Martins, I., Lima, C. F., & Pinheiro, A. P. (2022). Enhanced salience of musical sounds in singers and instrumentalists. *Cognitive, Affective, and Behavioral Neuroscience, 22*,1044-1062. https://link.springer.com/article/10.3758/s13415-022-01007-x

FUNDING: This work was funded by national funds through the Portuguese Foundation for Science and Technology (FCT) in the scope of the project PTDC/PSI-GER/28274/2017, awarded to C.F.L., and co-funded by the European Regional Development Fund (ERDF) through the Lisbon Regional Operational Programme (LISBOA-01-0145-FEDER-028274) and the Operational Programme for Competitiveness and Internationalisation (POCI-01-0145-FEDER-028274). The work also received support from other FCT grants (PTDC/MHC-PCN/0101/2014).

Enhanced Salience of Musical Sounds in Singers and Instrumentalists

Inês Martins [a], César F. Lima [b], & Ana P. Pinheiro [a]

[a] CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal

[b] Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal

Word Count: 9083

Correspondence concerning this article should be addressed to:

Ana P. Pinheiro,

Faculdade de Psicologia, Universidade de Lisboa,

Alameda da Universidade,

1649-013 Lisboa, Portugal.

Contact: appinheiro@psicologia.ulisboa.pt

Abstract

Music training has been linked to facilitated processing of emotional sounds. However, most studies have focused on speech, and less is known about musicians' brain responses to other emotional sounds and in relation to instrument-specific experience. The current study combined behavioral and EEG methods to address two novel questions related to the perception of auditory emotional cues: whether and how long-term music training relates to a distinct emotional processing of nonverbal vocalizations and music; and whether distinct training profiles (vocal *vs.* instrumental) modulate brain responses to emotional sounds from early to late processing stages. Fifty-eight participants completed an EEG implicit emotional processing task, in which musical and vocal sounds differing in valence were presented as non-target stimuli. After this task, participants explicitly evaluated the same sounds regarding the emotion being expressed, their valence, and arousal. Compared to nonmusicians, musicians displayed enhanced salience detection (P2), attention orienting (P3), and elaborative processing (Late Positive Potential) of musical (*vs.* vocal) sounds in ERP data. The explicit evaluation of musical sounds was also distinct in musicians: accuracy in the emotional recognition of musical sounds was similar across valence types in musicians, who also judged musical sounds to be more pleasant and more arousing than nonmusicians. Specific profiles of music training (singers *vs.* instrumentalists) did not relate to differences in the processing of vocal *vs.* musical sounds. Together, these findings reveal that music has a privileged status in the auditory system of long-term musically trained listeners, irrespective of their instrument-specific experience.

Key-words: emotion; music; voice; music expertise; ERP.

1. Introduction

Emotional sounds, such as human vocalizations and music, are an integral part of our environment and impact on our behavior and communication with others (Frühholz et al., 2016). We often smile when hearing someone laughing, but we may quickly escape from a menacing growl and feel sadness when listening to Chopin's Nocturnes. The emotional properties of sounds are efficient at attracting attention and lead to enhanced perceptual analysis compared with neutral sounds (Pinheiro, Barros, Dias, et al., 2017; Pinheiro, Barros, et al., 2016). A rapid discrimination of emotional from neutral information, typically within the first 200 ms after stimulus onset (Liu et al., 2012; Sauter & Eimer, 2010; Schirmer, 2018), is essential for our understanding of the environment and intentions of others. Differences between emotional and neutral cues are detected even when emotion is task-irrelevant or outside the focus of attention (Liu et al., 2012; Paulmann & Kotz, 2008; Paulmann et al., 2012; Pinheiro, Barros, Dias, et al., 2017; Pinheiro, Barros, et al., 2016; Pinheiro, Barros, Vasconcelos, et al., 2017; Schirmer et al., 2005), and when cognitive resources are depleted by a second task (Lima et al., 2019). Emotional salience detection is therefore automatic to an important extent (i.e., relatively independent of attention).

Due to their role in communication and social interactions, voices seem to have a privileged status in the auditory system (Belin et al., 2004). Behavioral (e.g., Castiajo & Pinheiro, 2019; Lima et al., 2019), neuroimaging (e.g., Bestelmeyer, Kotz, & Belin, 2017; Giordano et al., 2021), and event-related potential (ERP) studies (e.g., Paulmann et al., 2012; Paulmann & Kotz, 2008; Pinheiro et al., 2016; Pinheiro, Barros, Vasconcelos, et al., 2017; Schirmer et al., 2005) support the notion that vocal emotional processing involves dynamic interactions between bottom-up sensory and top-down cognitive mechanisms: the sensory processing of emotionally relevant acoustic cues; the automatic detection of the emotional salience of the voice; and the cognitive evaluation of its emotional significance (Bestelmeyer et al., 2014; Schirmer & Kotz, 2006; Wildgruber et al., 2006). These processing stages are respectively reflected in ERP modulations in the N1, a fronto-centrally distributed negativity peaking around 100 ms post-stimulus onset (Näätänen & Picton, 1987); in the P2, a fronto-centrally distributed positivity peaking around 200 ms post-stimulus onset (Crowley & Colrain, 2004); and in the late positive potential (LPP), a central-posterior positivity emerging roughly after 400 ms post-stimulus onset (Moran et al., 2013), often linked to higher-order cognitive processes such as meaning evaluation or access to memory representations (Paulmann et al., 2013b). P2 amplitude modulations have been shown to reflect the earliest response to the emotional salience of a stimulus (e.g., Paulmann & Kotz, 2008; Proverbio et al., 2020), even though some studies report an earlier differentiation between neutral and emotional sounds reflected in the N1 (e.g., reduced N1 amplitude in response to emotional *vs.* neutral vocalizations – Castiajo & Pinheiro, 2021; Liu et al., 2012). P2 and LPP amplitude enhancements were observed in response to emotional compared to neutral voices (Liu et al., 2012; Proverbio et al., 2020), reflecting enhanced salience detection in emotional sounds, followed by higher-order evaluation of their emotional significance.

Music is also an important channel for auditory emotional communication. Voices and music rely on similar acoustic cues (e.g., pitch, intensity, high-frequency energy) to convey emotions (Curtis & Bharucha, 2010; Ilie & Thompson, 2006; Juslin & Laukka, 2003; Paquette et al., 2018, 2020). For instance, the acoustic features that underlie the expression of happiness, sadness, tenderness, anger, and fear are partly similar in music and voice (Juslin & Laukka, 2003). In line with this, emotional cues conveyed by musical sounds are also decoded rapidly and consistently across listeners (Bigand et al., 2005; Fritz et al., 2009). Furthermore, decoding emotional meaning from voices and music recruits a similar core neural network involving, for example, the amygdala, inferior frontal cortex, insula, and cerebellum (Frühholz et al., 2016), indicating similarities in functional processing requirements associated with distinct types of affective sounds. Given that voices and music rely on similar acoustic cues and neural resources, a testable prediction is that long-term musical experience is associated with differences in how musical and vocal emotional sounds are perceived.

