Juslin & Västfjäll, 2008; Patel, 2008b; Peretz, 2010). One strategy to test this hypothesis is to determine whether musical expertise transfers positively to emotional prosody. To achieve expert performance, musicians spend a massive amount of time in music learning and deliberate practice, usually 10 or more years (e.g., Ericsson & Lehmann, 1996; Ericsson & Towne, 2010). A central component of this practice concerns the production and perception of fine-grained modulations of complex acoustic patterns for expressive purposes. Indeed, years of music training correlate positively with the accuracy with which emotions are perceived in music (Lima & Castro, 2011; Livingstone, Muhlberger, Brown, & Thompson, 2010). If mechanisms are shared for emotion recognition in music and speech, then musicians should exhibit enhanced processing of emotional prosody. Regarding pitch, an important prosodic cue of emotion, Moreno and colleagues (2009) have shown that music training induces enhanced pitch processing. However, emotions expressed in speech are integrated patterns of multiple acoustic properties, not just pitch, and so emotional prosody requires direct examination. To our knowledge, only two studies analyzed the effects of musical expertise on emotional prosody, and they yielded conflicting results. Thompson, Schellenberg, and Husain (2004) reported suggestive evidence of a positive effect. There was an advantage of musicians (young adults) compared with untrained participants in a forced-choice identification of four emotions (happiness, sadness, fear and anger) in tone sequences that were melodic analogues of spoken sentences (Experiment 1); however, the untrained participants performed at chance-level in all emotions—they might not have understood the task, or the stimuli might not have been effective in conveying the emotions. With sentences in English and in a foreign language unknown to the participants (Tagalog), the advantage of musicians was restricted to two emotions, sadness and fear (not happiness or anger), and the main effect of expertise was not significant (Experiment 2). Finally (Experiment 3), in an emotion discrimination task (happy vs. sad,

fearful vs. angry pairs) children assigned to music lessons for a year had better accuracy than children without music lessons, but only for the subset on the discrimination between fearful and angry stimuli. These results are based on stimuli produced by a single female speaker, and this limits their generalizability. More recently, Trimmer and Cuddy (2008) reported negative evidence. One hundred undergraduate students with different degrees of music training were presented with sentences, and gliding tone analogues derived from the sentences, expressing happiness, sadness, fear, anger, or neutrality through prosody; their task was to rate the prominence of each emotion on 11-point scales. Music training did not correlate with the recognition of emotional prosody, and it was not a significant predictor in regression analyses. Interestingly, it was emotional intelligence that correlated and predicted the results on emotional prosody, specifically the experiential dimension of the Mayer-Salovey-Caruso emotional intelligence test (that assesses perceiving and facilitating emotions; the other dimension is strategic, and deals with understanding and managing emotions). The authors concluded that learning music is not associated with benefits in the recognition of emotional prosody and speculated that there might be a cross-modal system, unrelated to music training, that is responsible for recognizing emotions. However, it was not specified how many participants were indeed musically trained, just how many years of practice, on average, they had: 6.5 years. This is relatively low in comparison with studies on the effects of music training on language processing with adult participants, where musicians typically have an average of 12 or more years of training (e.g., Marques et al., 2007; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Schön, Magne, & Besson, 2004; Thompson et al., 2004). Because the vocal expression and discrimination of emotions is a basic biological function and thus, in all likelihood, efficient even in the absence of specific training (Hauser & McDermott, 2003; Patel, 2008a; Scherer, Johnstone, & Klasmeyer, 2003), it is possible that a fairly extensive period of musical practice is needed to detect differences in adults, especially in an offline behavioral task.

Taking into account the inconsistencies of the available evidence, the hypothesis that musical expertise has an impact on emotional speech prosody remains open, and we set out to reexamine it in the present study. Forty highly trained musicians, with 12 years of training on average, were compared with 40 musically naïve adults (hereafter, controls) in the recognition of emotional speech prosody. With respect to previous studies, important differences in the procedure were introduced. First, we covered a wider range of emotions by including the six emotions usually considered in emotion perception research: anger, disgust, fear, happiness, sadness, surprise, plus neutrality. Second, the stimuli underwent previous perceptual and acoustic validation (Castro & Lima, 2010), so that it was known beforehand that they effectively conveyed the intended expressions. Third, two additional measures were taken, reaction times (RTs) and intensity ratings. RTs were collected to control for possible speed-accuracy trade-offs; musicians are trained to carefully analyze acoustic signals, and they might have longer response latencies than controls (Chartrand & Belin, 2006). Intensity ratings were collected in order to control for the possibility that musicians might have increased responsiveness to emotional salience. Fourth, participants were from two age groups, young and middle-aged adults, to verify whether the possible effect of expertise is general and long-lasting, or whether

it is contingent on intensive formal training at the moment of testing, as was the case of most young musicians. Because previous research indicates that basic acoustic properties of vocal stimuli, such as F0 and duration, are used to identify emotions (acoustic cues predict the participants' categorizations, e.g., Banse & Scherer, 1996; Juslin & Laukka, 2001), we also conducted multiple regression analyses to explore how acoustic cues would be relied upon by musicians and controls. Finally, apart from general intellectual functioning and cognitive control, participants were also assessed for personality and sociocommunicative traits because these may influence emotion processes (Ashwin, Chapman, Colle, & Baron-Cohen, 2006; Hamann & Canli, 2004). If the recognition of emotions in speech prosody is responsive to musical expertise, then an advantage of highly trained musicians should be found. This would be evidence in favor of partly shared mechanisms in the domains of music and language.

## Method

### Participants

We tested 80 participants distributed into four groups according to musical expertise (musicians and controls) and age (younger and middle-aged), 20 per group (10 women). Table 1 presents their demographic and background characteristics. Younger participants ranged between 18 and 30 years of age, and middle-aged ones between 40 and 60 years. Musicians were instrumentalists who played piano ( $n = 18$ ), flute ( $n = 5$ ), violin ( $n = 5$ ), guitar ( $n = 3$ ), double bass ( $n = 2$ ), clarinet ( $n = 1$ ), drums ( $n = 1$ ), cello ( $n = 1$ ), oboe ( $n = 1$ ), viola ( $n = 1$ ), trombone ( $n = 1$ ), or accordion ( $n = 1$ ); in addition to instrumental training, two of them had vocal training in classical singing. They had at least eight years of musical training started in childhood and practiced regularly at the moment of testing. Younger musicians were advanced music students or professional musicians; older ones were music teachers and/or orchestral performers. They were recruited from local music schools and orchestras, including Conservatório de Música do Porto, Escola Superior de Música e Artes do Espectáculo (ESMAE/IPP), and Orquestra Sinfónica do Porto Casa da Música. Musicians in

the two age groups were similar regarding years of training, age of training onset, and average hours of weekly instrumental practice ( $F_s < 1$ ; see Table 1). Controls had not had formal music lessons, nor played any instruments. The four groups were matched for education, as assessed by the number of years attending school ( $F < 1$ ). All the participants were Portuguese speakers, and reported no speech disorders, no major psychiatric or neurological illnesses, and no head trauma or substance abuse. The groups rated themselves similarly for hearing acuity in a scale from 1, very good, to 6, very bad ( $M = 2.1$ ).

