

# The music of language: An ERP investigation of the effects of musical training on emotional prosody processing <sup>☆</sup>



Ana P. Pinheiro <sup>a,b,\*</sup>, Margarida Vasconcelos <sup>a</sup>, Marcelo Dias <sup>a</sup>, Nuno Arrais <sup>c</sup>, Óscar F. Gonçalves <sup>a,d</sup>

<sup>a</sup>Neuropsychophysiology Lab, CIPsi, School of Psychology, University of Minho, Braga, Portugal

<sup>b</sup>Cognitive Neuroscience Lab, Department of Psychiatry, Harvard Medical School, Boston, MA, USA

<sup>c</sup>Music Department, Institute of Arts and Human Sciences, University of Minho, Braga, Portugal

<sup>d</sup>Spaulding Center of Neuromodulation, Department of Physical Medicine & Rehabilitation, Spaulding Rehabilitation Hospital and Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

## ARTICLE INFO

### Article history:

Accepted 22 October 2014

Available online 21 November 2014

### Keywords:

Event-related potentials

Emotional prosody

Language

Musical training

## ABSTRACT

Recent studies have demonstrated the positive effects of musical training on the perception of vocally expressed emotion. This study investigated the effects of musical training on event-related potential (ERP) correlates of emotional prosody processing.

Fourteen musicians and fourteen control subjects listened to 228 sentences with neutral semantic content, differing in prosody (one third with neutral, one third with happy and one third with angry intonation), with intelligible semantic content (semantic content condition – SCC) and unintelligible semantic content (pure prosody condition – PPC).

Reduced P50 amplitude was found in musicians. A difference between SCC and PPC conditions was found in P50 and N100 amplitude in non-musicians only, and in P200 amplitude in musicians only. Furthermore, musicians were more accurate in recognizing angry prosody in PPC sentences.

These findings suggest that auditory expertise characterizing extensive musical training may impact different stages of vocal emotional processing.

© 2014 Elsevier Inc. All rights reserved.

## 1. Introduction

*“From a musical perspective, speech is a concert of phonemes and syllables, melodically inflected by prosody”*

[Brandt, Gebrian, & Slevc, 2012, p. 4]

The musician’s brain is considered a model of experience-driven neuroplasticity (Hoening et al., 2011; Moreno, Marques, Santos, Santos, & Besson, 2009; Münte, Altenmüller, & Jäncke, 2002). The advantages of musical training and expertise on different cognitive domains have been consistently demonstrated (e.g., Brochard, Dufour, & Despres, 2004; Kraus & Chandrasekaran, 2010; Patel, 2011; Schellenberg, 2005). Some studies indicate that this expertise might translate into enhanced language and speech perception abilities, such as emotional prosody perception. For example,

<sup>☆</sup> *Statement of Significance to the Neurobiology of Language:* This study provided event-related potential and behavioral evidence suggesting that musical training has an effect on different stages underlying the processing of emotional prosody in speech with intelligible vs. unintelligible semantic content.

\* Corresponding author at: Neuropsychophysiology Lab, CIPsi, School of Psychology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal. Fax: +351 253604658.

E-mail address: [ana.pinheiro@psi.uminho.pt](mailto:ana.pinheiro@psi.uminho.pt) (A.P. Pinheiro).

extensive musical training is associated with enhanced perceptual abilities in tasks of speech perception in adverse listening conditions (e.g., Strait, Kraus, Skoe, & Ashley, 2009; Strait, Parbery-Clark, Hittner, & Kraus, 2012) and, specifically, with differences in phoneme perception in early processing stages (e.g., Kuhn, Elmer, Meyer, & Jancke, 2013; Ott, Langer, Oechslin, Meyer, & Jancke, 2011). Furthermore, musical training is related to enhanced prosodic perception abilities, including enhanced subcortical (e.g., brainstem) encoding of language pitch patterns (e.g., Patel & Iversen, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007), enhanced pitch discrimination abilities during speech perception (Moreno et al., 2009), and facilitated recognition of vocal emotional expressions (e.g., Bhatara, Tirovolas, Duan, Levy, & Levitin, 2011; Lima & Castro, 2011; Strait et al., 2009; Thompson, Schellenberg, & Husain, 2004).

Being able to appreciate and accurately decode emotional information from speech stimuli is a critical aspect of social life (Thönnessen et al., 2010). Emotional prosody, the melody of speech, is the non-verbal vocal expression of emotion, being instantiated by a combination of acoustic parameters, including frequency-related (fundamental frequency [F0] or perceived pitch), temporal (speech rate or rhythm) and loudness (intensity)

information (e.g., Banse & Scherer, 1996; Scherer, 2003; Schirmer & Kotz, 2006). Successful spoken communication requires the integration of these multiple auditory cues, at a millisecond speed (e.g., Paulmann & Kotz, 2008; Schirmer & Kotz, 2006). Evidence on the evaluation of emotional prosody suggests that listeners are adept at extracting emotional information from speech stimuli, with performance being about four to five times above chance level (reviewed in Banse & Scherer, 1996; Scherer, 2003). Relative to individuals without musical training, musicians seem to be more accurate in the recognition of emotions in speech prosody (Lima & Castro, 2011; Thompson et al., 2004), which suggests that the perception of emotion in language and in music relies on common mechanisms.

There is strong evidence demonstrating that the perception of emotional prosody is a multi-stage process, in which different sub-stages are differentially represented in the brain (Bestelmeyer, Maurage, Rouger, Latinus, & Belin, 2014; Frühholz, Ceravolo, & Grandjean, 2012; Kotz, Meyer, & Paulmann, 2006; Kotz & Paulmann, 2007, 2011; Paulmann & Kotz, 2008; Paulmann, Seifert, & Kotz, 2010; Schirmer & Kotz, 2006; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Witteman, Van Heuven, & Schiller, 2012). Schirmer and Kotz (2006) proposed that emotional prosody processing starts with the sensory processing of the acoustic signal, followed by the detection of emotional salience in a vocal signal, and ending with the cognitive evaluation of its emotional significance. The first stage is mediated by bilateral auditory areas (e.g., Hart, Hall, & Palmer, 2003; Rauschecker, 1998), and is indexed by the N100 event-related potential (ERP) component (Paulmann & Kotz, 2008; Paulmann et al., 2010; Pinheiro et al., 2013, 2014). The auditory N100 is believed to reflect sensory and perceptual processing and to be modulated by attention (Hillyard, Hink, Schwent, & Picton, 1973; Näätänen, Gaillard, & Mäntysalo, 1978; Rosburg, Boutros, & Ford, 2008). The second stage recruits temporal areas, including the superior temporal gyrus and the anterior temporal sulcus (Grandjean et al., 2005; Kotz et al., 2003; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Morris, Scott, & Dolan, 1999; Phillips et al., 1998), and is indexed by the P200 ERP component. The last stage recruits frontal areas, including the inferior frontal gyrus and the orbitofrontal cortex (Buchanan et al., 2000; Gandour et al., 2003; Hornak, Rolls, & Wade, 1996; Schirmer, Zysset, Kotz, & Yves von Cramon, 2004; Wildgruber, Pihan, Ackermann, Erb, & Grodd, 2002; Wildgruber et al., 2004, 2005). Of note, more recent studies demonstrated emotion effects before the P200, at the P50 level (with emotional vocalizations – Liu et al., 2012; with single words – Pinheiro et al., 2014). Behavioral data may provide an indirect probe of the third stage of emotional prosody processing, in which integrative and decision processes take place (Paulmann & Kotz, 2008; Paulmann et al., 2010; Pinheiro et al., 2013, 2014). Importantly, these stages are reciprocally connected, such that sensory bottom-up processes may impact upon more integrative stages, and may additionally be modulated by top-down mechanisms (e.g., Ethofer et al., 2006; Leitman et al., 2010, 2011; Schirmer & Kotz, 2006; Witteman et al., 2012).