Music training has been associated with advantages in sound processing, possibly as a result of structural and functional brain changes driven by long-term experience playing an instrument (Herholz & Zatorre, 2012; Lappe et al., 2008; Musacchia et al., 2007; Pantev & Herholz, 2011). However, associations between music training and socioemotional skills such as the detection and recognition of emotions remain less explored (Martins et al., 2021). Some studies revealed that musicians show facilitated processing and increased recognition accuracy of musical emotions when compared to nonmusicians (Castro & Lima, 2014; Nolden et al., 2017; Sharp et al., 2019). The positive associations between music training and emotion processing do not seem to be restricted to musical sounds, as they also extend to vocal emotions (Fuller et al., 2014; Lima & Castro, 2011; Pinheiro et al., 2015; Thompson et al., 2004). However, similar effects are not observed in the visual modality, for faces or multimodal stimuli (Correia et al., 2020; Farmer et al., 2020). The positive link between music training and vocal emotional processing is supported by evidence from electrophysiological (Pinheiro et al., 2015), fMRI (Park et al., 2015), and behavioral (Lima & Castro, 2011; Parsons et al., 2014; Thompson et al., 2004) studies. Whereas some studies reported increased recognition accuracy for a wide range of discrete emotions (e.g., anger, disgust, fear, happiness, sadness, surprise) irrespective of participants' age (Lima & Castro, 2011), others reported more selective benefits (e.g., anger – Pinheiro et al., 2015). Crucially, it remains to be clarified whether the specific domain of musical experience, namely vocal or instrumental training, affects how auditory emotions (vocal and musical sounds) are processed. Recent models of perception postulate that listening to sounds is not a passive process, simply determined by the properties of the stimulus. Instead, the prior expectations and experience of the listener change how the brain filters and interpret sounds (Denham & Winkler, 2020). Examining whether and how different forms of musicianship are associated with distinct neural and behavioral responses to emotional sounds will therefore contribute to our understanding of how long-term sensorimotor experiences shape auditory perception and, in particular, the perception of sounds that are emotionally relevant. This has further implications for everyday behavior and communication, as well as for debates on the use of music as a tool in clinical and educational contexts (Dumont et al., 2017; Grau-Sánchez et al., 2020). Furthermore, the possibility of transfer effects of music training to a central aspect of socioemotional processing, i.e., the ability to decode emotional meaning expressed by others, has received less attention compared to transfer to speech perception (e.g., Bidelman et al., 2014) and domain-general cognitive abilities (e.g., Strong & Mast, 2019).

Singers and instrumentalists develop distinct sets of motor expertise and auditory-motor repertoires (Christiner & Reiterer, 2015; Krishnan et al., 2018). Studies comparing vocal experts and instrumentalists documented differences in the structure (Halwani et al., 2011) and function (Christiner & Reiterer, 2015; Elbert et al., 1995; Kleber et al., 2010; Nikjeh et al., 2008) of sensorimotor brain regions that relate to instrument-specific sensorimotor expertise. For instance, an fMRI study comparing beatboxers and guitarists revealed increased activity in sensorimotor regions during auditory perception as a function of the music musicians can produce: when analyzing neural activity in hand and mouth motor regions, beatboxers recruited mouth areas while passively listening to beatboxing, whereas guitar players recruited hand areas while listening to guitar music (Krishnan et al., 2018). Instrument-specific differences were also observed in sensory regions such as the primary auditory cortex (Dick et al., 2011; Margulis et al., 2009; Pantev, Engelien, et al., 2001; Shahin et al., 2008), which becomes tuned to specific acoustic features (e.g., spectrotemporal) of the trained instrument (Margulis et al., 2009). Accordingly, associations between music training and auditory processing were found to be stronger for sounds related to the practiced instrument compared to other sounds (Fujioka et al., 2006; Pantev, Roberts, et al., 2001). These findings thus raise the possibility that the type of music training might be linked to differences in the processing of auditory emotions (i.e., vocalizations *vs.* instrumental sounds, respectively).

To test this hypothesis, here we examined associations between long-term music training and emotional sound processing. Most studies addressing these associations have focused on prosodic speech perception (e.g., Lima & Castro, 2011; Pinheiro et al., 2015). Much less is known about musicians' brain responses to other types of emotional sounds, which is necessary to specify the mechanisms underpinning the links between musical experience and auditory emotions, i.e., between musical and non-musical domains. We focused on both musical sounds and purely nonverbal vocalizations, such as laughter or crying, for which advantages of music training have been identified (Martins et al., 2021). Furthermore, these sound categories have generated rigorous validation work (Belin et al., 2008; Paquette et al., 2013) that allows for a stricter control of stimulus properties in experimental tasks using EEG. We first wanted to examine general associations between musicianship status (musicians *vs.* nonmusicians) and brain responses to musical and vocal emotional sounds. Second, in an exploratory analysis, we tested whether these associations depend on the type of training experience, by comparing singers with instrumentalists.

Participants performed an implicit emotion processing task while their electrophysiological responses were being measured: human nonverbal vocalizations and musical bursts expressing happiness, sadness, or no emotion (neutral) were presented, and in a minority of trials a different target sound was played instead (monkey vocalizations and tambourine sounds), which participants were instructed to detect. We chose an implicit task such that any differences between groups in response to the experimental sounds could be attributed to automatic aspects of auditory perception, rather than to motor or higher-order decisional processes. The ERP components N1, P2, and LPP were the focus of our analysis.

Additionally, after the EEG experiment, participants completed a behavioral task in which they explicitly judged the emotional properties of the sounds, including their emotional category, valence, and arousal.

From a mechanistic standpoint, the links between music and facilitated emotional sound processing have been related to acoustic similarities between musical and vocal stimuli (Juslin & Laukka, 2003). Musicians seem to be better than nonmusicians at decoding sound features such as timbre (Chartrand & Belin, 2006; Pantev, Roberts, et al., 2001), pitch (Kishon-Rabin et al., 2001; Micheyl et al., 2006; Munzer et al., 2002; Pitt, 1994; Spiegel & Watson, 1984; Tervaniemi et al., 2005), duration (Chobert et al., 2014; Jeon & Fricke, 1997; Rammsayer & Altenmüller, 2006), and rhythm (Rüsseler et al., 2001). If long-term music practice is indeed related to facilitated acoustic processing, consistent with the idea of overlapping sensory pathways for processing musical and vocal emotional information (Patel, 2014), we would expect group differences in early ERP components, namely reduced N1 amplitude in response to both emotional vocalizations and musical bursts in musicians compared to nonmusicians. Previous studies have reported reduced N1 amplitudes in response to emotionally salient compared to neutral sounds (Jessen & Kotz, 2011; Liu et al., 2012; Sauter & Eimer, 2010), as well as in response to predictable compared to unpredictable stimuli: these studies link N1 attenuation to facilitated acoustic processing.

However, music training could also relate to socio-emotional processing at a higher-order, supra-modal level of processing (Martins et al., 2021), and this could be reflected in enhanced processing of auditory sounds at later processing stages, irrespective of stimulus category (vocal *vs.* musical). If this hypothesis is true, we should observe effects of music training in late ERP components associated with the cognitive evaluation of the emotional significance of a stimulus (i.e., the LPP) and in behavioral data reflecting the explicit evaluation of the sounds. This would be reflected in enhanced LPP amplitudes to vocal and musical sounds in musicians when compared to nonmusicians, particularly regarding the stimuli with emotional significance. Such a finding would be consistent with evidence indicating that music expertise relates to distinct brain responses to vocal emotions in prefrontal areas associated with domain-general mentalizing and social processes (Park et al., 2015), which are part of a network thought to be involved in LPP generation (Liu et al., 2012). Additionally, based on previous behavioral studies (Castro & Lima, 2014; Lima & Castro, 2011), we expected increased recognition accuracy of emotions in voices and music as a function of music training, with accuracy rates being higher in musicians when compared to nonmusicians.

Another possibility is that music training is associated with enhanced effects at intermediate stages of auditory emotional processing (Schirmer & Kotz, 2006), when emotional salience is rapidly derived from relevant acoustic cues, including pitch, loudness or duration (P2). This would be reflected in enhanced P2 responses to vocal and musical sounds with an emotional quality in musicians.