To assess global intellectual functioning, we administered the Montreal Cognitive Assessment test (MOCA; www.mocatest.org), a screening test which covers the main neurocognitive domains, and a timed version of the Raven's Advanced Progressive Matrices (20-min version; Hamel & Schmittmann, 2006). Musicians and controls did not differ on these measures ( $F_s < 1$ ). Older participants had lower scores, both in the MOCA [ $F(1, 76) = 5.47, p < .05, \eta_p^2 = .07$ ] and Raven's Matrices [ $F(1, 76) = 19.5, p < .01, \eta_p^2 = .2$ ; see Table 1], a result that is consistent with findings on cognitive aging (e.g., Salthouse, 2009). Cognitive control was assessed with a Stroop test; the version we used included reading color names without incongruity (baseline), and an incongruous color-naming task (Stroop effect; Trenerry, Crosson, Deboe, & Leber, 1995; Portuguese version, Castro, Martins, & Cunha, 2003). Musicians were faster than controls in the two conditions [baseline:  $F(1, 76) = 16.2, p < .001, \eta_p^2 = .18$ ; Stroop effect:  $F(1, 76) = 6.7, p < .05, \eta_p^2 = .08$ ], but their advantage in the critical incongruous condition did not reach significance when reading speed was entered as a covariate in an ANCOVA [ $F(1, 75) = 3.7, p = .06, \eta_p^2 = .05$ ]. An age-related decline was observed in the incongruous condition [ $F(1, 76) = 7.7, p < .01, \eta_p^2 = .09$ ], but not in the baseline ( $F < 1$ ). As a control for personality we used the Ten-Item Personality Inventory, a brief measure of the Big-5 personality domains (Gosling, Rentfrow, and Swann Jr., 2003). Because autistic traits are linked with structural and functional differences in brain regions involved in emotional processing in neurotypical individuals (Di Martino et al., 2009; Hagen et al., 2011) we also administered the questionnaire The Autism-

Table 1  
*Demographic and Background Characteristics of the Participants ( $n = 20 \times 4$ )*

Characteristics	Musicians		Controls	
	Young	Middle-aged	Young	Middle-aged
Age in years	23.4 (3.6)	48.4 (4.8)	22.7 (2.8)	47 (4)
Education in years	15.4 (1.8)	16.5 (3.6)	15.7 (1.5)	17.1 (4)
Music training in years	11.3 (3.1)	12.6 (3.2)	—	—
Age of training onset (years)	9.2 (2.5)	8.4 (4.5)	—	—
Average practice hours per week	12.7 (11.6)	12.3 (11)	—	—
MOCA score <sup>1</sup>	28.1 (1.6)	27.8 (1.2)	28.2 (1.3)	26.8 (1.9)
Raven's APM (problems solved) <sup>2</sup>	19.7 (4.9)	16.8 (4.5)	20.5 (4.7)	13.9 (5.1)
Stroop test <sup>3</sup>				
Baseline (words/second)	2.37 (0.3)	2.39 (0.3)	2.09 (0.3)	2.14 (0.3)
Stroop (colors/second)	1.08 (0.2)	1.04 (0.2)	1.04 (0.2)	0.87 (0.1)

*Note.* Standard deviations in parentheses.

<sup>1</sup> MOCA is scored between 0 and 30. <sup>2</sup> Raven's Advanced Progressive Matrices raw scores (range: 0–36). <sup>3</sup> Number of items named per second: number of correct responses/time taken to perform the task (in seconds).

Spectrum Quotient (Baron-Cohen, Whaalwright, Skinner, Martin, & Clubley, 2001), a measure of sociocommunicative traits associated with the autistic spectrum in neurotypical adults. For both tests, we used the Portuguese versions available from the websites of the original authors. The four groups did not differ on any dimension of these measures (all  $F_s < 1$ ).

Written informed consent was obtained from all subjects, who were paid for their participation.

### Stimuli and Task

The stimuli were selected from a database on emotional prosody in Portuguese that was submitted to extensive perceptual and acoustic validation (for details, Castro & Lima, 2010). For stimuli to be included in the database, the intended emotional expression had to be correctly identified in a forced-choice task at least three times above chance level (the percentage of correct identifications was 73% on average); acoustic properties, including mean F0, F0 variability ( $SD$ ), and total duration, were significant predictors of the category membership of the stimuli in a discriminant analysis, and the emotion-specific acoustic profiles were in agreement with previous descriptions (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2003). The stimuli selected for the present study consisted of sentences with emotionally neutral semantic content (e.g., “Ela viajou de comboio.” She traveled by train; “O quadro está na parede.” The painting is on the wall) produced by two female speakers in seven emotional tones—anger, disgust, fear, happiness, sadness, surprise, and neutrality. The duration of the sentences was about 1.5 seconds each ( $M = 1480$  ms;  $SD = 243$ ). Seventy sentences were used, 10 tokens per emotional tone; half of them were produced by one speaker and the other half by the other speaker.

Participants were told that they would listen to short sentences that were neutral regarding semantic content and were asked to focus their attention on the tone of voice; the labels of the seven emotional tones were then introduced and briefly explained to ensure that they were adequately understood. Participants were instructed to perform two consecutive judgments for each sentence: a forced-choice identification of the emotional tone and an intensity judgment rating how much the expression was present in the stimulus, on a scale from 1, low intensity, to 7, high intensity. Responses were given on a seven-button response pad from Cedrus Corporation, model RB-730, attached to an Apple MacBook Pro computer running SuperLab version 4.0 (Abboud, Schultz, & Zeitlin, 2006). The emotion labels were affixed below the buttons in a fixed order, as were, above the buttons, the numbers 1 to 7 for the intensity judgments. Participants were instructed to respond as accurately and quickly as possible in the forced-choice decision; there was no time limit for the intensity judgments. RTs were measured from stimulus onset till the first button press (the categorization response). Responses were given with the index finger of the dominant hand, that had to be kept in the same position, the center of the response pad, throughout the session. Each trial consisted of the following sequence: a fixation cross for 500 ms; the auditory presentation of the sentence with a blank screen until the emotion was identified; the seven-point scale then appeared on the screen until intensity was rated; and an interstimulus interval of 3000 ms. The stimuli were pseudorandomized and divided into two blocks of 35 trials each; the presentation order

was counterbalanced across participants. There was a block of four practice trials to familiarize with the task, during which feedback was given; the practice stimuli were not included in the experimental blocks. Stimuli were presented via high-quality headphones individually adjusted to a comfortable hearing level.