Given that segmental and suprasegmental cues closely interact during speech perception, it is critical to understand the influence of semantic information during the extraction of emotional information from acoustic cues (see also Dietrich, Ackermann, Szameitat, & Alter, 2006; Kotz & Paulmann, 2007). A previous ERP study analyzing the time course of emotional prosody processing (Pinheiro et al., 2013) demonstrated that neutral and emotional prosody are processed differently both when embedded in sentences with intelligible or unintelligible semantic content, and that the presence of semantic information in the speech signal influences the three stages of emotional prosody processing.

Another study (Schwartz & Pell, 2012) demonstrated that prosodic cues alone are sufficient for an accurate recognition of emotions in spoken language.

However, it is still not clear which stages of processing emotional prosody are affected by musical expertise and what the relative contribution of prosodic and semantic cues is during emotional prosody processing in musicians vs. non-musicians. The understanding of how these processes take place in the musicians' brain is particularly important, considering that the expression of emotion through music and speech relies on similar acoustic profiles, with pitch assuming a central role (e.g., Coutinho & Dibben, 2013; Escoffier, Zhong, Schirmer, & Qiu, 2013; Hausen, Torppa, Salmela, Vainio, & Särkämö, 2013; Huron, 2008; Juslin & Laukka, 2003; Murray & Arnott, 1993), and that emotional content from voice and music is similarly represented in the brain (e.g., Escoffier et al., 2013). Furthermore, previous studies suggested that musical expertise might influence the brain architecture of emotional prosody processing in language (Schön, Magne, & Besson, 2004). Considering the strong links between music and speech perception (e.g., Hausen et al., 2013), this study investigated the effect of musical training on the electrophysiological correlates of emotional prosody processing. We presented neutral, happy and angry prosodic sentences with intelligible (semantic content condition – SCC) and unintelligible (pure prosody condition – PPC) semantic content. The use of a PPC version of the task allowed to overcome the problem of a potential interaction between affective and semantic content during speech prosody processing and aimed to facilitate cross-cultural comparisons of the results (as suggested by Belin, Bestelmeyer, Latinus, & Watson, 2011).

If musical training leads to fine neural tuning to acoustic features that are important for vocal communication (Strait et al., 2009) and increases sensitivity to emotional prosodic cues (e.g., Bhatara et al., 2011; Lima & Castro, 2011; Strait et al., 2009; Thompson et al., 2004), we should observe facilitated decoding of emotional salience from acoustic cues in musicians relative to controls. In particular, based on reports of enhanced pitch perception abilities in musicians (Moreno et al., 2009; Patel & Iversen, 2007; Wong et al., 2007), we expected this facilitated processing to be more pronounced in the PPC condition, which represents a more difficult perceptive condition.

At the electrophysiological level, we expected to find group differences in the first 200 ms after sentence onset, reflecting differential sensory processing of the prosodic signal (P50 and N100) and the detection of emotional salience from prosodic cues (P200). Furthermore, we hypothesized differential modulatory effects of the sentences' semantic status on P50, N100 and P200 amplitude, dependent on musical expertise. Behaviorally, we hypothesized increased accuracy in the recognition of emotional prosody in musicians relative to non-musicians, particularly in the PPC condition.

## 2. Method

### 2.1. Participants

Fourteen musicians and fourteen healthy controls with no musical training, matched for age, handedness and socioeconomic status, participated in the experiment (see Table 1). Musicians and non-musicians did not differ in terms of education level ( $U = 65.50$ ,  $p = .076$ ) or age ( $t(26) = -.102$ ,  $p = .920$ ).

Common inclusion criteria were: (a) European Portuguese as the first language; (b) right handedness (Oldfield, 1971); (c) no history of electroconvulsive treatment; (d) no history of neurological illness; (e) no history of DSM-IV diagnosis of drug or alcohol abuse (APA, 2000); (f) no present medication for medical disorders that

**Table 1**  
Socio-demographic characteristics of musician and non-musician participants.

	Musicians (N = 14)	Non-musicians (N = 14)
Age (years)	23.00 (4.82) <sup>†</sup> Range 18–35	22.86 (2.07) <sup>†</sup> Range 21–29
Gender	9 Males	7 Males
Education Level	1 = 12; 2 = 1; 3 = 1	1 = 7; 2 = 7

Note: <sup>†</sup>M (SD) values are shown. Education level categories were defined as the following: 1 = bachelor degree, 2 = master degree, and 3 = Ph.D.

could have deleterious effects on electroencephalogram (EEG) morphology, as well as neurological, and/or cognitive functioning consequences; (g) no hearing, vision or upper body impairment – relevant for neuropsychological function; (h) no history of psychiatric disorder in oneself or in first-degree relatives; and (i) an ability and desire to cooperate with the experimental procedures, as demonstrated by given written informed consent.

Participants in the musicians group were required to have a minimum of 8 years musical training with daily instrument practice. Most of the musicians started musical practice around the age of 10 years. They had 13.36 ( $SD = 4.89$ ) years of musical practice, on average, at the time of the testing, and were assessed with a music aptitude test: the *Advanced Measures of Music Audiation* (AMMA – Gordon, 1989) (see Table 2). None of the musicians reported having absolute pitch. Participants in the control group were college students. Before participation in the study, all participants had the procedures fully explained to them and read and signed an informed consent form to confirm their willingness to participate in the study.

## 2.2. Stimuli

Stimuli were 228 sentences pseudo-randomly presented in two conditions: semantic content condition (SCC) and pure prosody condition (PPC) (see Pinheiro et al., 2013 for a detailed description of stimuli development and validation). In order to create the SCC sentences, 114 sentences with neutral semantic content were generated by an actress, with happy, angry and neutral intonation. The recordings were made in a quiet room with an Edirol R-09 recorder and a CS-15 cardioid-type stereo microphone, with a sampling rate of 22 kHz and 16-bit quantization.

All sentences had neutral semantic content (describing simple daily actions), similar syntactic structure (subject + verb + object) and length (4 words), and all started with a proper noun (50% a male noun and 50% a female noun). The emotional intonation of sentences was previously assessed in a validation study with 125 participants (Pinheiro et al., 2011). Only those sentences with an inter-rater agreement of at least 90% were selected for each prosody type (38 neutral, 38 happy and 38 angry).

Auditory stimuli were acoustically analyzed using Praat (Boersma & Weenink, 2006) software. Mean duration was 1.88 ( $SD = 0.18$ ), 2.00 ( $SD = 0.16$ ), and 1.79 ( $SD = 0.13$ ) seconds for

**Table 2**  
Characterization of musical experience and expertise of musicians.

Variable	Musicians (N = 14)	
	M (SD)	Range
Years of musical training (years)	13.36 (4.89)	8–26
Onset of musical training (years)	9.64 (3.73)	3–16
AMMA total	58.36 (6.11)	52–71
AMMA rhythm	29.58 (3.08)	25–35
AMMA tonal	28.79 (3.56)	24–36

neutral, happy and angry sentences, respectively. Mean F0 was 203.97 ( $SD = 5.11$ ), 448.01 ( $SD = 33.16$ ), and 293.44 ( $SD = 32.51$ ) Hz for neutral, happy and angry sentences, respectively. Mean intensity was 80.00 ( $SD = 2.32$ ), 77.00 ( $SD = 1.67$ ), and 77.00 ( $SD = 1.83$ ) decibels for neutral, happy and angry sentences, respectively.

The same sentences were used in the PPC. In order to assure that the sentences sounded as natural as possible, the phones of each sentence (from the list of 114 “natural” sentences) were manually segmented in Praat (Boersma & Weenink, 2006). F0 was automatically extracted in Praat at four points of each segment (20%, 40%, 60% and 80%). Occasional F0 error measurements were manually corrected. Based on the procedures of Ramus and Mehler (1999), duration and F0 values were then transferred to MBROLA (Dutoit, Pagel, Pierret, Bataille, & Van Der Vreken, 1996) for concatenative synthesis by using the European Portuguese (female) diphone database. All fricatives were replaced with the phone [s], all stop consonants with [t], all glides with [j], all stressed vowels with [æ] and all unstressed vowels with [ə], assuring that the synthesis of new sentences preserved characteristics such as global intonation, syllabic rhythm and broad phonotactics (Ramus & Mehler, 1999).