We also hypothesized that different types of music training could be reflected in distinct ERP responses to vocal *vs.* musical sounds. Differences in brain adaptation have been reported in response to intense and long-term instrument-specific training (e.g., Halwani et al., 2011). For example, singers train their vocal apparatus much more extensively than instrumentalists. This is reflected in experience-dependent plasticity in the functional networks subserving voice perception, which could translate into facilitated processing of vocal sounds. Accordingly, singing-related advantages have been identified in vocal pitch regulation (Zarate & Zatorre, 2008), which is critical for the generation of vocal emotions (Juslin & Laukka, 2003), as well as in speech imitation tasks of unfamiliar utterances (Christiner & Reiterer, 2015). Notably, these advantages were not identified in instrumentalists, suggesting that singing and instrument playing can be viewed as separable musical abilities (Christiner & Reiterer, 2015). On the other hand, instrumentalists show enhanced activity in sensorimotor regions (e.g., inferior frontal gyrus) when listening to the musical instrument they can play (Krishnan et al., 2018). Because these brain regions are also recruited in auditory emotional perception (Fruhholz et al., 2016), we speculate that emotional evaluation of auditory signals could be facilitated for musical (*vs.* vocal) sounds. Therefore, we expected more pronounced ERP effects (e.g., reduced N1 or enhanced LPP) to vocal emotional sounds in singers, and to musical emotional sounds in instrumentalists (Krishnan et al., 2018).

2. Methods

2.1. Participants

Twenty-three nonmusicians (12 female) and forty-three musicians (21 female) took part in this study. Due to EEG artifacts, eight participants had to be excluded. The final sample included 58 participants. Nonmusicians were recruited in a university context (*n* = 22). Musicians were singers (*n* = 19) or instrumentalists (*n* = 17) with at least six years of formal training and regular practice in singing or playing a musical instrument, respectively[1] (see Table 1). Instrumentalists played violin (*n* = 8), guitar (*n* = 6), piano (*n* = 3), piano and guitar (*n* = 1), transverse flute (*n* = 1), bass (*n* = 1), percussion (*n* = 1), or cello (*n* = 1). Two instrumentalists reported participating in an amateur choir for the last four years. Some singers also had instrumental training of guitar (*n* = 3), piano (*n* = 3), and oboe (*n* = 1), but identified themselves primarily as singers. Instrumentalists and singers did not differ in years of music training (*p* = .414), age of training onset (*p* = .533), or musical sophistication (*p* = .462) assessed with the Goldsmiths Musical Sophistication Index (Gold-MSI; Lima et al., 2018). Groups

---

[1] Musicians were recruited from Master programs in Musical Sciences, local music schools (including *Conservatório de Música de Lisboa* and *Escola Superior de Música de Lisboa*), and local choirs (e.g., *Setúbal Voz*, *Coro Sinfónico Lisboa Cantat*, *Coro Gulbenkian*, and *Coro da Universidade de Lisboa*). Seventeen musicians (5 singers and 12 instrumentalists) were college students at the time of the recruitment.

differed in chronological age ($F(2, 63) = 7.599$, $p = .001$, $\eta_p^2 = .241$): singers were slightly older than instrumentalists ($p = .002$) and nonmusicians ($p = .008$). However, they did not differ in years of education ($F(2, 56) = 1.900$, $p = .159$, $\eta_p^2 = .063$).

All participants were native speakers of European Portuguese and reported normal hearing, as well as no history of neurological or psychiatric disorders. According to the Edinburgh Handedness Inventory (Espírito-Santo et al., 2017), 21 nonmusicians were right-handed, one was left-handed, and one was ambidextrous; thirty-eight musicians were right-handed, one was left-handed, and four were ambidextrous. Participants did not present psychopathological symptoms, assessed with the Brief Symptom Inventory (Canavarro, 1999). The mood state before the experiment, which was assessed with the Positive and Negative Affect Schedule (Galinha & Pais-Ribeiro, 2005), was comparable across groups ($p > .05$).

The study was approved by the institutional review board of Faculty of Psychology, University of Lisbon. Written informed consent was collected from all participants. Participants received a 20€ voucher or course credit for their participation in the study.

<INSERT TABLE 1 HERE>

2.2. Stimuli

The experiment included non-target and target stimuli. Non-target stimuli (the focus of the ERP analysis) consisted of 24 nonverbal emotional vocalizations and 24 musical emotional bursts. Neutral, sad, or happy (eight stimuli per condition) vocal stimuli were selected from the Montreal Affective Voices battery (MAV; Belin et al., 2008; Vasconcelos et al., 2017). Half of them were produced by female speakers and half by male speakers. Neutral, sad, or happy (eight stimuli per category) musical sounds were selected from the Musical Emotional Bursts battery (MEB; Paquette et al., 2013). Thirteen of them were played on a clarinet and 11 on a violin. Since normative data on affective ratings of the MEB stimuli are not available for Portuguese listeners, we conducted a pilot study (*N* = 24) to collect valence and arousal ratings. The ratings informed stimulus selection (see Table 2 and Table S1 in the Supplementary Material). Most importantly, we aimed to ensure that positive and negative sounds did not differ in perceived arousal, considering the effects of stimulus arousal in early and late ERP components (e.g., Paulmann, Bleichner, & Kotz, 2013). Arousal ratings differed between the three emotional conditions as determined by a one-way ANOVA ($F(2, 21) = 14.984$, $p < .001$). Bonferroni-corrected post-hoc comparisons revealed that neutral musical sounds were perceived as less arousing than happy ($p < .001$) and sad ($p = .017$) musical sounds, but happy and sad musical sounds did not differ in arousal ratings ($p = .081$). Valence ratings also differed when comparing the three emotional conditions ($F(2, 21) = 18.035$, $p < .001$). Happy musical sounds were considered more pleasant than neutral ($p < .001$) and sad ($p < .001$) musical sounds. Considering the normative arousal and valence ratings of the MAV vocalizations (Vasconcelos et al., 2017) selected for the current study, both arousal ($F(2, 21) = 78.369$, $p < .001$) and valence ($F(2, 21) = 226.798$, $p < .001$) ratings differed for the three emotion conditions. Neutral vocal sounds were perceived as less arousing than sad ($p = .015$) and happy ($p <.001$) sounds, but there were no differences in perceived arousal of happy and sad vocalizations ($p = .729$). Neutral vocal sounds were also perceived as less pleasant than happy vocalizations ($p < .001$) but more pleasant than sad vocalizations ($p < .001$), whereas happy vocalizations were perceived as more pleasant than sad ($p < .001$) vocalizations.

Target stimuli were 10 monkey vocalizations and 10 tambourine sounds (presented in a block of MAV or MEB sounds, respectively). Monkey vocalizations were downloaded from http://www.monkeymania.co.uk/, http://www.findsounds.com/ and http://www.soundbible.com/, and tambourine sounds from https://www.pond5.com, and subsequently edited into shorter segments with durations between 0.85 and 3.47 seconds.

All sounds were normalized for peak intensity (90% of maximum amplitude) using Adobe Audition 3.0 (Adobe Systems. Inc. SanJose, CA).

<INSERT TABLE 2 HERE>

2.3. Procedure

In the EEG experiment, participants sat comfortably 100 cm away from a desktop computer monitor in a sound-attenuated and electrically shielded room. Stimulus presentation was controlled with Presentation 16.3 (Neurobehavioral Systems, Inc.). Stimuli were delivered via Sennheiser CX 300-II headphones.

The auditory stimuli were distributed over two experimental blocks with 140 trials each (see Supplementary Material). The order of the blocks (one containing the MAV and the other containing the MEB sounds) was counterbalanced across participants. Within each block, stimuli were pseudo-randomized so that no more than three sounds expressing the same emotion (or played on the same instrument) were presented in consecutive trials. Furthermore, target stimuli were also not presented in consecutive trials. Each MAV or MEB sound was repeated five times, and each monkey voice and tambourine sound was repeated twice to increase signal-to-noise ratio, resulting in a total of 140 trials per block. Target sounds therefore corresponded to ≈14% of the trials. Participants were instructed to press a button each time they detected a monkey vocalization or tambourine sound using the right index finger. Participants were also instructed to center their gaze at a fixation cross shown at the center of the screen throughout the experiment. The trial structure is illustrated in Figure 1.