## Results

### Recognition Accuracy

A response was considered correct when it matched the intended expression of the utterance. The proportion of correct identifications for each emotion, as well as the proportion of inaccurate responses and its distribution across categories, were computed individually. Table 2 presents the average proportion of correct identifications in diagonal cells (in bold) and the distribution of errors in rows, for each group. Overall accuracy was around .72 correct. Recognition rates ranged from .26, for disgust in older controls, to .92, for surprise in younger musicians. Relatively low recognition of disgust is frequently reported in speech prosody research (Banse & Scherer, 1996; Scherer, 2003).

First we examined the effects of musical expertise and age on correct identifications. Raw scores were arcsine transformed and submitted to an ANOVA with emotion as repeated-measures factor (anger, disgust, fear, happiness, sadness, surprise and neutrality), expertise (musicians and controls) and age (younger and middle-aged) as between-subjects factors. Accuracy was higher in musicians (.77) than in controls (.68), as indicated by a significant main effect of expertise [ $F(1, 76) = 9.87, p < .01, \eta_p^2 = .11$ ]. The advantage of musicians was similar for all emotions and age groups (interactions Expertise  $\times$  Emotion, Expertise  $\times$  Age, and Expertise  $\times$  Age  $\times$  Emotion  $ns, ps > .05$ ).<sup>1</sup> It was also independent of the speaker who produced the stimuli, as examined by an additional ANOVA with speaker as repeated-measures factor (main effect of speaker and interaction Speaker  $\times$  Expertise  $ns, ps > .05$ ). Concerning age differences, older participants were less accurate than younger ones, but not for all emotions [main effect of age,  $F(1, 76) = 23.98, p < .001, \eta_p^2 = .24$ ; Age  $\times$  Emotion interaction,  $F(6, 456) = 4.16, p < .001, \eta_p^2 = .05$ ]. An age-related change was found for disgust ( $p < .01$ ), fear ( $p < .01$ ), and anger ( $p = .06$ ), but not for happiness, sadness, surprise, and neutrality ( $ps > .4$ ), as revealed by post hoc Tukey’s HSD tests. This effect adds to recent evidence that emotion recognition in prosody changes with age (e.g., Mill, Allik, Realo, & Raivo, 2009; Paulmann, Pell, & Kotz, 2008). Importantly, the effects of expertise and age were not an artifact of general cognitive differences, because they were replicated when the scores on the MOCA test, Raven’s APM, and Stroop test (both baseline and incongruous

<sup>1</sup> As shown in Table 2, the recognition of sadness was not enhanced in older musicians (.79 correct) versus controls (.85). Given that the main effect was robust and that the interactions were not significant, this probably reflects measure variability. We also sought to examine whether the type of musical expertise played a role, but this was limited by the fact that musicians played a varied set of instruments (see Methods). In an exploratory fashion, we categorized musicians according to instrumental group [keyboard,  $n = 18$ ; strings,  $n = 12$ ; woodwinds,  $n = 7$ ; others,  $n = 3$  (drums, accordion, and trombone)] and computed an ANOVA on accuracy rates; the performance was similar across instrumental groups ( $p > .3$ ).



Table 2  
Distribution of Responses for Each Emotion (Horizontal Lines) as a Function of Expertise and Age

Age group/ Emotion	Distribution of responses (proportion)													
	Controls						Musicians							
	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutrality	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutrality
Young														
Anger	.69 (.25)	0.09	0.02	0.05	0.01	0.07	0.09	.84 (.16)	0.06	0.02	0.03	0.00	0.02	0.05
Disgust	0.03	.55 (.33)	0.00	0.17	0.00	0.17	0.09	0.03	.69 (.25)	0.00	0.15	0.01	0.07	0.08
Fear	0.03	0.01	.80 (.25)	0.01	0.01	0.12	0.04	0.01	0.00	.82 (.16)	0.02	0.01	0.13	0.02
Happiness	0.01	0.01	0.01	.71 (.26)	0.01	0.22	0.04	0.02	0.02	0.00	.76 (.25)	0.01	0.22	0.01
Sadness	0.01	0.02	0.05	0.00	.83 (.19)	0.00	0.10	0.02	0.00	0.04	0.00	.87 (.15)	0.00	0.08
Surprise	0.01	0.00	0.02	0.05	0.00	.85 (.14)	0.09	0.00	0.01	0.00	0.03	0.00	.92 (.13)	0.05
Neutrality	0.01	0.01	0.01	0.00	0.14	0.01	.84 (.13)	0.02	0.01	0.00	0.00	0.08	0.00	.90 (.13)
Middle-aged														
Anger	.50 (.24)	0.15	0.02	0.17	0.01	0.06	0.11	.72 (.22)	0.09	0.02	0.07	0.01	0.03	0.07
Disgust	0.08	.26 (.25)	0.01	0.36	0.01	0.21	0.09	0.09	.42 (.25)	0.01	0.23	0.00	0.22	0.04
Fear	0.05	0.06	.48 (.28)	0.03	0.05	0.28	0.07	0.02	0.04	.65 (.3)	0.04	0.03	0.19	0.05
Happiness	0.01	0.02	0.01	.62 (.19)	0.01	0.32	0.04	0.03	0.01	0.01	.67 (.15)	0.01	0.28	0.01
Sadness	0.02	0.02	0.03	0.00	.85 (.17)	0.01	0.08	0.02	0.02	0.04	0.03	.79 (.23)	0.00	0.11
Surprise	0.01	0.03	0.01	0.06	0.01	.81 (.18)	0.09	0.01	0.00	0.01	0.05	0.01	.88 (.12)	0.06
Neutrality	0.01	0.03	0.03	0.01	0.2	0.02	.72 (.27)	0.02	0.02	0.00	0.01	0.15	0.00	.81 (.11)

Note. Diagonal lines indicate correct identifications (standard deviations in parentheses).

conditions) were entered as covariates in an ANCOVA [main effect of expertise,  $F(1, 72) = 10.57, p < .01, \eta_p^2 = .13$ ; main effect of age,  $F(1, 72) = 7.16, p < .01, \eta_p^2 = .09$ ; interaction Age  $\times$  Emotion,  $F(6, 432) = 3.21, p < .01, \eta_p^2 = .04$ ; all other interactions *ns*,  $ps > .2$ ].<sup>2</sup>

Some emotions were better recognized than others [main effect of emotion,  $F(6, 456) = 37.18, p < .001, \eta_p^2 = .33$ ]: sadness was better recognized than all the other emotions except fear ( $ps < .05$ ), and disgust was the most difficult one ( $ps < .05$ ). To confirm that the recognition rates were above chance-level, we compared the obtained accuracy for each emotion and group with the accuracy expected by chance, which is .143. Pairwise *t* tests ( $df = 19$ ) revealed that, in the four groups, all emotions were identified above chance (all  $ps < .0001$ ) except disgust in older participants, where  $p = .05$ .