## 2.3. Procedure

Each participant was seated comfortably at a distance of 100 cm from a computer monitor in a sound-attenuating chamber, with a button box in front of them. The experimental session was divided in two blocks each composed by 114 pseudorandomized sentences. SCC sentences were presented first to all participants (Block 1), followed by PPC sentences (Block 2), since the latter represent more difficult stimuli for which no sensory memory templates exist. Stimuli were presented binaurally through headphones and were not repeated during the experiment. Stimuli presentation, timing of events and recording of subjects' responses were controlled by Presentation software (version 16.3; Neurobehavioral Systems, Inc., Albany, NY, USA). Before each experimental block, participants were given a brief training with feedback (5 sentences of each prosody type), to make sure that they understood the instructions and got familiarized with the task and with the response box.

Before each sentence onset, a fixation cross was presented centrally on the screen for 1000 ms and remained there during sentence presentation. At the end of the sentence, the cross was replaced by a question mark followed 1000 ms later by a question which appeared for 2.5 s. Fig. 1 shows the design of an experimental trial.

Participants were asked to decide if the sentence was spoken in a neutral, positive or negative intonation by pressing one of three keys (the order of the keys was counterbalanced across subjects). Each response key was marked with an emoticon in order to minimize working memory demands. A short pause was provided after 57 sentences, and no feedback was provided. The experimental session lasted approximately 45 min. ERPs were recorded while sentences were presented.

## 2.4. EEG data acquisition and analysis

The electroencephalogram (EEG) was recorded using the Quick-Amp EEG recording system (Brain Products, Munich, Germany) with 32 electrodes mounted on an elastic cap (Easy Cap), according to the 10–20 system using an average reference. Electrodes impedance was kept below 5 k $\Omega$ . The electrooculogram (EOG) was recorded from electrodes placed at the outer canthus of each eye (horizontal EOG) and from sites below and above the right eye (vertical EOG). A ground electrode was placed at Fpz. The

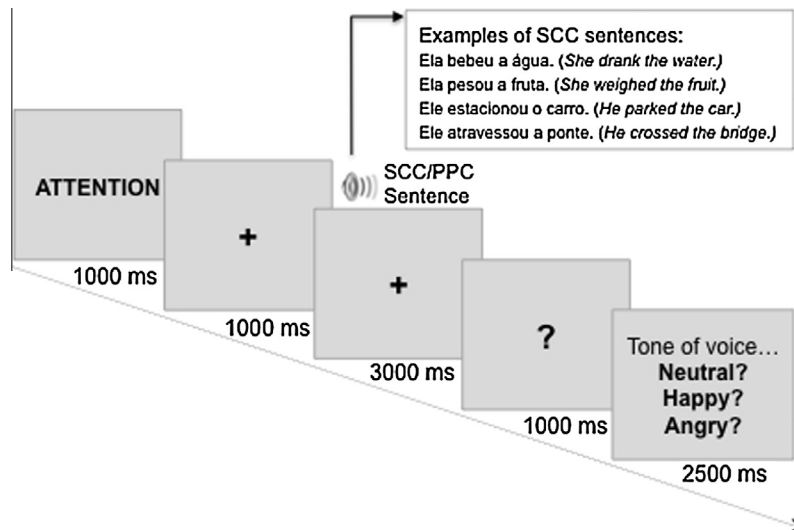


Fig. 1. Schematic illustration of an experimental trial.

EEG signal was recorded continuously and digitized at 250 Hz. Participants were asked to avoid eye and head movements during sentences' presentation.

The EEG data were processed using Brain Analyzer software (Brain Products, Inc., 2000). EEG epochs containing eye blinks or movement artifacts exceeding  $\pm 100 \mu\text{V}$  were not included in individual ERP averages. After artifact rejection, at least 80% of trials per condition per subject entered the analyses (Controls – Neutral SCC = 85.71%; Happy SCC = 84.10%; Angry SCC = 85.94%; Neutral PPC = 88.25%; Happy PPC = 87.79%; Angry PPC = 87.79%; Musicians – Neutral SCC = 81.34%; Happy SCC = 80%; Angry SCC = 81.11%; Neutral PPC = 82.88%; Happy PPC = 82.13%; Angry PPC = 83.62%). Separate ERPs for each condition were created for each participant. Averages were computed using a 200-ms prestimulus baseline and 1500 ms after the onset of the sentence, spanning the length of a sentence. This approach was adopted following all existing ERP studies of emotional prosody processing using non-spliced sentences (Paulmann, Bleichner, & Kotz, 2013; Paulmann & Kotz, 2008; Paulmann et al., 2010; Pinheiro et al., 2011, 2013).

After the inspection of grand average waveforms and considering the existing ERP literature on emotional prosody processing (Paulmann & Kotz, 2008; Paulmann et al., 2010, 2013; Pinheiro et al., 2011, 2013, 2014), three peaks were selected for analysis: P50, N100 and P200. Mean amplitude was calculated between 40 and 100 ms (P50), 130 and 220 ms (N100), 220 and 380 ms (P200) post-stimulus onset. Peak latency was calculated using the same windows. Since maximal effects were observed at fronto-central electrode sites, consistent with previous reports (e.g., Paulmann & Kotz, 2008; Pinheiro et al., 2013), P50, N100 and P200 were measured at frontal (Fz, F3/4) and central electrodes (Cz, C3/4).

## 2.5. Statistical analyses

For the statistical analysis, the IBM SPSS Statistics 20.00 (SPSS, Corp., USA) software package was used. Only significant results are presented ( $p < 0.05$ ).

### 2.5.1. ERP data

Repeated-measures analyses of variance (ANOVA) were computed for the group comparisons of P50, N100 and P200 amplitude and latency, with sentence condition (SCC, PPC), emotion (neutral, happy, angry), and electrodes (Fz, F3/4; Cz, C3/4) as within-subject

factors, and group (musicians, controls) as between-subjects factor.

### 2.5.2. Accuracy data

A repeated measures ANOVA with semantic status and emotion as within-subjects factors and group as between-subjects factor tested group differences in behavioral accuracy.

Analyses were corrected for non-sphericity using the Greenhouse–Geisser method (the original  $df$  is reported). All significance levels are two-tailed with the preset significance alpha level of  $p < 0.05$ . Main effects were followed with pairwise comparisons between conditions, using the Bonferroni adjustment for multiple comparisons.

### 2.5.3. Correlation analyses

The association between number of years/onset of musical training and ERP (P50, N100, and P200 amplitude) and behavioral measures (hit rates) was determined by calculating Pearson product-moment correlation coefficients, uncorrected for multiple comparisons.

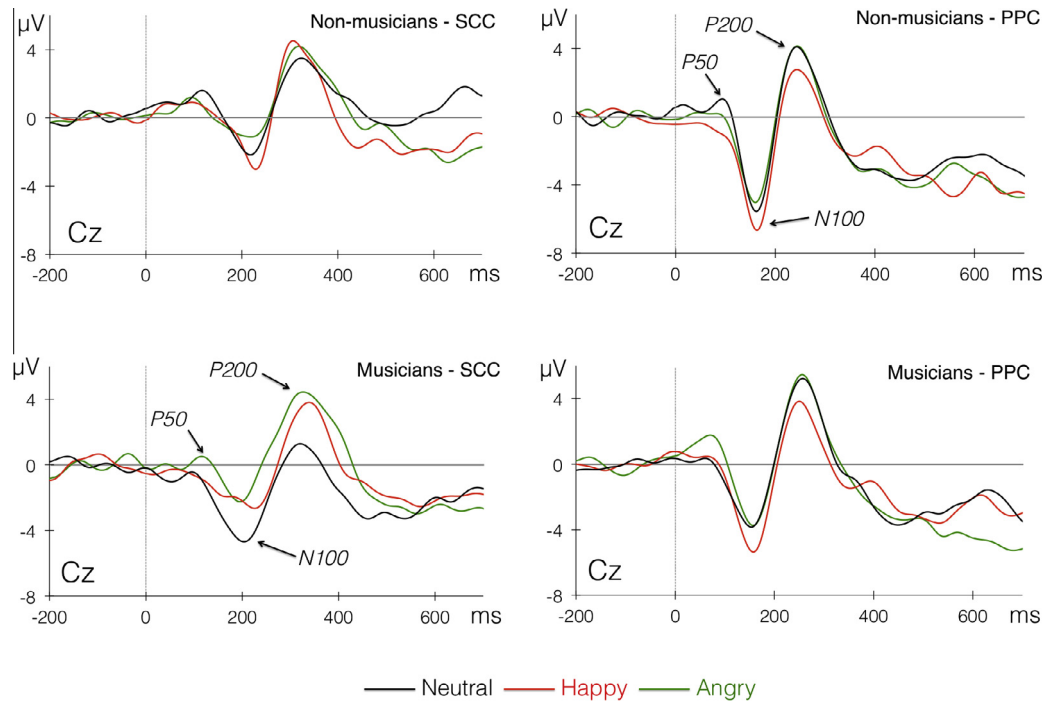
## 3. Results

### 3.1. ERP results

The analysis of ERP data revealed differences in how musicians and non-musicians processed prosodic stimuli with intelligible and unintelligible semantic content, at the level of P50, N100 and P200 (see Figs. 2 and 3, & Table 3).