<INSERT FIGURE 1 HERE>

At the end of the EEG task, participants completed a behavioral task in which they evaluated the affective properties of the sounds. They listened to vocal and musical sounds again in separate blocks, and rated their valence and arousal using a 9-point Likert scale (1 = extremely unpleasant [valence] or extremely calm [arousal]; 9 = extremely pleasant [valence] or extremely aroused [arousal]). Participants also performed a three-alternative forced-choice categorization of the sounds (happy, sad, or neutral).

The stimuli used in the experimental tasks are available at https://neuralbasesofcommunication.eu/download/. The experiment was not preregistered.

2.4. EEG Data Acquisition and Analysis

EEG data were recorded continuously at a digitization rate of 512 Hz using a 64-channel BioSemi Active Two system and stored on hard disk for later analysis. Blinks and eye movements were monitored through electrodes placed over the left and right temple and one below the left eye.

EEG data were processed offline using Letswave 6, an open-source Matlab toolbox (https://www.letswave.org/). EEG data were band-pass filtered at 0.1-30 Hz filter (1601 Hamming windowed filter) and referenced offline to the average of the left and right mastoids. Epochs started 200 ms before stimulus onset and ended 1000 ms after stimulus onset for each condition: happy/sad/neutral vocal sounds, and happy/sad/neutral musical sounds. EEG was baseline corrected using a -200 to 0 ms pre-stimulus interval. Segments were screened for eye movements, muscle artifacts, electrode drifting, and amplifier blocking. The vertical electrooculogram (EOG) was derived by subtracting the activity measured at an electrode positioned below the left eye from an electrode positioned above it. The horizontal EOG was derived by subtracting the activity measured in electrodes placed at the outer canthi of the eyes. Eye movements were corrected using the method described by Gratton, Coles, and Donchin

(Gratton et al., 1983). EEG epochs with amplitudes exceeding +/-100 microvolts were rejected. After artifact rejection, at least 70% of the segments per condition per participant entered the analyses (vocal happy: *M* = 38.88, *SD* = 1.39; vocal sad: *M* = 38.66, *SD* = 1.37; vocal neutral: *M* = 38.98, *SD* = 1.54; musical happy: *M* = 38.48, *SD* = 1.96; musical sad: *M* = 38.48, *SD* = 1.87; and musical neutral: *M* = 38.78, *SD* = 1.44). Conditions did not differ in the number of non-rejected epochs (*F*(5, 342) = 2.479, *p* = .955).

The waveforms revealed three pronounced ERP peaks: a negative peak at approximately 100 ms (N1), a positive peak at approximately 200 ms (P2), and another at approximately 400 ms (P3[2]). Mean amplitude was calculated between 120-190 ms (N1), 220-290 ms (P200), and 350-450 ms (P3). We also analyzed late emotional effects (LPP) in two post-stimulus latency windows to more accurately capture differences between the conditions: 500-700 ms and 701-900 ms. Based on visual inspection and previous studies (Jessen & Kotz, 2011; Liu et al., 2012; Pinheiro, Rezaii, et al., 2016), four regions of interest (ROI) were selected: frontal (F3, Fz, F4); frontocentral (FC3, FCz, FC4), central (C3, Cz, C4), and centroparietal ROI (CP3, CPz, CP4).

Mean amplitudes of the N1, P2, P3, and LPP were analyzed with repeated-measures analyses of variance (ANOVAs) including stimulus type (vocal sounds, musical sounds), emotion (happiness, sadness, neutral), and ROI (frontal, frontocentral, central, centroparietal) as within-subjects factors, and music expertise (nonmusicians, musicians) as between-subjects factor. When significant effects of music expertise were observed, an additional exploratory ANOVA was performed with the two musician groups (instrumentalists and singers) as between-subjects factor to determine whether the expertise effect was general or modulated by type of training. As chronological age differed between instrumentalists and singers, and socioeconomic variables can represent confounding variables when testing the association

---

[2] Even though we expected to elicit a P3b in response to target sounds, the non-target sounds differed in emotional valence. Due to their relevance, they may have elicited an involuntary switching of attention or distraction from the primary task, i.e., a P3a, even if the task did not require explicit emotion categorization (Delplanque et al., 2006; Polich, 2007).

between music training and emotional processing (Martins et al., 2021), age was included in the models as a covariate. Main effects and interactions were followed-up with Bonferroni-corrected pairwise comparisons. When necessary, analyses were corrected for sphericity violations using the Greenhouse-Geisser adjustment.

2.5. Behavioral Data Analysis

Emotion recognition scores were corrected for possible response biases using unbiased hit rates or H_u (Wagner, 1993). H_u scores were obtained by computing the squared frequency of correct identifications of a given emotion category, divided by the number of stimuli in that emotion category, and multiplied by the number of times that the corresponding response category was used. H_u scores vary between 0 and 1: $H_u = 0$ when no stimulus from a given emotion is correctly recognized, and $H_u = 1$ when all the stimuli from a given emotion (e.g., happy voices) are correctly recognized and the corresponding response category (e.g., happiness) is always correctly used (i.e., there are no false alarms).

H_u scores, valence ratings, and arousal ratings, were submitted to repeated-measures ANOVAs with stimulus type (vocal sounds, musical sounds) and emotion (happiness, sadness, neutral) as within-subjects factor, and music expertise (nonmusicians, musicians) as between-subjects factor. Similar to the analysis of EEG data, when significant effects of expertise were observed, an additional ANOVA was performed with musician groups (instrumentalists and singers) as between-subjects factor.

3. Results

3.1. ERP Responses to Vocal and Musical Sounds

In the EEG experiment, there were no significant differences in button presses between experimental blocks including vocal (*M* = 96.44%, *SD* = 6.61) or musical (*M* = 96.74%, *SD* =

11.08) sounds (*t*(130) = -.191, *p* = .849), or between musicians (*M* = 97.03%, *SD* = 1.05) and nonmusicians (*M* = 95.76%, *SD* = 1.13; *t*(130) = -.766, *p* = .445). These high accuracy rates indicate that participants were paying attention to the target sounds while their brain responses were being recorded. Groups did not differ in the P3 response to target sounds in both experimental blocks (see Supplementary Material).

<INSERT FIGURES 2 AND 3 HERE>

Grand average waveforms in musicians and nonmusicians showing ERP responses as a function of emotion and stimulus type are shown in Figures 2 and 3. Figure 4 shows the spatial distribution of the ERP effects. In the next sections, we first report effects that were independent of music expertise (*general effects*) for each ERP component indexing a distinct stage of auditory emotional perception. We then report the findings on the comparison between musicians and nonmusicians (*group effects*). The ERP findings are intended to specify the temporal dynamics of operations leading to the extraction of emotional meaning from vocal *vs.* musical stimuli processed under implicit task instructions.

<INSERT FIGURE 4 HERE>

3.1.1. The sensory processing of vocal and musical sounds: N1

General effects: A main-effect of stimulus type ($F(1, 56) = 38.734$, $p < .001$, $\eta_p^2 = .409$) and emotion ($F(2, 112) = 5.468$, $p = .005$, $\eta_p^2 = .089$) were observed, as well as an interaction between the two factors ($F(2, 112) = 20.152$, $p < .001$, $\eta_p^2 = .265$). The N1 was more negative in response to musical compared to vocal sounds but only when they had emotional valence (happy – $p < .001$; sad – $p < .001$; neutral – $p = .760$; Table 3). When comparing emotion effects for each stimulus type separately, we observed that the N1 was more negative in response to neutral compared to emotional (happy – *p* < .001; sad – *p* < .001) vocalizations in line with previous studies (e.g., Liu et al., 2012), whereas it was more negative in response to sad compared to neutral musical sounds (*p* = .003).