To get an in-depth view on the effect of musical expertise, we analyzed whether it was general across participants or rather produced by a subset of musicians. Figure 1 plots the distribution of musicians and controls across performance levels, separately for younger and older participants. The number of musicians and controls in each level was significantly different ( $\chi^2 = 14.1, df = 5, p < .05$ ). Most of the musicians clustered in the upper levels of performance (72.5% musicians vs. 40% controls above .70 correct), and most of the controls clustered below (60% controls vs. 27% musicians with less than .70 correct;  $\chi^2 = 8.56, df = 1, p < .01$ ).

The pattern of inaccurate responses for each emotion was also briefly analyzed (see Table 2, rows). Anger was misidentified mainly as disgust, happiness, or neutrality, and disgust was misidentified as happiness or surprise. Fear and happiness were both misidentified as surprise, and surprise as neutrality and happiness. Neutrality and sadness were confused with each other. This pattern of confusions has been observed in previous research with other stimuli (Adolphs, Damasio, & Tranel, 2002; Juslin & Laukka, 2001; Pell, 2002), as well as with a larger group of the stimulus set used here (Castro & Lima, 2010). It is probably associated with similarities in the emotion-specific acoustic profiles (ibid.). To examine whether the distribution of inaccurate responses varied across groups, we computed separate ANOVAs for each emotion, with inaccurate emotion category as repeated-measures factor (six categories), and expertise and age as between-subjects factors. The distribution of the misclassifications of musicians and controls was similar across emotions and age groups (main effects of expertise

<sup>2</sup> The main findings reported here were replicated in an additional analysis using a measure correcting for possible response biases, the unbiased hit rate or *Hu*, computed according to Wagner (1993). This is a measure of the joint probability that a given emotion category is correctly recognized when it is presented and that a response category is correct when it is used. The ANOVA computed on *Hu* scores showed significant main effects of expertise [ $F(1, 76) = 9.86, p < .01, \eta_p^2 = .11$ ] and age [ $F(1, 76) = 24.64, p < .001, \eta_p^2 = .24$ ], as well as the interaction Age  $\times$  Emotion [ $F(6, 456) = 4.14, p < .001, \eta_p^2 = .05$ ]. The age-related decrement was significant for anger, disgust, fear, and happiness ( $ps < .01$ ), not for sadness, surprise, and neutrality ( $ps > .6$ ). Note that differently from the main analysis on correct responses, with *Hu* the age-related decline was significant for happiness. This indicates that the stability observed with the raw scores may reflect a bias of older participants to use the response category happiness (see the pattern of inaccurate responses).

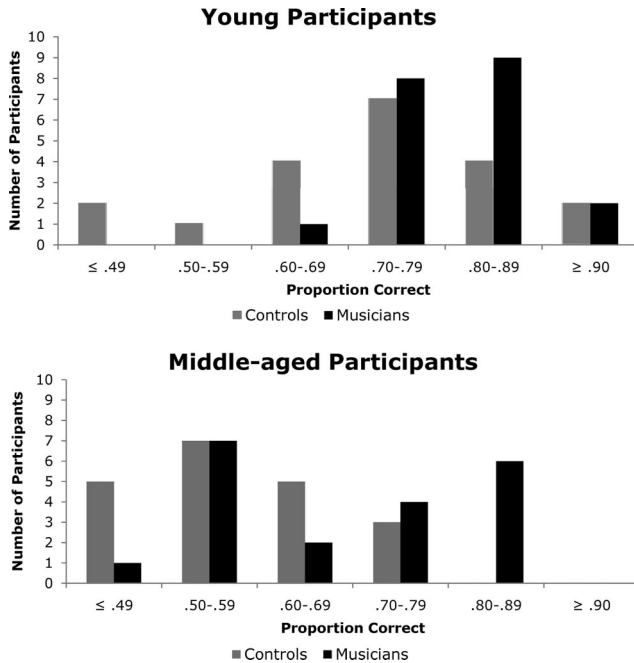


Figure 1. Distribution of young and middle-aged participants (controls and musicians) as a function of performance level.

and interactions Expertise  $\times$  Age *ns*,  $F_s < 1$ ). This confirms that musicians and controls perceived the general emotional properties of the misattributed stimuli in a similar way. The misidentifications of younger and older participants were similar for fear, sadness, and surprise (interactions Age  $\times$  Emotion *ns*,  $F_s < 1$ ), but not in the four remaining emotions. Older participants mislabeled anger as happiness more often than younger ones [interaction Age  $\times$  Emotion,  $F(5, 380) = 2.43$ ,  $p < .05$ ,  $\eta_p^2 = .03$ ; post hoc,  $p < .01$ ], and disgust more often as happiness and surprise [interaction Age  $\times$  Emotion,  $F(5, 380) = 5.05$ ,  $p < .01$ ,  $\eta_p^2 = .06$ ; post hoc,  $ps < .01$ ]. They also mislabeled happiness more often as surprise [interaction Age  $\times$  Emotion,  $F(5, 380) = 2.58$ ,  $p < .05$ ,  $\eta_p^2 = .03$ ; post hoc,  $p < .01$ ], and neutrality as sadness [interaction Age  $\times$  Emotion,  $F(5, 380) = 3.06$ ,  $p < .05$ ,  $\eta_p^2 = .04$ ; post hoc,  $p < .01$ ]. No other effects were significant ( $ps > .05$ ).

### Intensity Judgments, and Latencies

The mean intensity ratings given to correct identifications were submitted to an ANOVA with emotion as repeated-measures factor and expertise and age as between-subjects factors. Ratings were similar for both age and expertise groups (main effects of expertise and age *ns*,  $ps > .05$ ), with the sole exception that musicians judged disgust stimuli as slightly less intense than controls did [4.5 versus 5.4, respectively; interaction Expertise  $\times$  Emotion,  $F(6, 456) = 2.93$ ,  $p < .01$ ,  $\eta_p^2 = .04$ ; post hoc,  $p = .09$ ]. Stimuli were judged to express emotions with high intensity, 5.3 on average (maximum = 7): younger and older controls,  $M = 5.6$  and 5.3, respectively; corresponding values for musicians,  $M = 5$  and 5.1.<sup>3</sup> Intensity ratings differed across emotions [ $F(6, 456) = 30.31$ ,  $p < .001$ ,  $\eta_p^2 = .29$ ]: disgust was rated as less intense (4.9) than happiness (5.4) and surprise (5.5;  $ps < .05$ ). The four remaining

emotions were rated similarly: 5.3 for anger and sadness, and 5.2 for fear and neutrality ( $ps > .1$ ). The other interactions were *ns* ( $F_s < 1$ ). The global similarity between groups in intensity judgments indicates that the advantage of musicians in accuracy does not reflect general differences in the perceived emotional salience of the stimuli.