#### 3.1.1. P50

**3.1.1.1. P50 amplitude.** A significant group by sentence condition interaction indicated group differences at a very early processing stage ( $F(1,26) = 7.787, p = .01$ ). P50 was more positive in controls relative to musicians in the SCC ( $p = .011$ ); also, a distinction in P50 amplitude between SCC and PPC conditions was observed in controls only: P50 was more positive in the SCC relative to PPC condition ( $p = .042$ ). This difference was only marginally significant in the musicians group ( $p = .083$ ), in which a distinct pattern was observed: P50 tended to be more positive in the PPC relative to SCC condition. No significant main effect or interactions involving the hemisphere factor were observed ( $p > .05$ ).



**Fig. 2.** Grand average waveforms at Cz for neutral, happy, and angry prosody in SCC and PPC conditions, in non-musician vs. musician groups.

**3.1.1.2. P50 latency.** A main effect of sentence condition ( $F(1,26) = 385.492, p < .001$ ) indicated an earlier P50 peak for the PPC relative to SCC condition.

### 3.1.2. N100

**3.1.2.1. N100 amplitude.** Significant main effects of sentence condition ( $F(1,26) = 19.410, p < .001$ ) and of emotion ( $F(2,52) = 4.441, p = .031$ ) were observed. N100 was more negative in the PPC relative to SCC condition ( $p < .001$ ), and in the happy relative to angry prosody condition ( $p = .005$ ).

A significant group by sentence condition interaction ( $F(1,26) = 4.270, p = .049$ ) indicated a difference in N100 amplitude between SCC and PPC conditions in controls only: N100 was more negative in the PPC relative to SCC condition ( $p < .001$ ). No difference between sentence conditions was observed in the musicians group ( $p = .110$ ). Furthermore, no significant main effect or interactions involving the hemisphere factor were observed ( $p > .05$ ).

**3.1.2.2. N100 latency.** N100 peaked earlier in the PPC relative to SCC condition (main effect of sentence condition –  $F(1,26) = 894.068, p < .001$ ).

### 3.1.3. P200

**3.1.3.1. P200 amplitude.** A significant main effect of emotion was found ( $F(2,52) = 4.989, p = .011$ ): P200 was more positive in the angry relative to happy condition ( $p = .016$ ). Emotion interacted with sentence condition ( $F(2,52) = 3.411, p = .048$ ): for neutral prosody only, P200 was more positive in the PPC relative to SCC condition; also, in the PPC condition, P200 was more positive for neutral relative to happy prosody ( $p = .027$ ) and for angry relative to happy prosody ( $p = .002$ ).

A marginally significant group by sentence condition interaction was observed ( $F(1,26) = 3.826, p = .061$ ). Nonetheless, pairwise comparisons revealed a significant difference in P200 amplitude between SCC and PPC conditions in musicians only: P200 was more positive in the PPC relative to SCC condition ( $p = .05$ ). No differences between sentence conditions were

observed in the control group ( $p > .05$ ). Also, no significant main effect or interactions involving the hemisphere factor were observed ( $p > .05$ ).

**3.1.3.2. P200 latency.** P200 peaked earlier in the PPC relative to SCC condition (main effect of sentence condition –  $F(1,26) = 1205.146, p < .001$ ).

## 3.2. Recognition accuracy

A main effect of sentence condition ( $F(1,26) = 95.594, p < .001$ ) indicated differences in prosody recognition between SCC and PPC sentences: more accurate responses were found in SCC relative to PPC sentences in both groups. A significant sentence condition by emotion interaction ( $F(2,52) = 32.547, p < .001$ ) showed differences in recognition accuracy that depended on emotion type. The analysis of emotion effects for each sentence condition type revealed the following: in the SCC condition, accuracy was increased in the recognition of angry relative to neutral prosody ( $p = .025$ ); in the PPC condition, reduced accuracy was found in the recognition of angry relative to both neutral ( $p < .001$ ) and happy ( $p = .031$ ) prosody types.

Importantly, group differences were highlighted by a significant group by sentence condition by emotion interaction ( $F(2,52) = 3.136, p = .052$ ). In spite of an overall higher number of correct responses in musicians relative to non-musicians (see Table 4), differences were only significant for the recognition of angry prosody in the PPC condition: musicians were more accurate than controls ( $p = .022$ ).

## 3.3. Correlations between ERP and behavioral measures

We found a significant association between ERP amplitude and number of years of musical training: the more extensive the musical training, the less positive the P50 amplitude in the SCC (neutral SCC prosody at F3:  $r = -.380, p = .046$ ; and angry SCC prosody at F3:  $r = -.438, p = .02$ ), and the less negative the N100 amplitude