An interaction between stimulus type and ROI (*F*(3, 168) = 14.854, *p* < .001, η_p^2 = .210) confirmed the frontocentral distribution of the N1 effects: the N1 was larger over central electrodes for both vocal and musical sounds compared to frontal (vocal and musical: *p* < .001), frontocentral (vocal: *p* < .001; musical: *p* = .003), and centroparietal electrodes (musical: *p* < .001; but not vocal: *p* = .072); the N1 was also larger over frontocentral compared to frontal electrode sites (vocal and musical: *p* < .001).

Group effects: The effect of group was not significant (*F*(1, 56) = .105, *p* = .747). Furthermore, no significant interactions involving the group factor were observed (stimulus type x group: *p* = .641; emotion x group: *p* = .245; stimulus type x emotion x group: *p* = .194).

The analysis of N1 amplitude modulations allowed us to test whether music training is associated with facilitated acoustic processing of auditory emotions. This hypothesis was not confirmed.

<INSERT TABLE 3 HERE>

3.1.2. Emotional salience detection in vocal and musical sounds: P2

General effects: The analysis yielded a main effect of stimulus type (*F*(1, 56) = 7.123, *p* = .010, η_p^2 = .113) and emotion (*F*(2, 112) = 8.379, *p* < .001, η_p^2 = 130), and an interaction between the two factors (*F*(2, 112) = 4.905, *p* = .009, η_p^2 = .081). Vocalizations elicited a more positive P2 than musical sounds when they had a happy (*p* < .001) or neutral (*p* = .035) quality, but the two sound types did not differ in the case of sadness (*p* = .828). The P2 was enhanced in response to happy compared to both sad (p = .001) and neutral (p = .001) vocalizations, as well as in response to sad compared to neutral musical sounds (p = .007; Table 4).

An interaction between stimulus type and ROI (F(3, 168) = 12.867, p < .001, η_p^2 = .187) revealed that the P2 effects were largest over frontocentral electrodes for both stimulus types compared to frontal (vocal and musical: p < .001), central (vocal: p < .001; musical: p = .022), and centroparietal electrodes (vocal and musical: p < .001). The P2 effects were less prominent over centroparietal compared to frontal (vocal and musical: p < .001), frontocentral (vocal: p < .001; musical p = .022), and central (vocal and musical: p = .001) electrode sites.

Group effects: The effect of group did not reach statistical significance (F(1, 56) = 2.832, p = .098). Nonetheless, a significant interaction between group and stimulus type (F(1, 56) = 8.436, p = .005, η_p^2 = .131) showed a P2 amplitude enhancement in response to musical sounds in musicians when compared to nonmusicians (p = .012), but not in response to vocal sounds (p = .624). Furthermore, nonmusicians showed more positive P2 amplitudes in response to vocal compared to musical sounds (p = .001), whereas in the case of musicians the P2 did not distinguish between the two sound categories (p = .849). A significant interaction between group, stimulus type, and ROI (F(3, 168) = 6.959, p < .001, η_p^2 = .111) indicated that the P2 amplitude enhancement in response to musical sounds in musicians was consistent across the four ROIs (frontal: p = .021; frontocentral: p = .011; central: p = .008; centroparietal: p = .015), as well as the P2 amplitude enhancement in response to vocal sounds in nonmusicians (frontal: p < .001; frontocentral: p < .001; central: p = .001; centroparietal: p = .020). Follow-up separate repeated-measures ANOVA comparing instrumentalists and singers did not yield a main effect of type of music training (F(1, 33) = .148, p = .703) or significant interactions involving this factor (stimulus type x type of music training: p = .113; emotion x type of music training: p = .928; stimulus type x emotion x type of music training: p = .564).

The analysis of P2 modulations examined whether music training links to early facilitated processing of motivationally or emotionally relevant auditory stimuli. Our hypothesis was partially confirmed: compared to nonmusicians, musicians showed facilitated detection of emotional salience in music, but not in vocalizations.

<INSERT TABLE 4 HERE>

3.1.3. Attention orienting to vocal and musical sounds: P3

General effects: Main effects of stimulus type (*F*(1, 56) = 10.658, *p* = .002, η_p^2 = .160) and emotion (*F*(2, 112) = 3.479, *p* = .034 η_p^2 = .058) were observed: vocalizations elicited a more positive P3 than musical sounds (*p* = . 002) and happy sounds tended to elicit enhanced P3 amplitudes relative to neutral sounds (*p* = .060). An interaction between stimulus type and ROI (*F*(3, 168) = 5.357, *p* = .002 , η_p^2 = .087) revealed that that the P3 was enhanced for vocal compared to musical sounds across all ROIs (frontal: *p* = .005; frontocentral: *p* = .001; central: *p* = .001; centroparietal: *p* = .017).

Group effects: The effect of group was not significant (*F*(1, 56) = 1.045, *p* = .311). However, a significant interaction between group and stimulus type (*F*(1, 56) = 4.164, *p* = .046, η_p^2 = .069) was observed: a more positive P3 was elicited by vocalizations compared to musical sounds (*p* .< 001) in nonmusicians, whereas the P3 response did not distinguish between the two stimulus types in musicians (*p* = .325). Follow-up separate repeated-measures ANOVA did not show differences between instrumentalists and singers (*F*(1, 34) = .012, *p* = .914). A significant interaction between stimulus type, group, and ROI (*F*(3, 168) = 3.366, *p* = .020, η_p^2 = .057) revealed that the enhanced P3 for vocal compared to musical sounds in the control group was consistent across the four ROIs (frontal: *p* = .003; frontocentral: *p* < .001; central: *p* = .001;

centroparietal: p = .020). There were no other significant interactions involving the group factor (emotion x group: p = .770; stimulus type x emotion x group: p = .313).

The P3 analysis shed light on attention orienting to sounds as a function of their emotional salience. We observed that attention orienting differed as a function of music expertise: in musicians it was similar for both vocal and musical sounds, whereas in nonmusicians it was enhanced for vocal sounds only.

3.1.4. The cognitive evaluation of the emotional significance of vocal and musical sounds: Late Positive Potential

3.1.4.1. LPP: 500-700 ms

General effects: Significant main effects of stimulus type (F(1, 56) = 73.402, p < .001, η_p^2 = .567) and emotion (F(2, 112) = 6.669, p = .002, η_p^2 = .106) were observed, as well as an interaction between the two factors (F(2, 112) = 3.389, p =.037, η_p^2 = .057). More positive amplitudes were observed in response to vocal relative to musical sounds irrespective of their emotional quality (p < .001). Furthermore, amplitudes were more positive in response to both happy (p < .001) and sad (p = .007) vocalizations compared to their neutral counterparts, whereas the LPP did not distinguish between emotional conditions in the case of musical sounds (p > .362). A significant interaction between stimulus type and ROI (F(3, 168) = 18.274, p < .001, η_p^2 = .246) revealed that for both types of stimuli, the LPP was largest over centroparietal electrode sites compared to frontal (vocal and musical: p < .001), frontocentral (vocal and musical: p < .001), and central electrodes (vocal and musical: p < .001). The LPP was also increased over central electrode sites compared to frontal and frontocental sites (vocal and musical: p < .001).