With respect to RTs, we calculated an ANOVA to verify whether the advantage of musicians was associated with longer response latencies. Errors and outliers (RTs exceeding the mean of each participant by 2 *SD*) were not included in the analysis. On average, participants took 3666 ms to respond. Older participants were slower (4079 ms) than younger ones [3253 ms;  $F(6, 456) = 30.31$ ,  $p < .001$ ,  $\eta_p^2 = .29$ ], but musicians did not differ from controls (3812 vs. 3512 ms, respectively,  $p = .1$ ). There was a main effect of emotion [ $F(6, 456) = 5.4$ ,  $p < .001$ ,  $\eta_p^2 = .07$ ]. It took longer to respond to disgust (3909 ms) than to happiness (3660 ms), sadness (3601 ms), surprise (3503 ms), and neutrality (3543;  $ps < .05$ ); RTs for anger and fear were similar (3701 and 3747 ms, respectively). None of the interactions reached significance ( $ps > .3$ ). The correlation between accuracy and latency was not significant,  $r = -0.14$ ,  $p > .05$ , confirming the absence of a speed-accuracy trade-off.

### Acoustic Cues as Predictors of Emotion Categorization

As combinations of acoustic cues predict emotion categorization performed by listeners (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2001; Sauter, Eisner, Calder, & Scott, 2010), we conducted multiple regression analyses to explore how the responses of musicians and controls were related to patterns of acoustic cues. First, each emotion portrayal was measured regarding the following acoustic parameters: F0, including mean (*M*), variation (*SD*), minimum, maximum, and perturbations (jitter); voice intensity, including *M* and *SD*; voice quality, as indexed by the proportion of high-frequency energy (cut-off 500 Hz); and temporal aspects, including total duration and proportion of pauses. These cues were taken as predictors in the analyses, with the exception of intensity *SD* and F0 maximum. Intensity *SD* was excluded because it was invariant across emotions ( $p > .1$ ), and F0 maximum because it was highly correlated with F0 mean and variation (Juslin & Laukka, 2001). The dependent variable was the number of participants who chose each emotion category (as in Banse & Scherer, 1996; e.g., for a certain stimulus, if 20 participants responded happy, the happiness score for that stimulus was 20; if none of the participants responded happy, the score was 0). Two simultaneous multiple regression analyses were calculated for each emotion, one for controls and another for musicians. Table 3 presents the main results of these analyses: beta weights and the proportion of

<sup>3</sup> Because the stimuli were judged as highly intense, a question that arises is whether the effect of expertise on accuracy would also be observed at lower levels of intensity. To approach this concern, we split the stimuli in each emotion into a more intense and a less intense group ( $n = 5$  in each) and repeated the ANOVA including intensity as repeated-measures factor. The main effect of expertise was again significant,  $F(1, 76) = 10.23$ ,  $p < .01$ ,  $\eta_p^2 = .12$ , and it did not interact with intensity,  $p > .05$ . This result indicates that the advantage of musicians is robust across intensity levels. However, future studies would benefit from an experimental manipulation of this factor.

Table 3

Results From Multiple Regression Analyses of the Listener's Utilization of Acoustic Cues (Lines) for the Categorization of Emotions (Columns), Separately for Controls and Musicians

Voice cue	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Neutrality
<b>Controls</b>							
F0 (mean)	0.26	-0.24	.83**	.42**	-.83**	0.21	-.47**
F0 (SD)	-0.19	0.3	-.76**	0.21	0.15	0.27	-0.01
F0 (minimum)	-.3**	0.03	-0.09	-.25*	0.17	.32**	0.01
Jitter	0.02	-0.05	.28**	-.2*	-0.1	-0.02	0.09
Intensity (mean)	-0.17	-0.02	0.11	-0.01	-.28**	-0.01	.4**
HF 500 Hz	.54**	0.18	-0.15	-0.04	-.21**	-0.07	-0.09
Total duration	-.38**	.3**	0.09	0.16	-0.12	-0.03	0.1
Pause proportion	-0.03	.27*	0.05	0.16	-0.09	-.26**	0.06
Adjusted R <sup>2</sup>	.42**	.18**	.41**	.39**	.49**	.32**	.30**
<b>Musicians</b>							
F0 (mean)	0.23	-0.27	.83**	.42**	-.84**	0.17	-.47**
F0 (SD)	-0.14	.33*	-.72**	0.18	0.19	0.28	-0.08
F0 (minimum)	-.28**	0.01	-0.1	-.22*	0.19	.34**	-0.01
Jitter	-0.01	-0.05	.3**	-0.17	-0.12	-0.06	0.1
Intensity (mean)	-0.16	-0.04	0.08	0.01	-.27**	-0.04	.4**
HF 500 Hz	.5**	0.09	-0.15	-0.01	-.22**	-0.12	-0.04
Total duration	-.4**	.32**	0.04	.19*	-0.13	-0.04	0.13
Pause proportion	0.01	.23*	0.06	0.09	-0.04	-.25*	0.01
Adjusted R <sup>2</sup>	.41**	.17**	.38**	.32**	.45**	.31**	.36**

Note. Values correspond to beta weights. Adjusted R<sup>2</sup> is also shown. HF = high frequency energy.

\* $p \leq .1$ . \*\* $p < .05$ .

variance explained by the acoustic measures (adjusted  $R^2$ ). All the emotion categories were significantly predicted by some combination of acoustic cues ( $ps < .05$ ). The explained variance ranged from .17 for disgust in musicians to .49 for sadness in controls. An inspection of Table 3 shows that the constellation of cues yielding significant beta weights was unique for each emotion, thus confirming that listeners rely on different acoustic profiles to categorize emotions (e.g., Juslin & Laukka, 2001; Juslin & Laukka, 2003). Concerning the impact of musical expertise, these analyses revealed two important findings: (a) the predictive strength of acoustic measures was similar in controls and musicians [the explained variance was the same across groups, .36 and .34, respectively;  $t(12) = .31, p > .7$ ], which suggests that both groups were similarly efficient and consistent in the utilization of the cues analyzed here; (b) the emotion-specific patterns of cues that predicted responses were largely common in both groups—cues reaching significant beta weights for anger, fear, sadness, surprise, and neutrality were the same in musicians and controls; for happiness, jitter was slightly more predictive in controls and duration in musicians, and for disgust F0 variation was slightly more predictive in musicians than in controls. The similarity in the relative weight of the acoustic cues that predicted categorization by musicians and controls suggests that the inference rules they used to recognize emotional prosody were not qualitatively different. Therefore, enhanced recognition accuracy in musicians probably reflects quantitative rather than qualitative differences in how emotional prosody is processed.