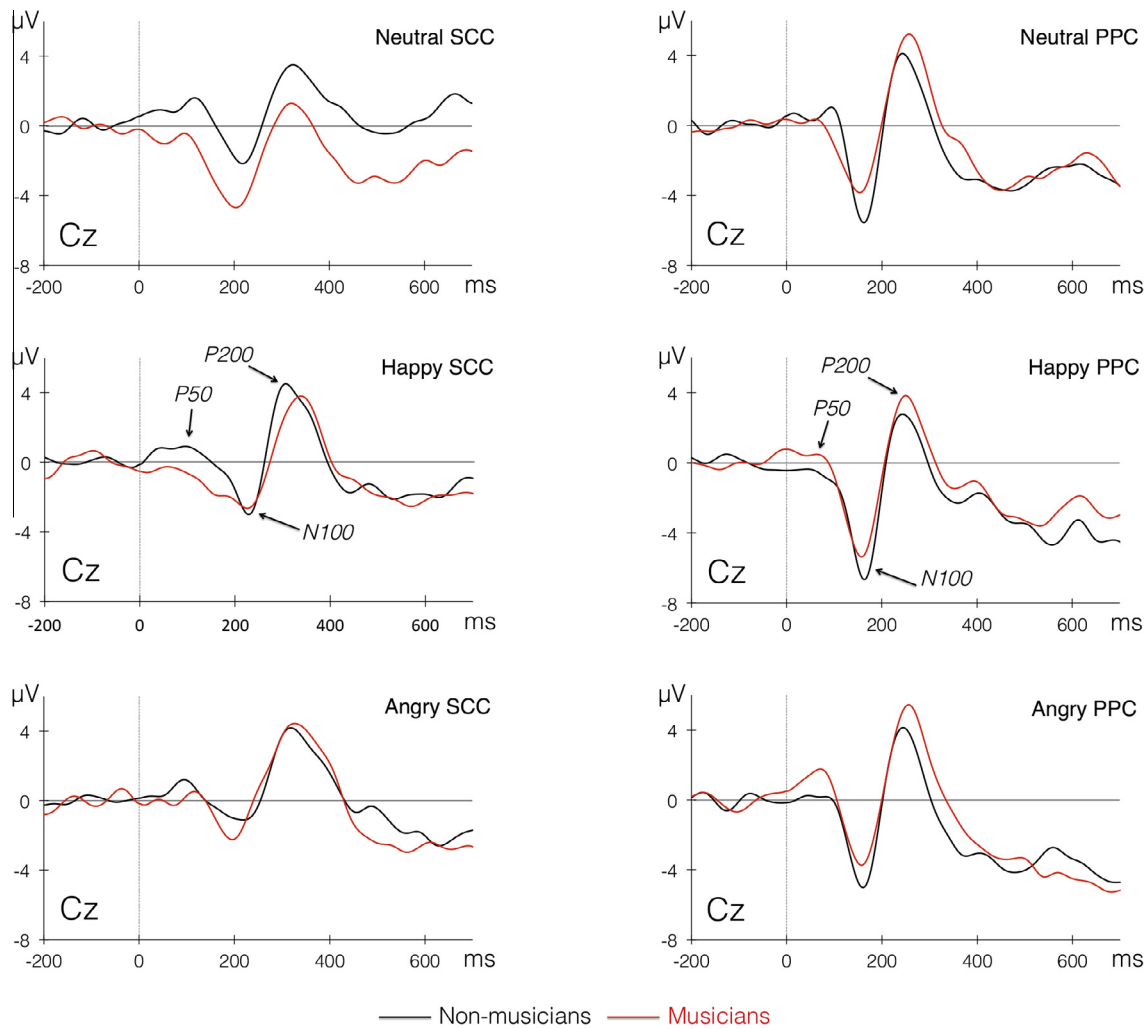


Fig. 3. Grand average waveforms at Cz showing group contrasts for neutral, happy, and angry prosody in SCC and PPC conditions.

**Table 3**  
Mean P50, N100 and P200 amplitude in the SCC and PPC conditions, in musicians and non-musicians.

Sentence condition	Group	P50 M (SD)	N100 M (SD)	P200 M (SD)
SCC	MUS	-0.152 (2.027)	-1.763 (3.043)	1.95 (3.753)
	NON-MUS	1.039 (1.896)	-0.976 (2.440)	2.297 (2.876)
PPC	MUS	0.622 (1.732)	-2.805 (2.684)	3.196 (3.419)
	NON-MUS	0.119 (2.141)	-3.858 (3.386)	1.868 (3.178)

Note: Data represent the average of all the electrodes included in the statistical analysis (Fz/3/4, Cz/3/4); MUS = musicians; NON-MUS = non-musicians.

**Table 4**  
Percentage of correct responses in the recognition of neutral, happy and angry prosody in SCC and PPC sentences, in non-musicians vs. musicians.

Sentence condition	Emotion	Group	
		Non-musicians M (SD)	Musicians M (SD)
SCC	Neutral	96.06 (1.69)	95.48 (2.59)
	Happy	95.71 (8.30)	98.06 (1.88)
	Angry	96.42 (2.67)	98.06 (3.62)
PPC	Neutral	89.26 (8.00)	92.58 (5.23)
	Happy	83.52 (14.34)	85.16 (11.33)
	Angry	72.39 (11.33)	82.58 (12.35)

in the PPC (neutral PPC prosody at F4:  $r = .379, p = .047$ ). An association between P50 amplitude and onset of musical training was also found: the later the year, the more positive the P50 amplitude

in the PPC (happy PPC at Fz:  $r = .652, p = .011$ ). No significant associations were observed for P200 amplitude.

Furthermore, the higher the number of years of musical training, the higher the number of correct responses for angry PPC prosody recognition ( $r = .451, p = .016$ ). The associations did not reach statistical significance when applying a Bonferroni-type correction.

#### 4. Discussion

The ERP and behavioral findings showed group differences that spanned three stages of emotional prosody processing (Schirmer & Kotz, 2006) and interacted with the semantic status of sentences, confirming our hypothesis. In spite of similar modulatory effects of emotion on the electrophysiological processing of sentences in

musicians and controls, qualitative group differences were observed in how the semantic status of sentences modulated P50, N100 and P200 amplitude, as revealed by significant interactions between group and sentence condition. Consistent with previous studies (Paulmann & Kotz, 2008; Pinheiro et al., 2011, 2013, 2014), ERP effects in both groups were observed within the first 200 ms after sentence onset, corroborating the sensitivity of P50, N100 and P200 components to prosodic manipulations in speech sounds (Liu et al., 2012; Paulmann & Kotz, 2008; Paulmann et al., 2010, 2013; Pinheiro et al., 2011, 2013, 2014). Furthermore, the presence of semantic information in the speech signal delayed the onset of the ERP effects, as revealed by earlier latency effects in the PPC relative to SCC sentence conditions for all components under analysis. Furthermore, while an early differentiation between SCC and PPC conditions was observed in the first 100 ms after a sentence onset (P50 and N100) in controls only, this distinction was observed later in musicians as demonstrated by P200 findings. Besides a differential amplitude pattern dependent on sentences' semantic status, the ERP results also revealed a quantitative difference for P50 amplitude only: P50 for SCC sentences was reduced in musicians relative to non-musicians. Quantitative differences were again observed at a later stage of emotional prosody processing, indexed by recognition accuracy: musicians were better than non-musicians in identifying the emotional tone of angry PPC sentences only.

Below, we discuss these findings based on a multi-stage model of emotional prosody processing (Schirmer & Kotz, 2006).

#### 4.1. Effects of musical expertise on the sensory processing of a prosodic signal (P50 and N100)

The P50 and N100 components might be considered indices of the sensory processing of prosodic cues (Schirmer & Kotz, 2006). In our study, P50 was more positive in controls relative to musicians in the SCC condition. Furthermore, while less negative P50 and N100 amplitude was found in the SCC relative to PPC condition in controls, no differences were observed in P50 or N100 amplitude between SCC and PPC conditions in the musicians group. The differences in P50 and N100 amplitude between controls and musicians suggest that musical training has an impact on the extraction of basic acoustic properties from a speech signal based on semantic status (SCC vs. PPC) but not on emotion type.

Considering the P50 component, the understanding of its functional significance is limited by the few reports of P50 effects in studies of speech processing (Lebib, Papo, de Bode, & Baudonnière, 2003; Marinkovic et al., 2003; Pinheiro et al., 2014; Sanders & Neville, 2003a, 2003b). The P50 has been primarily discussed in studies of auditory gating (e.g., Boutros, Korzyukov, Jansen, Feingold, & Bell, 2004). However, studies with simple tones shed little light on the actual underlying mechanisms of processing more complex sounds, such as speech. In comparison with simple tones, vocal stimuli require a more extensive analysis of the acoustic patterns, such as voice onset times, formant transition and pitch (e.g., Vouloumanos, Kiehl, Werker, & Liddle, 2001). There is some evidence pointing to the role of P50 as an index of the formation of sensory memory traces at the level of the primary auditory cortex (Haenschel, Vernon, Dwivedi, Gruzeliel, & Baldeweg, 2005). Its amplitude seems to be primarily modulated by the physical properties of the eliciting stimulus, such as intensity and frequency range of sounds (Chen, Ninomiya, & Onitsuka, 1997; Ninomiya, Sato, Onitsuka, Hayashida, & Tashiro, 2000). However, more recent studies highlighted the role of top-down factors on P50 amplitude, such as expectancy (White & Yee, 2006). Some studies also indicate attention effects before 100 ms post-stimulus onset, at the P50 level (Jerger, Biggins, & Fein, 1992). In that sense, the P50 has been considered an index of preattentive mechanisms, while the N100

indexes early attentive mechanisms and the P200 reflects a late attentive process (Näätänen, 1990).

Considering the existing evidence on the functional significance of the P50 ERP component, a plausible explanation for the differences in P50 amplitude between the two sentence conditions in non-musicians considers differences in the physical properties of SCC and PPC. However, the two sentence types were matched on F0, intensity, and duration, making this hypothesis less likely. However, it is plausible that the two sentence conditions differently attracted attentional resources at a very early processing stage in controls, but not in musicians. More studies are needed to support this hypothesis.