Group effects: The main effect of group did not reach statistical significance (F(1, 56) = 2.415, p = .126). Nonetheless, a significant group by stimulus type interaction (F(1, 56) = 5.482,

p = .023, η_p^2 = .089) indicated an amplitude enhancement (i.e., more positive) for musical sounds in musicians compared to nonmusicians (p = .039; Table 5). However, the two groups did not differ in the ERP response to vocal sounds (p = .471). A significant interaction between group, stimulus type, and ROI (F(3, 168) = 4.779, p = .003, η_p^2 = .079) showed that the LPP was enhanced in response to musical sounds in musicians compared to nonmusicians over frontal (p = .045), frontocentral (p = .032), and central (p = .035) regions (the difference was only marginally significant over centroparietal sites – p = .075). Follow-up separate repeated-measures ANOVA did not show differences between instrumentalists and singers (F(1, 33) = 1.031, p = .317).

<INSERT TABLE 5 HERE>

3.1.4.2. LPP: 701-900 ms

General effects: The ANOVA yielded significant effects of stimulus type (F(1, 56) = 49.871, $p < .001$, η_p^2 = .471), emotion (F(2, 112) = 5.818, p = .004, η_p^2 = .471), and an interaction between the two factors (F(2, 112) = 6.641, p = .002, η_p^2 = .106). More positive amplitudes were observed in response to both happy ($p < .001$) and sad (p = .001) compared to neutral vocalizations (Table 6), but the LPP did not distinguish between emotional conditions in the case of musical sounds (all *ps* > .220). A significant interaction between stimulus type and ROI (F(3, 168) = 19.790, $p < .001$, η_p^2 = .261) confirmed the centroparietal distribution of the LPP effects: for both vocal and musical sounds, the LPP was enhanced over centroparietal electrode sites compared to frontal (vocal and musical: $p < .001$), frontocentral (vocal and musical: $p < .001$), and central (vocal and musical: $p < .001$) sites, and enhanced over central sites compared to both frontal (vocal and musical: $p < .001$) and frontocentral (vocal and musical: $p < .001$) ones.

Group effects: No main effect of group was observed ($F(1, 56) = .1.690$, $p = .199$) or significant interactions involving the group factor (stimulus type x group: $p = .115$; emotion x group: $p = .894$; stimulus type x emotion x group: $p = .568$).

The analysis of LPP modulations clarified whether later sustained stimulus evaluation processes are modulated by music training, in line with the hypothesis that music expertise relates to facilitated socio-emotional processing at a higher-order level of processing. We partially confirmed our hypothesis by observing that the early phase of the LPP was enhanced for musical sounds in musicians compared to nonmusicians, i.e., this stimulus category led to enhanced sustained emotional evaluation.

<INSERT TABLE 6 HERE>

3.2. Behavioral Evaluation of Vocal and Musical Sounds

The behavioral data informed on processes related to the cognitive evaluation of the emotional significance of vocal *vs.* musical sounds. In the post-EEG behavioral task, emotion categorization accuracy was generally high for nonmusicians and musicians (see Table 7). Accuracy was higher for vocalizations than for musical sounds, as indicated by a main effect of stimulus type ($F(1, 64) = 91.985$, $p < .001$, $\eta_p^2 = .590$). However, there were no differences between musicians and nonmusicians ($F(1, 64) = .126$, $p = .724$). The only significant effect involving music expertise was a three-way interaction between music expertise, stimulus type, and emotion ($F(2, 128) = 6.095$, $p =.003$, $\eta_p^2 = .087$). Musicians were more accurate at recognizing neutral compared to happy ($p < .001$) and sad ($p = .002$) vocalizations; in the case of nonmusicians, the only significant difference was between neutral and sad vocalizations ($p = .011$). In the case of musical sounds, musicians recognized all emotions with similar accuracy (all *p*s > .473); however, nonmusicians were more accurate at recognizing happy compared to neutral (*p* < .001) and sad (*p* < .001) musical sounds. Follow-up analyses comparing instrumentalists and singers did not yield significant main effects or interactions related to instrument-specific expertise (all *p*s > .26).

<INSERT TABLE 7 HERE>

Valence ratings followed the expected pattern for vocalizations and music (see Table 8): they were highest for happy stimuli, intermediate for neutral, and lowest for sad stimuli. Specifically, a main effect of emotion (*F*(2, 128) = 420.557, *p* < .001 , η_p^2 = .868) revealed that happy stimuli were rated as more pleasant than sad (*p* < .001) and neutral (*p* < .001) stimuli, whereas neutral stimuli were considered more pleasant than sad ones (*p* < .001). There were differences between musicians and nonmusicians, as indicated by a main effect of expertise (*F*(1, 64) = 5.719, *p* = .020, η_p^2 = .082), and by an interaction between music expertise and stimulus type (*F*(1, 64) = 11.889, *p* = .001, η_p^2 = .157). The two groups rated vocalizations similarly (*p* = .756) but musicians considered musical sounds to be generally more pleasant than nonmusicians (*p* = .001). Follow-up analyses comparing instrumentalists and singers did not yield significant main effects or interactions related to instrument-specific expertise (all *p*s > .15).

<INSERT TABLE 8 HERE>

Arousal ratings also followed the expected pattern (see Table 8): emotional vocal and musical sounds were perceived as more arousing compared to neutral sounds (all *p*s < .001), as indicated by a significant interaction between stimulus type and emotion (*F*(2, 128) = 11.815, *p* < .001, η_p^2 = .156). There were no differences between musicians and nonmusicians (*F*(1, 64) = .743, *p* = .392). The only effect involving music expertise was a significant three-way interaction between music expertise, stimulus type, and emotion ($F(2, 128) = 3.243$, $p = .042$, $\eta_p^2 = .048$). In the case of happy and neutral sounds, musicians (*vs.* nonmusicians) considered music to be more arousing than vocalizations (*p*s < .002). Follow-up analyses comparing instrumentalists and singers did not yield significant main effects or interactions related to instrument-specific expertise (all *p*s > .272).

The behavioral data analysis confirmed an advantage associated with music training in explicit decoding of emotions from musical stimuli, which were also generally perceived as more pleasant and arousing by musicians compared to nonmusicians.

3.3. Supplementary Analyses

In supplementary analyses, we tested whether the ERP findings were modulated by other aspects of music training, namely mixed vocal and instrumental training or practice of violin (since violin sounds were intermixed with clarinet sounds in the experimental task). The effects of music training on the P2 remained significant (see Supplementary Material). Furthermore, we tested whether the ERP findings were modulated by individual differences in musical sophistication (indexed by Gold-MSI scores), including differences in how participants engage with music and the skill they display for different types of musical activities. The Gold-MSI total score was added as a covariate in the statistical model. The ERP findings were not modulated by differences in musical sophistication (see Supplementary Material).

We also tested the association between ERP and behavioral data with a correlational analysis (see Supplementary Material). No significant associations were found in musicians (*p* > .05).

4. Discussion

The current study combined ERP and behavioral methods to ask whether long-term music training relates to distinct responses to emotional sounds. We compared how musicians and nonmusicians process emotional voices and music under implicit task instructions, and how they explicitly evaluate the affective properties of these sounds. An additional goal was to investigate whether instrument-specific expertise (singers *vs.* instrumentalists) selectively modulates brain and behavioral responses to vocal and musical sounds.

We partially confirmed our hypothesis that music training is linked to facilitated emotional sound processing. We observed that this facilitation is specific to musical stimuli but independent of instrument-specific expertise. Specifically, we observed enhanced salience detection (P2), attention orienting (P3), and elaborative processing (LPP) of musical sounds in musicians compared to nonmusicians. At the behavioral level, we found that accuracy in the emotional recognition of musical sounds was similar across valence types in musicians, who also judged musical sounds to be more pleasant and more arousing than nonmusicians.