## Discussion

Does the perception of emotion in music and language engage shared mechanisms? We examined this question by determining whether musical expertise influenced the processing of emotional

speech prosody. A robust effect of expertise was obtained: musicians were more accurate than musically naïve listeners in recognizing emotions in speech prosody. This effect was general across the seven emotional tones tested (six universal emotions and neutrality), was observed in two age groups (young and middle-aged adults), and was widespread across the participant sample. It cannot be accounted for by socioeducational background, general intellectual level or personality characteristics, longer analysis of the stimuli, or increased responsiveness to emotional salience (intensity judgments), because these were similar in musicians and controls. Additionally, both groups perceived the general emotional properties of the stimuli similarly: they did not differ in the pattern of misclassifications, nor in the profile of acoustic cues that were predictive of their categorization responses. This suggests that the neurocognitive mechanisms involved in processing emotional prosody are responsive to extensive music training, such that musical expertise is associated with cross-domain benefits in the ability to recognize emotions in speech prosody. Another finding was an age-related change in the recognition of anger, disgust, and fear (not happiness, sadness, surprise, or neutrality). This is consistent with previous research showing that the recognition of emotions in speech prosody, particularly negative ones, changes with age decreasing with an early onset in the middle years (e.g., Mill et al., 2009; Paulmann, Pell, & Kotz, 2008).

The effect of musical expertise on emotional speech reported here clarifies previous conflicting results. The initial findings by Thompson and colleagues (2004) were replicated and extended in important ways: (a) with well-controlled stimuli that were accurately perceived by musicians and controls, and that were produced by more than one speaker, the advantage of musicians was general, not restricted to a subset of emotions; (b) the effect was obtained in a wider range of emotions, specifically in the six categories

usually tested in emotion perception research; (c) it was observed in younger as well as middle-aged musicians, thus indicating that it is long-lasting; and (d) potential confounds not analyzed previously have been excluded, namely possible differences in RTs and in perceived intensity. In contrast, the present results stand at odds with Trimmer and Cuddy's (2008) negative findings. A likely explanation for this discrepancy is the level of music training of the participants in both studies. In the present study, expert musicians had at least eight years of formal training, whereas in the other study the musically trained subjects presumably had a lower level of expertise (on average 6.5 years of training). In line with this interpretation is the fact that in other studies where positive effects of musical training were found, musicians also had quite an extended practice: 13 years in Thompson et al.'s (2004), at least 14 years in Marques et al.'s (2007), 15 years in Schön et al.'s (2004), and 16 years in Parbery-Clark et al.'s (2009). It is likely that extensive training is required to detect experience-related behavioral differences in adults, especially in domains tapping important adaptive abilities such as the recognition of emotions through voice (Hauser & McDermott, 2003; Patel, 2008a; Scherer, Johnstone, & Klasmeyer, 2003). Studies showing effects of musical expertise after short training periods, including one year or even less (e.g., Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006; Moreno et al., 2009; Schellenberg, 2004), have used neural measures with children participants, who are more plastic and presumably have a less varied exposure to life experiences. Another factor that may have played a role is that in our study the chances of observing the effect were increased because more emotions, and more stimuli per emotion, were included than in Trimmer and Cuddy's study.

From a neurocognitive perspective, the observed effect is consistent with the hypothesis that the recognition of emotions in speech prosody and in music involves shared mechanisms, and converges with other evidence that music and language interact at different levels of processing (e.g., Koelsch, Gunter, Wittfoth, & Sammler, 2005; Moreno et al., 2009; Parbery-Clark et al., 2009; Schön et al., 2010; Wong, Skoe, Russo, Dees, & Kraus, 2007). That this is the case for emotional processing is compatible with previous findings. Electrophysiological studies have shown that music training affects pitch processing in speech (Marques et al., 2007; Schön, Magne, & Besson, 2004), and pitch is one of the acoustic properties that is linked to the differentiation of emotions in both domains. The present results also fit nicely with the theoretical view put forward by Juslin and colleagues on the mechanisms subserving the emotional responses to music (Juslin et al., 2010; Juslin & Västfjäll, 2008). One of them, emotional contagion, is based on the link between speech and music: because expressive music mimics the acoustic patterns of emotional speech, it engages the neural processes devoted to vocal emotions; by contagion, the perceived expression triggers the corresponding emotional experience (*ibid.*). However, arguing that emotion processing in speech and music engages shared mechanisms does not entail two simplistic overgeneralizations: that the cross-domain overlap is total, and that no musical abilities rely on modular mechanisms. As noted by Juslin and colleagues, the richness of emotional reactions to music is not only rooted on the affinity with speech, it also involves other mechanisms that are unrelated to speech (brain stem reflexes, episodic memory, musical expectancies, visual imagery, evaluative conditioning, and rhythmic entrainment). And affective judgments for music and speech might

involve a combination of domain-general and domain-specific processes, as suggested by Ilie and Thompson (2006). As for modularity, both music and language are highly complex systems with several interacting components—some of them might be domain-specific, and some domain-general. For example, nonemotional and emotional musical abilities appear to depend on separate neuropsychological mechanisms (e.g., Peretz, Gagnon, & Bouchard, 1998).