The finding of less positive P50 amplitude for SCC sentences in musicians than in non-musicians may indicate top-down effects on sensory processing of a prosodic signal, resulting from long-term musical training and use of fine-grained acoustic information (e.g., Kraus & Chandrasekaran, 2010; Parbery-Clark, Skoe, & Kraus, 2009), such as, for example, increased familiarity with specific sound parameters (e.g., pitch). Indeed, there is strong evidence indicating that “music is a resource that tunes the brain for auditory fitness” (Kraus & Chandrasekaran, 2010, p. 599). The effects of musical training on such an early stage of prosody processing were further demonstrated by the significant association between years of musical training/onset of musical training and P50 amplitude: the more extensive the training, the less positive the P50 amplitude for SCC sentences, and the earlier the training onset, the less positive the P50 amplitude for PPC sentences. This finding keeps with the notion that musicians are better at preattentively extracting information from auditory stimuli (Koelsch, Schröger, & Tervaniemi, 1999). It is also consistent with more recent studies pointing to an advantage of musicians over nonmusicians in early stages of phoneme perception, evidenced by increased responsiveness of the auditory cortex to vowels and consonant-vowel syllables (Kuhnis et al., 2013), and similar auditory encoding of voiced and unvoiced stimuli within 100 ms after stimulus onset (while voiced and unvoiced stimuli are differently processed by nonmusicians – Ott et al., 2011). Future studies should investigate whether this advantage is associated with transfer effects from musical training to speech processing (*speech-specific* advantage), or with neuroplastic changes in the auditory cortex resulting in increased sensitivity to subtle acoustic cues (*generalized* acoustic advantage) (e.g., Kuhnis, Elmer, & Jancke, 2014).

Qualitative but not quantitative group differences were also observed at the N100 level. This component is related to early auditory encoding and its amplitude is modulated by attention (increased automatic attention allocation is related to increased N100 amplitude – e.g., Rosburg et al., 2008), as well as by task difficulty (prior preparation for performing a demanding task is related to increased N100 – e.g., Näätänen & Picton, 1987). In our study, we did not observe a main effect of group on N100 amplitude. Similar N100 amplitude in musicians and non-musicians was reported by Marie, Delogu, Lampis, Belardinelli, and Besson (2011), in a study testing the discrimination of tonal and segmental variations in an unknown tonal language (Mandarin), and by Marques, Moreno, and Besson (2007) in a task of detection of pitch variations in a foreign language (European Portuguese), suggesting similar perceptual processing in the two groups. However, in our study N100 findings revealed amplitude differences between sentence conditions in controls, but not in musicians. Increased N100 to PPC relative to SCC sentences in individuals without musical training may reflect increased attention allocation to PPC sentences, plausibly related to increased perceptual/cognitive demands imposed by this sentence condition. In other words, increased N100 amplitude to PPC stimuli may be related to an increased effort to extract acoustic properties relevant to emotional prosody recognition. A negativity tendency was also

found for P50 PPC amplitude (less positive amplitude for PPC relative to SCC). The more difficult processing of PPC was further suggested by latency data, indicating later peak latency (P50, N100 and P200) in the PPC relative to SCC condition. Unintelligible speech stimuli (PPC) lack a long-term representation in the auditory memory (even though matching SCC sentences in acoustic complexity). The lower prototypicality of PPC sentences might increase perceptual demands and result in increased N100 amplitude (see also [Lattner et al., 2003](#), for a discussion of ERP differences between stimuli differing in prototypicality). Nonetheless, in musicians we found similar processing of prosodic cues with (SCC) or devoid of semantic content (PPC) at the N100 level, which might indicate similar attentional resources allocation to the two sentence conditions, plausibly related to facilitated perceptual processing resulting from auditory expertise. In a similar line, [Ott et al. \(2011\)](#) reported similar N100 for voiced and unvoiced stimuli in musicians, but not in nonmusicians (i.e., N100 amplitude was increased for voiced relative to nonvoiced stimuli). The correlational analysis supported the role of musical training on reduced N100 for PPC: the higher the number of years of musical training, the less negative the N100 amplitude. It is possible that musical expertise facilitated the processing of pitch variations in PPC sentences. Previous studies demonstrated increased sensitivity to pitch resulting from years of musical practice, which includes: increased sensitivity to pitch changes in melodic contours (e.g., [Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006](#); [Trainor, Desjardins, & Rockel, 1999](#)), superior representation of voice pitch cues (e.g., [Bidelman, Gandour, & Krishnan, 2011](#); [Wong et al., 2007](#)), or facilitated lexical pitch processing in a novel tone language (e.g., [Chandrasekaran, Krishnan, & Gandour, 2009](#); [Marie et al., 2011](#)). Furthermore, musicians seem to more effectively represent the most salient elements in sounds, which suggests more efficient sound-to-meaning relationships, or more effective extraction of regularities in an auditory signal ([Kraus & Chandrasekaran, 2010](#)). Since N100 generators are located primarily in supratemporal plane and superior temporal gyrus ([Näätänen & Picton, 1987](#)), these differences might also reflect functional and structural brain changes in the temporal cortex resulting from musical training (e.g., [Fujioka et al., 2006](#); [Pantev et al., 1998](#); [Schneider et al., 2002](#); [Trainor, Shahin, & Roberts, 2003](#)).

#### 4.2. Effects of musical expertise on the detection of the emotional salience of a prosodic signal (P200)

In the context of prosody processing, P200 was proposed to index the detection of the emotional salience of a vocal stimulus ([Paulmann & Kotz, 2008](#); [Pinheiro et al., 2013, 2014](#); [Schirmer & Kotz, 2006](#)). P200 may thus reflect a more selective response to those stimuli evaluated as significant after sensory analysis. More recent studies on emotional prosody processing points to the fact that within 200 ms after stimulus onset, emotionally relevant details such as arousal and valence are detected, and emotional category-related knowledge is inferred (e.g., [Paulmann & Kotz, 2008](#); [Paulmann et al., 2013](#)). In our study, no significant interactions involving group and emotion factors were observed, pointing again to similar modulatory effects of emotion in both groups. However, P200 amplitude differed between SCC and PPC conditions in the musicians group only, although at a marginally significant level. The difference in P200 amplitude between SCC and PPC, irrespective of prosody type, in musicians suggests that the integration of acoustic information into an emotional percept or “gestalt” occurs differently when the semantic content of the speech stimuli is intelligible (SCC) or unintelligible (PPC). Since P200 is generated in associative auditory cortical regions, including the planum temporale ([Godey, Schwartz, De Graaf, Chauvel, & Liegeois-Chauvel, 2001](#)), these findings might again point to

functional and structural brain changes in the temporal cortex resulting from musical training (e.g., [Fujioka et al., 2006](#); [Pantev et al., 1998](#); [Schneider et al., 2002](#); [Trainor et al., 2003](#)). Nonetheless, considering that the group by sentence condition interaction was only marginally significant, these results should be interpreted with caution.

Together, the ERP results suggest that the effects of musical expertise in emotional prosody processing are not necessarily identical during different stages of prosody perception. They additionally suggest that the processing of more complex vocal sounds can be largely determined by factors such as type of sound and auditory expertise of the listener, supporting a dynamic view of auditory processing (e.g., [Kraus & Chandrasekaran, 2010](#)), i.e., perceptual processing reflects a dynamic interplay between the environment (the stimulus) and the listener.

#### 4.3. Effects of musical expertise on the cognitive evaluation of a prosodic signal (behavioral data)

Behavioral data were expected to provide an indirect probe of the third stage of prosody processing, indexing processes related to the cognitive evaluation of the emotional significance of prosodic stimuli in both SCC and PPC conditions ([Schirmer & Kotz, 2006](#)). In spite of representing a later stage of emotional prosody processing, the explicit evaluation of a prosodic signal reflects the central contributions of processes occurring in earlier stages, as discussed in Section 1 (e.g., [Davis & Johnsrude, 2007](#); [Schirmer & Kotz, 2006](#)). In other words, the mainly qualitative differences observed in earlier stages related to the extraction of acoustic cues from speech stimuli and to the detection of its emotional salience may have an impact on higher-level cognitive processes involved in emotional prosody recognition (e.g., [Pinheiro et al., 2013](#); [Schirmer & Kotz, 2006](#)).