In the following sections, we first discuss general effects related to manipulations of sound type and emotion, and then focus on specific effects of music training on emotional decoding, based on multi-stage models of auditory emotional processing (Frühholz et al., 2016; Schirmer & Kotz, 2006).

Decoding emotional meaning from vocal vs. musical sounds

We found that vocal and musical emotions are associated with differences in early and late stages of sound processing. The N1 amplitude was enhanced in response to musical (*vs.* vocal) sounds, whereas the P2, P3, and LPP were enhanced in response to vocal (*vs.* musical) sounds. These findings suggest differences in the processes underlying emotional meaning decoding from music *vs.* human voices, corroborating previous studies (e.g., Proverbio et al., 2020). The ERP differences between vocal and musical sounds were valence-specific and aligned with the behavioral ratings of sounds: the P2 was increased to happy and neutral voices, which were rated as more pleasant than happy and neutral musical sounds; the LPP (701-900 ms) was increased to emotional compared to neutral voices but no valence-specific modulations were observed in the case of musical sounds. Generally enhanced P2, P3, and LPP amplitudes for vocal compared to musical sounds may reflect the enhanced biological relevance of voices, which are often used as a means of communication in social interactions (see also Proverbio et al., 2020). That is, human voices are associated with enhanced salience detection (P2), attention orienting (P3), and late sustained attention (LPP) compared to music, confirming the privileged status of voices in the auditory system (Belin et al., 2004). The P2 findings also suggest that extracting negative valence from vocal and musical sounds (i.e., similar P2 for sad vocal and musical sounds) relies on a common mechanism. Since an implicit task was used, condition effects reflect the result of an automatic extraction of emotional salience from task-irrelevant auditory information.

Music training and auditory emotional processing

Partially confirming our hypotheses, the ERP findings revealed that music training was selectively associated with facilitated processing of musical sounds: in general, musicianship, but not instrument-specific expertise, was associated with differences in the P2, P3, and LPP components irrespective of valence. That is, sounds that were identified as potentially more relevant by musicians compared to nonmusicians – musically expressed emotions – were also the focus of more elaborated processing. These findings are consistent with the idea that auditory perception is influenced by the experience of the listener, and is not merely driven by sound properties (Krishnan et al., 2018).

Whereas in nonmusicians the P2 was increased in response to vocal compared to musical sounds, the P2 response did not discriminate between the two sound categories in musicians, which were associated with a similarly enhanced P2 response. A similar pattern was found in the P3. These findings are consistent with previous ERP studies indicating that the processing of musical sounds seems to be different in musicians when compared to nonmusicians. For example, Rigoulot and collaborators (Rigoulot et al., 2015) reported increased P2 amplitudes in response to musical compared to vocal sounds in musicians, whereas nonmusicians revealed an increased P2 in response to vocal compared to musical sounds. Given that the P2 has been related to automatic salience detection (Liu et al., 2012; Paulmann & Kotz, 2008; Pinheiro et al., 2013), the current findings suggest that long-term music practice is associated with the perceived salience of musical sounds. The lack of P2 differences between voices and music suggests that musical sounds are as salient as vocalizations in individuals with long-term music training, and therefore both sound types capture similar attentional resources. Plausibly, the enhanced precision of sensorimotor representations for music in experts (*vs.* nonmusicians) could have accounted for the easier salience detection in musical sounds (Krishnan et al., 2018). The subsequent LPP enhancement to music in listeners with (*vs.* without) formal music training is consistent with the notion that stimuli identified as salient need to be more thoroughly processed and engage sustained attention (Paulmann et al., 2013a). However, we should note that only the P2 findings remained significant after accounting for the effects of mixed vocal and instrumental training, as well as of the specific practice of violin. The robustness of the P2 effects suggests that music has a privileged status in the auditory system, similarly to voices, as a function of long-term music training.

The lack of valence-related differences between musicians and nonmusicians in the ERP data may be explained by a more controlled selection and matching of vocal and musical sounds in the current study. Importantly, the MEB was developed to represent the musical counterpart of the MAV sounds. Additionally, sound selection took into consideration the matching of acoustic and arousal properties in the two sound categories (see Methods). Notwithstanding, music training related to differences in explicit emotion recognition, as well as in the ratings of valence and arousal of emotional sounds. Musicians were similarly accurate in the emotion categorization of musical sounds irrespective of their valence, but nonmusicians were more accurate in the categorization of happy (*vs.* sad and neutral) music. This suggests that some emotion categories were easier to discriminate than others in those who did not have formal music training, whereas all categories were easily recognized by experts. Furthermore, musical sounds were perceived as more pleasant and arousing by musicians compared to nonmusicians. Musicians rated musical sounds as more pleasant than vocal sounds, as well as happy and neutral music as more arousing than happy and neutral voices. These findings suggest that music expertise is also associated with how emotional meaning is explicitly decoded from musical and vocal sounds. In previous studies, music expertise was linked to enhanced emotion recognition in music (Castro & Lima, 2014) and voices (Correia et al., 2020; Lima & Castro, 2011; Thompson et al., 2004). However, some of these studies were restricted to a subgroup of participants (Parsons et al., 2014) or a subset of emotions (Pinheiro et al., 2015; Thompson et al., 2004), whereas others reported null effects (Başkent et al., 2018; Dibben et al., 2018; Mualem & Lavidor, 2015; Park et al., 2015; Trimmer & Cuddy, 2008; Weijkamp & Sadakata, 2016). One possibility is that larger sample sizes are required for an association between emotion recognition accuracy and music training to emerge.

Although we did not find a significant association between the ERP and behavioral findings, the P2 and LPP results are congruent with valence and arousal explicit ratings: musicians rated music sounds as more pleasant and arousing than vocal sounds and their ERP responses were also enhanced for music compared to vocalizations. We note that even though the P2 and LPP are modulated by perceived sound valence and arousal (Delaney-Busch et al., 2016; Liu et al., 2012; Paulmann & Kotz, 2008; Paulmann et al., 2013), the two types of measures tap into implicit (ERP) *vs.* explicit (behavioral) aspects of emotional sound processing. Moreover, ERPs inform on the dynamic unfolding of operations related to deriving emotional significance from acoustic cues before a response is made, or in the absence of a response (as in the current EEG task), and may therefore reflect processes that are independent of later decision stages.

Together, the ERP effects suggest that music training does not necessarily relate to early low-level auditory-perceptual enhancements (which would be reflected in N1 differences), putatively related to enhanced experience in the identification of spectrotemporal and musical properties of sounds. Instead, advantages of music training in emotion decoding might be associated with facilitated processing in intermediate and higher-order stages of auditory emotional processing, namely enhanced salience detection (P2), attention orienting (P3), sustained attention and elaborative processing (LPP), and explicit evaluation of the emotional significance of *musical* sounds (accuracy, valence, and arousal ratings). In contrast with our hypothesis, our findings also suggest that the correlates of long-term music experience on emotional sound processing are independent of the specific training profile (singers *vs.* instrumentalists). Given previous fMRI studies (e.g., Krishnan et al., 2018) showing that the neural systems involved in music perception are modulated by previous sensorimotor experience, this finding was not anticipated. However, we need to point out differences in the temporal resolution of fMRI and EEG, which may have accounted for this apparent inconsistency. Additionally, the musical stimuli presented in the current study were not representative of the different types of musical instruments played by the instrumentalists in our sample. Furthermore, the vocal stimuli used in our task consist of vocalizations frequently perceived in daily social interactions and do not represent the type of sounds typically produced by singers in their musical practice. Moreover, the lack of differences could also be explained by issues related to sample size and statistical power. Therefore, future studies should compare the effects of different types of instrumental training and include larger samples of musicians, as well as distinct types of emotional sounds (e.g., sung notes) covering the full spectrum of discrete emotions and thoroughly controlled for perceived valence and arousal (e.g., Paulmann et al., 2013). Furthermore, we note that a cross-sectional comparison between musicians and nonmusicians in the current study does not establish causality (Schellenberg, 2020). Future studies should extend our cross-sectional approach and test music training effects with a longitudinal design, random assignment, and active control groups (Martins et al., 2021).