Future studies will need to specify the stage(s) of processing where mechanisms for emotional processing are shared, or the locus of cross-domain transfer. Considering Schirmer and Kotz's (2006) model on emotional vocal processing, it might be as early as low-level auditory processes, or later in the integration of emotionally significant acoustic cues, or even later at a higher level of cognitive evaluation. Recently, Strait, Kraus, Skoe and Ashley (2009) observed that musical expertise influences subcortical neural responses to vocal signs of emotion: brainstem potentials to an infant's unhappy cry were enhanced and faster in musicians. Brainstem responses to speech and music have also been shown to be earlier and larger in musicians than in nonmusician controls (Musacchia, Sams, Skoe, & Kraus, 2007). Thus, one locus for the speech-music interaction might be at a quite early level of processing. However, it is uncertain how these low-level modulations would mediate the behavioral effect reported here, where integrative explicit categorization is involved. Globerson, Lavidor, Golan, Kishon-Rabin and Amir (2010) reported an association between pitch processing abilities and emotion categorization in prosody, but Leitman and colleagues (2010) failed to find such an association in healthy adults (though they found it in patients with schizophrenia). Importantly, in multiple regression analyses, we observed that musicians were not more consistent than controls in how they used low-level acoustic cues to categorize the stimuli. Another possibility is that the effect occurs in processing stages that involve the conceptual system of emotions. Barrett and colleagues (e.g., Barrett, 2006; Barrett, 2009; Lindquist & Barrett, 2010) propose that individuals differ in the "granularity" of their emotion concepts: persons with lower emotional granularity perceive affective states as broad and undifferentiated categories with low specificity, whereas those with higher granularity perceive more precise and differentiated emotional states. They also propose that emotional granularity can be trained, much as X-ray or wine experts have been trained to perceive subtle differences that novices are unaware of. Analogously, musicians might develop sharper (granular) emotion concepts for music because an important part of their training and professional activity as performers or teachers concerns the expression and perception of subtle modulations in musical expressiveness. Because the acoustic codes of emotions are similar in music and speech (Bowling et al., 2010; Curtis & Bharucha, 2010; Juslin & Laukka, 2003), musicians might rely on their finely grained concepts for musical emotions when perceiving emotional speech. Specifying which are the common stages of emotional processing in speech and music might also illuminate why musicians were superior to controls in the identification of disgust and surprise, emotions that are not frequently considered in music (e.g., Juslin, 2009; Juslin & Laukka, 2003; Zentner, Grandjean, & Scherer, 2008). If the locus of the effect is low-level auditory, then musicians would be expected to be more accurate irrespective of emotion category. However, enhanced recognition of supposedly nonmusical emotions is also



compatible with a higher-level locus of the effect: if musicians do develop sharply defined concepts for a specific set of emotions, this might indirectly benefit their performance on other categories by restricting the space of the classification problem (e.g., presented with disgust, it might be easier to reject happiness or fear as response candidates, thus enhancing a more finely tuned, and accurate, identification).

Two shortcomings of the current study should be discussed. The first concerns causality. Like most research on the effects of musical expertise (e.g., Bialystok & DePape, 2009; Brochard, Dufour, & Després, 2004; Gaser & Schlaug, 2003; Marques et al., 2007), ours was also based on a quasi-experimental design. Hence, we cannot definitely ascertain that the advantage of musicians stems from training; it might as well be a result of previous predispositions. However, this possibility is highly unlikely because our participants were matched in background socioeducational variables, general intellectual level and cognitive control, and personality traits. Furthermore, training studies indicate that learning music induces changes in brain structure and function (e.g., Fujioka et al., 2006; Hyde et al., 2009), and Moreno et al.'s (2009) longitudinal study provided experimental evidence of a causal relationship: 8-year-old children pseudorandomly assigned to six months of music lessons exhibited, posttraining, an advantage in processing linguistic pitch in comparison with matched children assigned to painting lessons. These results lend credence to the view that differences reported in cross-sectional designs may well be contingent upon training. This fits in well with research on expert performance suggesting that high levels of proficiency in one domain result from extended deliberate practice rather than from static initial predispositions (for reviews, see Ericsson & Lehmann, 1996; Ericsson & Towne, 2010). Another possibility worth considering regarding causality is that the musicians' enhanced processing of emotional prosody might stem from nonmusical dimensions of training (for a detailed discussion, see Schellenberg, 2006). Music training, like chess, is an out-of-school activity that is school-like in that it involves demanding "cognitive, adult-supervised activities that require serious and concerted effort on the part of the participant to acquire knowledge" (p. 465, *ibid.*). Such broad, music-unspecific, learning experiences are likely to foster cognitive abilities in general. Note however that in the present study the positive effect of musical expertise was not general: musicians had an advantage in recognizing emotional speech but performed like controls on the other measures of cognitive abilities.

The second shortcoming concerns specificity. Participants were tested for emotion recognition in prosody and we cannot exclude that musical expertise affects other communicative channels as well (e.g., facial expressions). Available evidence on the generality versus specificity of emotion recognition across communicative channels is mixed. Bänziger, Grandjean and Scherer (2009) compared the perception of emotion in voice, faces, and bodies, and a principal component analysis indicated two separate factors in emotion perception ability, one for the visual modality and another for the auditory one. Also, a dissociation between emotion perception in facial expressions and in speech prosody was reported by Adolphs, Tranel and Damasio (2001). On the other hand, Borod and colleagues (2000) reported a positive correlation between emotion recognition in prosody and in faces, though this turned out to be nonsignificant when demographic and cognitive characteris-

tics of the participants were entered as covariates. Moreover, the vocal expression of emotions may have a special phylogenetic status (e.g., Juslin & Laukka, 2003; Scherer et al., 2003). Future studies are needed to specify the extent to which musical expertise affects the recognition of emotions in different communicative channels.

In conclusion, this study provides empirical evidence in favor of a processing interplay between language and music: we establish a robust effect of musical expertise in the identification of emotional speech prosody. We believe that this is a starting point for further systematic research on the neurocognitive relations between emotion processing in speech and in music. This will have implications for both applied and fundamental psychological science. If speech and music set into action common resources, music therapy might be a useful device in language rehabilitation programs for patients who have difficulties at processing prosody (Golan, Baron-Cohen, & Hill, 2006; Pell & Leonard, 2003). From a fundamental perspective, it will contribute to better understanding how our mind/brain instantiates emotions, language and music.

## References

- Abboud, H., Schultz, W., & Zeitlin, V. (2006). SuperLab (Version 4.0) [Computer software]. San Pedro, CA: Cedrus Corporation.
- Adolphs, R., Damasio, H., & Tranel, D. (2002). Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion, 2*, 23–51.
- Adolphs, R., Tranel, D., & Damasio, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology, 15*, 396–404.
- Ashwin, E., Chapman, E., Cole, L., & Baron-Cohen, S. (2006). Impaired recognition of negative basic emotions in autism: A test of the amygdala theory. *Social Neuroscience, 1*, 349–363.
- Banase, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.
- Bänziger, T., Grandjean, D., & Scherer, K. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion, 9*, 691–704.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5–17.
- Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*, 20–46.
- Barrett, L. F. (2009). Variety if the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition and Emotion, 23*, 1284–1306.
- Besson, M., Magne, C., & Schön, D. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends in Cognitive Sciences, 6*, 405–407.
- Bialystok, E., & DePape, A-M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 565–574.
- Borod, J. C., Pick, L. H., Hall, S., Sliwinski, M., Madigan, N., Obler, L. K., . . . Tabert, M. (2000). Relationships among facial, prosodic, and lexical channels of emotional perceptual processing. *Cognition and Emotion, 14*, 193–211.
- Bowling, D. L., Gill, K., Choi, J. D., Prinz, J., & Purves, D. (2010). Major and minor music compared to excited and subdued speech. *The Journal of the Acoustical Society of America, 1*, 491–503.
- Brochard, R., Dufour, A., & Després, O. (2004). Effect of musical exper-