We found that, in both groups, accuracy was increased in the SSC relative to PPC condition (for all prosody types), confirming previous studies ([Kotz et al., 2003](#); [Pinheiro et al., 2013, 2014](#)). This finding suggests that understanding the semantic content of sentences influences the recognition of prosodic cues (see also [Marques et al., 2007](#) for further evidence on the interaction between semantic and prosodic processing). Importantly, accuracy rates confirmed previous studies on the recognition of vocal emotional expressions, i.e., that listeners can identify different emotions from speech above chance, even when the semantic content is unintelligible, such as in pseudowords ([Kotz et al., 2003](#); [Pinheiro et al., 2013, 2014](#)) or in a foreign language ([Pell, Monetta, Paulmann, & Kotz, 2009](#)). Nonetheless, contrary to previous studies (e.g., [Lima & Castro, 2011](#)), we did not find that musically trained individuals had an overall better performance in the recognition of emotional prosody when compared with untrained individuals. The advantage of musicians was restricted to a specific emotion type and depended on the semantic status of the utterances, i.e., their performance was better only in the recognition of angry prosody in PPC sentences. The role of musical training in the successful identification of angry prosody in more perceptually demanding conditions was reinforced by the significant association between years of musical training and accuracy for angry PPC. This finding fits well with previous evidence showing enhanced recognition of negative emotions in musicians ([Thompson et al., 2004](#)). Furthermore, previous studies demonstrated increased performance in musicians relative to non-musicians in the detection of prosodically incongruous sentence endings when the pitch variation of a final word was small and difficult to detect ([Marques et al., 2007](#); [Schön et al., 2004](#)). In our study, prosody in unintelligible speech was overall more difficult to discriminate than in normal speech. Furthermore, the relative importance of prosodic parameters, such as pitch, is likely



increased in the absence of semantic information (e.g., Kotz et al., 2003). As mentioned in Section 2, angry prosody stimuli occupied an intermediate position between neutral and happy prosody in terms of FO. It is possible that the two poles of the pitch continuum – neutral and happy prosody – were more easily recognized because they were perceptually more distinct. Therefore, the more difficult pitch discrimination in the angry prosody condition, particularly when devoid of semantic content, may explain the lower behavioral performance of non-musicians. On the contrary, the enhanced pitch discrimination abilities resulting from long-term musical training (e.g., Magne, Schön, & Besson, 2006; Schön et al., 2004; Wong et al., 2007) may have led to the better categorization of angry PPC sentences. Support for a relationship between enhanced pitch abilities and superior performance in emotional prosody recognition was recently provided by Globerson, Amir, Golan, Kishon-Rabin, and Lavidor (2013). However, we should note that the inclusion of a small number of prosody types might have obscured other differences between musicians and non-musicians.

#### 4.4. Limitations

Shortcomings of the current study include a relatively small sample size. Future studies should explore the relative contributions of specific acoustic parameters, such as pitch and duration, during emotional prosody processing in musicians relative to non-musicians. Furthermore, a wider range of emotions should be covered in future tasks (e.g., fear, disgust, sadness), for a more comprehensive understanding of the temporal course of distinct types of prosody and its relationship with musical training.

In addition, we cannot rule out that the relationship between musical experience and emotional prosody perception is reversed. Since the associations observed between ERP amplitude and musical experience did not reach significance when using a Bonferroni-type correction, future studies are needed to confirm these associations. Also, it is possible that other variables, such as emotional intelligence (Trimmer & Cuddy, 2008) or personality (Brück, Kreifelts, Kaza, Lotze, & Wildgruber, 2011), explain the observed findings. Nonetheless, an increasing number of studies suggest that the differences between musicians and non-musicians are more likely attributed to musical training *per se*, and not to a genetic predisposition for music (e.g., Besson, Chobert, & Marie, 2011). Future studies should address these possibilities.

## 5. Conclusions

Together, ERP and behavioral findings suggest that the auditory expertise underlying long-term musical training may impact both early and late stages of vocal emotional processing, particularly the extraction of sensory information from an acoustic signal (P50, N100), and the evaluation of the emotional significance of acoustic cues (recognition accuracy). They suggest that musical expertise leads to facilitated recognition of angry prosody in sentences carrying no intelligible semantic information. These results provide partial support for the hypothesis that music and language share neural resources (e.g., Hausen et al., 2013; Koelsch et al., 2002; Sluming et al., 2002; Strait et al., 2009) and for the transfer of training between music and language (e.g., Magne et al., 2006).

## Acknowledgments

This work was supported by Grants IF/00334/2012, PTDC/PSI-PCL/116626/2010 and PTDC/MHN-PCN/3606/2012 funded by *Fundação para a Ciência e a Tecnologia* (FCT, Portugal) and FEDER (*Fundo Europeu de Desenvolvimento Regional*) through the European programs QREN (*Quadro de Referência Estratégico Nacional*), and

COMPETE (*Programa Operacional Factores de Competitividade*), awarded to A.P.P.

We are grateful to all the participants of this study for their contribution to science. We are also grateful to Cristina Vieira and Ana Machado for their help with data acquisition.

## References

- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (4th ed.). Washington, DC: American Psychiatric Association.
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, 70(3), 614–636.
- Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, 102(4), 711–725.
- Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, 2(94), 1–12.
- Bestelmeyer, P. E., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to vocal expressions reveals multistep perception of auditory emotion. *The Journal of Neuroscience*, 34(24), 8098–8105.
- Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 37(3), 921–934.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, 23(2), 425–434.
- Boersma, P., & Weenink, D. (2006). *Praat: Doing phonetics by computer* (Version 4.4.12). <http://www.praat.org>.
- Boutros, N. N., Korzyukov, O., Jansen, B., Feingold, A., & Bell, M. (2004). Sensory gating deficits during the mid-latency phase of information processing in medicated schizophrenia patients. *Psychiatry Research*, 126(3), 203–215.
- Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in Psychology*, 3(327), 1–17.
- Brochard, R., Dufour, A., & Despres, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition*, 54(2), 103–109.
- Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011). Impact of personality on the cerebral processing of emotional prosody. *NeuroImage*, 58(1), 259–268.
- Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research*, 9(3), 227–238.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, 108(1), 1–9.
- Chen, C. H., Ninomiya, H., & Onitsuka, T. (1997). Influence of reference electrodes, stimulation characteristics and task paradigms on auditory P50. *Psychiatry and Clinical Neurosciences*, 51(3), 139–143.
- Coutinho, E., & Dikken, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition & Emotion*, 27(4), 658–684.
- Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, 229(1), 132–147.
- Dietrich, S., Ackermann, H., Szameitat, D. P., & Alter, K. (2006). Psychoacoustic studies on the processing of vocal interjections: How to disentangle lexical and prosodic information? *Progress in Brain Research*, 156, 295–302.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van Der Vreken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for noncommercial purposes. In *Proceedings of ICSLP'96* (Vol. 3, pp. 1393–1396).
- Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? *Human Brain Mapping*, 34(8), 1796–1810.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., et al. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *NeuroImage*, 30(2), 580–587.
- Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, 22(5), 1107–1117.
- 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(10), 2593–2608.
- Gandour, J., Wong, D., Dziedzic, M., Lowe, M., Tong, Y., & Li, X. (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping*, 18(3), 149–157.
- Globerson, E., Amir, N., Golan, O., Kishon-Rabin, L., & Lavidor, M. (2013). Psychoacoustic abilities as predictors of vocal emotion recognition. *Attention, Perception, & Psychophysics*, 75(8), 1799–1810.
- Godey, B., Schwartz, D., De Graaf, J. B., Chauvel, P., & Liegeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: A comparison of data in the same patients. *Clinical Neurophysiology*, 112(10), 1850–1859.
- Gordon, E. E. (1989). *Advanced measures of music audition*. Chicago, IL: G.I.A. Publications.

- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, 8(2), 145–146.
- Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *The Journal of Neuroscience*, 25(45), 10494–10501.
- Hart, H. C., Hall, D. A., & Palmer, A. R. (2003). The sound-level-dependent growth in the extent of fMRI activation in Heschl's gyrus is different for low- and high-frequency tones. *Hearing Research*, 179(1), 104–112.
- Hausen, M., Torppa, R., Salmela, V. R., Vainio, M., & Särkämö, T. (2013). Music and speech prosody: A common rhythm. *Frontiers in Psychology*, 4(566), 1–16.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, 182(108), 177–180.
- Hoenig, K., Müller, C., Herrnberger, B., Sim, E. J., Spitzer, M., Ehret, G., et al. (2011). Neuroplasticity of semantic representations for musical instruments in professional musicians. *NeuroImage*, 56(3), 1714–1725.
- Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia*, 34(4), 247–261.
- Huron, D. (2008). Science & music: Lost in music. *Nature*, 453(7194), 456–457.
- Jerger, K., Beggins, C., & Fein, G. (1992). P50 suppression is not affected by attentional manipulations. *Biological Psychiatry*, 31(4), 365–377.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129(5), 770.
- Koelsch, S., Gunter, T. C., Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, 17(2), 956–966.
- Koelsch, S., Schröger, E., & Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *NeuroReport*, 10(6), 1309–1313.
- Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, 86(3), 366–376.
- Kotz, S. A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: An overview and synopsis on the impact of study design. *Progress in Brain Research*, 156, 285–294.
- Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, 1151, 107–118.
- Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, 5(3), 108–125.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11(8), 599–605.
- Kuhnis, J., Elmer, S., & Jancke, L. (2014). Auditory evoked responses in musicians during passive vowel listening are modulated by functional connectivity between bilateral auditory-related brain regions. *Journal of Cognitive Neuroscience*. [http://dx.doi.org/10.1162/jocn\\_a.00674](http://dx.doi.org/10.1162/jocn_a.00674).
- Kuhnis, J., Elmer, S., Meyer, M., & Jancke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia*, 51, 1608–1618.
- Lattner, S., Maess, B., Wang, Y., Schauer, M., Alter, K., & Friederici, A. D. (2003). Dissociation of human and computer voices in the brain: Evidence for a preattentive gestalt-like perception. *Human Brain Mapping*, 20(1), 13–21.
- Lebib, R., Papo, D., de Bode, S., & Baudonnière, P. M. (2003). Evidence of a visual-to-auditory cross-modal sensory gating phenomenon as reflected by the human P50 event-related brain potential modulation. *Neuroscience Letters*, 341(3), 185–188.
- Leitman, D. I., Wolf, D. H., Laukka, P., Ragland, J. D., Valdez, J. N., Turetsky, B. I., et al. (2011). Not pitch perfect: Sensory contributions to affective communication impairment in schizophrenia. *Biological Psychiatry*, 70(7), 611–618.
- Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughhead, J., Valdez, J. N., et al. (2010). "It's not what you say, but how you say it": A reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, 4(19), 1–13.
- Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, 11(5), 1021–1031.
- Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport*, 23(2), 108–112.
- Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18(2), 199–211.
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, 23(10), 2701–2715.
- Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38(3), 487–497.
- Marques, C., Moreno, S., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 19(9), 1453–1463.
- Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, 41(10), 1410–1421.
- Moreno, S., Marques, C., Santos, A., Santos, M., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, 19(3), 712–723.
- Morris, J. S., Scott, S. K., & Dolan, R. J. (1999). Saying it with feeling: Neural responses to emotional vocalizations. *Neuropsychologia*, 37(10), 1155–1163.
- Münste, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, 3(6), 473–478.
- Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2), 1097–1108.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13(02), 201–233.
- Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42(4), 313–329.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, 24(4), 375–425.
- Ninomiya, H., Sato, E., Onitsuka, T., Hayashida, T., & Tashiro, N. (2000). Auditory P50 obtained with a repetitive stimulus paradigm shows suppression to high-intensity tones. *Psychiatry and Clinical Neurosciences*, 54(4), 493–497.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, 9(1), 97–113.
- Ott, C. G. M., Langer, N., Oechslin, M. S., Meyer, M., & Jancke, L. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Frontiers in Psychology*, 2(195), 1–10. <http://dx.doi.org/10.3389/fpsyg.2011.00195>.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392(6678), 811–814.
- Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience*, 29(45), 14100–14107.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2(142), 1–14.
- Patel, A. D., & Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, 11(9), 369–372.
- Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, 4(345), 1–10.
- Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport*, 19(2), 209–213.
- Paulmann, S., Seifert, S., & Kotz, S. A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience*, 5(1), 59–75.
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, 33(2), 107–120.
- Phillips, M. L., Young, A. W., Scott, S., Calder, A. J., Andrew, C., Giampietro, V., et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 265(1408), 1809–1817.
- Pinheiro, A. P., Del Re, E., Mezin, J., Nestor, P. G., Rauber, A., McCarley, R. W., et al. (2013). Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: An electrophysiological investigation. *Psychological Medicine*, 43(3), 603–618.
- Pinheiro, A. P., Galdo-Alvarez, S., Rauber, A., Sampaio, A., Niznikiewicz, M., & Gonçalves, O. F. (2011). Abnormal processing of emotional prosody in Williams syndrome: An event-related potentials study. *Research in Developmental Disabilities*, 32(1), 133–147.
- Pinheiro, A. P., Rezaii, N., Rauber, A., Liu, T., Nestor, P. G., McCarley, R. W., et al. (2014). Abnormalities in the processing of emotional prosody from single words in schizophrenia. *Schizophrenia Research*, 152(1), 235–241.
- Ramus, F., & Mehler, J. (1999). Language identification with suprasegmental cues: A study based on speech resynthesis. *The Journal of the Acoustical Society of America*, 105(1), 512–521.
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, 8(4), 516–521.
- Rosburg, T., Boutros, N. N., & Ford, J. M. (2008). Reduced auditory evoked potential component N100 in schizophrenia – A critical review. *Psychiatry Research*, 161(3), 259–274.
- Sanders, L. D., & Neville, H. J. (2003a). An ERP study of continuous speech processing: I. Segmentation, semantics, and syntax in native speakers. *Cognitive Brain Research*, 15(3), 228–240.
- Sanders, L. D., & Neville, H. J. (2003b). An ERP study of continuous speech processing: II. Segmentation, semantics, and syntax in non-native speakers. *Cognitive Brain Research*, 15(3), 214–227.
- Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science*, 14(6), 317–320.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, 40(1), 227–256.
- Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10(1), 24–30.
- Schirmer, A., Zysset, S., Kotz, S. A., & Yves von Cramon, D. (2004). Gender differences in the activation of inferior frontal cortex during emotional speech perception. *NeuroImage*, 21(3), 1114–1123.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, 5(7), 688–694.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341–349.

- Schwartz, R., & Pell, M. D. (2012). Emotional speech processing at the intersection of prosody and semantics. *PLoS ONE*, 7(10), e47279.
- Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage*, 17(3), 1613–1622.
- Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience promotes subcortical efficiency in processing emotional vocal sounds. *Annals of the New York Academy of Sciences*, 1169(1), 209–213.
- Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, 123(3), 191–201.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, 4(1), 46–64.
- Thönnessen, H., Boers, F., Dammers, J., Chen, Y. H., Norra, C., & Mathiak, K. (2010). Early sensory encoding of affective prosody: Neuromagnetic tomography of emotional category changes. *Neuroimage*, 50(1), 250–259.
- Trainor, L. J., Desjardins, R. N., & Rockel, C. (1999). A comparison of contour and interval processing in musicians and nonmusicians using event-related potentials. *Australian Journal of Psychology*, 51(3), 147–153.
- Trainor, L. J., Shahin, A., & Roberts, L. E. (2003). Effects of musical training on the auditory cortex in children. *Annals of the New York Academy of Sciences*, 999(1), 506–513.
- Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, 8(6), 838–849.
- Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: Event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, 13(7), 994–1005.
- White, P. M., & Yee, C. M. (2006). P50 sensitivity to physical and psychological state influences. *Psychophysiology*, 43(3), 320–328.
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, 156, 249–268.
- Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex*, 14(12), 1384–1389.
- Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., & Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. *Neuroimage*, 15(4), 856–869.
- Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage*, 24(4), 1233–1241.
- Witteman, J., Van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, 50(12), 2752–2763.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4), 420–422.