- tise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, *54*, 103–109.
- Carton, J. S., Kessler, E. A., & Pape, C. L. (1999). Nonverbal decoding skills and relationship well-being in adults. *Journal of Nonverbal Behavior*, *23*, 91–100.
- Castro, S. L., & Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudo-sentences for research on emotional prosody. *Behavior Research Methods*, *42*, 74–81.
- Castro, S. L., Martins, L., & Cunha, L. (2003). Neuropsychological screening with a Portuguese Stroop test [Abstract]. *The Clinical Neuropsychologist*, *17*, 104.
- Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*, 164–167.
- Curtis, M. E., & Bharucha, J. J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion*, *10*, 335–348.
- Di Martino, A., Shehzad, Z., Kelly, C., Roy, A. K., Gee, D. G., Uddin, L. Q., Gotimer, K., Klein, D. F., Castellanos, F. X., & Milham, M. P. (2009). Relationship between cingulo-insular functional connectivity and autistic traits in neurotypical adults. *American Journal of Psychiatry*, *166*, 891–899.
- Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273–305.
- Ericsson, K. A., & Towne, T. J. (2010). Expertise WIREs. *Cognitive Science*, *1*, 404–416.
- Fitch, W. T. (2006). On the biology and evolution of music. *Music Perception*, *24*, 85–88.
- Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., . . . Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology*, *19*, 573–576.
- Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical evoked fields in young children. *Brain*, *129*, 2593–2608.
- Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, *23*, 9240–9245.
- Globerson, E., Lavidor, M., Golan, O., Kishon-Rabin, L., & Amir, N. (2010, May). Psychoacoustic abilities as predictors of vocal emotion recognition. Paper presented at the Speech Prosody 5th International Conference, Chicago.
- Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *36*, 169–183.
- Gosling, S. D., Rentfrow, P. J., & Swann Jr., W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*, 504–528.
- Grandjean, D., Bänziger, T., & Scherer, K. R. (2006). Intonation as an interface between language and affect. *Progress in Brain Research*, *156*, 235–268.
- Hagen, E. A. H., Nummenmaa, L., Yu, R., Engell, A. D., Ewbank, M. P., & Calder, A. J. (2011). Autism spectrum traits in the typical population predict structure and function in the posterior superior temporal sulcus. *Cerebral Cortex*, *21*, 493–500.
- Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology*, *14*, 233–238.
- Hamel, R., & Schmittmann, V. D. (2006). The 20-minute version as a predictor of the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement*, *66*, 1039–1046.
- Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience*, *6*, 663–668.
- Hyde, K. L., Lerch, J. P., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience*, *29*, 3019–3025.
- Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*, 319–329.
- Juslin, P. N. (2009). Emotional responses to music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (pp. 131–140). New York: 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., Liljeström, S., Västfjäll, D., & Lundqvist, L. (2010). How does music evoke emotions? Exploring the underlying mechanisms. In P. N. Juslin, & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 605–642). New York: Oxford University Press.
- Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*, 559–575.
- Knösche, T. R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B., Witte, O. W., & Friederici, A. D. (2005). The perception of phrase structure in music. *Human Brain Mapping*, *24*, 259–273.
- Koelsch, S., Gunter, T., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, *17*, 1565–1577.
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*, 302–307.
- Leitman, D. I., Laukka, P., Juslin, P. N., Saccente, E., Butler, P., & Javitt, D. C. (2010). Getting the cue: Sensory contributions to auditory emotion recognition impairments in Schizophrenia. *Schizophrenia Bulletin*, *36*, 545–556.
- Lima, C. F., & Castro, S. L. (2011). Emotion recognition in music changes across the adult life span. *Cognition & Emotion*, *25*, 585–598.
- Lindquist, K. A., & Barrett, L. F. (2010). Emotional complexity. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 513–530). New York: The Guilford Press.
- Livingstone, S. T., Muhlberger, R., Brown, A. R., & Thompson, W. F. (2010). Changing musical emotion: A computational rule system for modifying score and performance. *Computer Music Journal*, *34*, 41–64.
- Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violations in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*, 1453–1463.
- Mill, A., Allik, J., Realo, A., & Valk, R. (2009). Age-related differences in emotion recognition ability: A cross-sectional study. *Emotion*, *9*, 619–630.
- Mithen, S. (2005). *The singing Neanderthals: The origins of music, language, mind and body*. London: Weidenfeld & Nicolson.
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, *19*, 712–723.
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences USA*, *104*, 15894–15898.
- Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear & Hearing*, *30*, 653–661.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*, 674–681.
- Patel, A. D. (2008a). *Music, language, and the brain*. New York: Oxford University Press.

- Patel, A. D. (2008b). A neurobiological strategy for exploring links between emotion recognition in music and speech. *Behavioral and Brain Sciences*, *31*, 589–590.
- Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). How aging affects the recognition of emotional speech. *Brain and Language*, *104*, 262–269.
- Pell, M. D. (2002). Evaluation of nonverbal emotion in face and voice: Some preliminary findings on a new battery of tests. *Brain and Cognition*, *48*, 499–504.
- Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: A role for the basal ganglia. *Cognitive, Affective, & Behavioral Neuroscience*, *3*, 275–288.
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*, 107–120.
- Peretz, I. (2008). Musical disorders: From behavior to genes. *Current Directions in Psychological Science*, *17*, 329–333.
- Peretz, I. (2010). Towards a neurobiology of musical emotions. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 99–126). Oxford: Oxford University Press.
- Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*, 688–691.
- Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*, 111–141.
- Salthouse, T. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*, 507–514.
- Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology*, *63*, 2251–2272.
- Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences, USA*, *107*, 2408–2412.
- Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*, 511–514.
- Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, *98*, 457–468.
- Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*, 235–248.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*, 227–256.
- Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. Davidson, K. R. Scherer, & H. Goldsmith (Eds.), *Handbook of the affective sciences* (pp. 433–456). New York: Oxford University Press.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341–349.
- Schön, D., Ystad, S., Kronland-Martinet, R., & Besson, M. (2010). The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience*, *22*, 1026–1035.
- Strait, D., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency - effects of training on subcortical processing of vocal expressions of emotion. *European Journal of Neuroscience*, *29*, 661–668.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*, 46–64.
- Trehub, S. E., & Hannon, E. R. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition*, *100*, 73–99.
- Trenerry, M. R., Crosson, B., Deboe, J., & Leber, W. R. (1995). *The Stroop neuropsychological screening test manual*. Tampa, FL: Psychological Assessment Resources.
- Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, *8*, 838–849.
- Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*, 720–752.
- Wagner, H. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*, 3–28.
- Wong, P., Skoe, E., Russo, N., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422.
- Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*, 494–521.

Received May 6, 2010

Revision received May 16, 2011

Accepted May 19, 2011 ■