Conclusion

The current study tested whether and how long-term music training relates to a distinct emotional processing of voices and music. Furthermore, it examined how specific training profiles (singers, instrumentalists) modulate the brain responses to emotional sounds from early to late processing stages. Our ERP findings indicate that, in trained musicians, implicit processing of musical emotions is facilitated, as reflected in enhanced salience detection, attention orienting, and elaborative processing of musical (*vs.* vocal) sounds. The behavioral findings confirm the advantages of music training on the explicit evaluation of the emotional significance of musical sounds. Of note, different types of music training (singers *vs.* instrumentalists) did not relate to differences in the processing of vocal *vs.* musical sounds. Together, these findings reveal that music has a privileged status in the auditory system, similarly to voices, as a function of long-term music training: musical sounds are, therefore, deemed as more salient by musicians compared to nonmusicians. However, they do not support the view that potential differences in auditory emotional processing related to music expertise are exclusively due to bottom-up sensory differences, namely a facilitated acoustic processing of emotional sounds.

The relationship between music and socio-emotional processes, such as emotion recognition, should be further examined in future studies including larger samples of musicians with distinct types of training, as well as different types of emotional sounds. The clarification of these links is relevant to broader debates on the neurocognitive associations between musical and non-musical domains.

Acknowledgements

This work was supported by the Portuguese Science Foundation (*Fundação para a Ciência e a Tecnologia* [FCT]; grant number PTDC/MHC-PCN/0101/2014 awarded to APP and grant numbers PTDC/PSI-GER/28274/2017 awarded to CL and APP). The Authors would like to thank all the participants who collaborated in the study, as well as Marta Pereira for assistance with data analyses.

Open Practices Statement

The study was not preregistered. The materials used in our experimental task are publicly available at https://neuralbasesofcommunication.eu/download/. The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

References

Başkent, D., Fuller, C. D., Galvin, J. J., Schepel, L., Gaudrain, E., & Free, R. H. (2018). Musician effect on perception of spectro-temporally degraded speech, vocal emotion, and music in young adolescents. *The Journal of the Acoustical Society of America, 143*(5), EL311–EL316. https://doi.org/10.1121/1.5034489

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. https://doi.org/10.3758/BRM.40.2.531

Bestelmeyer, P. E. G., Kotz, S. A., & Belin, P. (2017). Effects of emotional valence and arousal on the voice perception network. *Social Cognitive and Affective Neuroscience*, *12*(8), 1351–1358. https://doi.org/10.1093/scan/nsx059

Bestelmeyer, P. E. G., Maurage, P., Rouger, J., Latinus, M., & Belin, P. (2014). Adaptation to Vocal Expressions Reveals Multistep Perception of Auditory Emotion. *Journal of Neuroscience*, *34*(24), 8098–8105. https://doi.org/10.1523/JNEUROSCI.4820-13.2014

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*(4), 2662–2673. https://doi.org/https://doi.org/10.1111/ejn.12627

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition and Emotion*, *19*(8), 1113–1139. https://doi.org/10.1080/02699930500204250

Canavarro, M. C. (1999). Inventário de Sintomas Psicopatológicos [Brief Symptoms Inventory]: BSI. In L. S. A. M. R. Simões, M. Gonçalves (Ed.), *Testes e provas psicológicas em Portugal [Psychological Tests in Portugal]-vol. II.* (pp. 87–109). SHO/APPORT.

Castiajo, P., & Pinheiro, A. P. (2021). Acoustic salience in emotional voice perception and its relationship with hallucination proneness. *Cognitive, Affective, & Behavioral Neuroscience*. https://doi.org/https://doi.org/10.3758/s13415-021-00864-2

Castiajo, Paula, & Pinheiro, A. P. (2019). Decoding emotions from nonverbal vocalizations:

How much voice signal is enough? *Motivation and Emotion*, *43*(5), 803–813. https://doi.org/10.1007/s11031-019-09783-9

Castro, S. L., & Lima, C. F. (2014). Age and Musical Expertise Influence Emotion Recognition in Music. *Music Perception*, *32*(2), 125–142. https://doi.org/10.1525/mp.2014.32.2.125

Chartrand, J.-P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neuroscience Letters*, *405*(3), 164–167. https://doi.org/https://doi.org/10.1016/j.neulet.2006.06.053

Chobert, J., François, C., Velay, J.-L., & Besson, M. (2014). Twelve Months of Active Musical Training in 8- to 10-Year-Old Children Enhances the Preattentive Processing of Syllabic Duration and Voice Onset Time. *Cerebral Cortex*, *24*(4), 956–967. https://doi.org/10.1093/cercor/bhs377

Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience*, *9*, 482. https://www.frontiersin.org/article/10.3389/fnhum.2015.00482

Correia, A. I., Castro, S. L., MacGregor, C., Müllensiefen, D., Schellenberg, E. G., & Lima, C. F. (2020). Enhanced recognition of vocal emotions in individuals with naturally good musical abilities. *Emotion*. https://doi.org/10.1037/emo0000770

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744. https://doi.org/10.1016/j.clinph.2003.11.021

Curtis, M. E., & Bharucha, J. (2010). The minor third communicates sadness in speech, mirroring its use in music. *Emotion*, *10*(3), 335–348. https://doi.org/10.1037/a0017928

Delaney-Busch, N., Wilkie, G., & Kuperberg, G. (2016). Vivid: How valence and arousal influence word processing under different task demands. In *Cognitive, Affective &*

Behavioral Neuroscience (Vol. 16, Issue 3, pp. 415–432). Springer. https://doi.org/10.3758/s13415-016-0402-y

Delplanque, S., Silvert, L., Hot, P., Rigoulot, S., & Sequeira, H. (2006). Arousal and valence effects on event-related P3a and P3b during emotional categorization. *International Journal of Psychophysiology*, *60*(3), 315–322. https://doi.org/10.1016/j.ijpsycho.2005.06.006

Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception: challenges and unresolved questions. *European Journal of Neuroscience*, *51*(5), 1151–1160. https://doi.org/https://doi.org/10.1111/ejn.13802

Dibben, N., Coutinho, E., Vilar, J. A., & Estévez-Pérez, G. (2018). Do Individual Differences Influence Moment-by-Moment Reports of Emotion Perceived in Music and Speech Prosody? *Frontiers in Behavioral Neuroscience*, *12*, 184. https://www.frontiersin.org/article/10.3389/fnbeh.2018.00184

Dick, F., Lee, H. L., Nusbaum, H., & Price, C. J. (2011). Auditory-Motor Expertise Alters “Speech Selectivity” in Professional Musicians and Actors. *Cerebral Cortex*, *21*(4), 938–948. https://doi.org/10.1093/cercor/bhq166

Dumont, E., Syurina, E. V, Feron, F. J. M., & van Hooren, S. (2017). Music Interventions and Child Development: A Critical Review and Further Directions. *Frontiers in Psychology*, *8*, 1694. https://www.frontiersin.org/article/10.3389/fpsyg.2017.01694

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased Cortical Representation of the Fingers of the Left Hand in String Players. *Science*, *270*(5234), 305–307. https://doi.org/10.1126/science.270.5234.305

Espírito-Santo, H., Pires, C. F., Garcia, I. Q., Daniel, F., Silva, A. G. da, & Fazio, R. L. (2017). Preliminary validation of the Portuguese Edinburgh Handedness Inventory in an adult sample. *Applied Neuropsychology: Adult*, *24*(3), 275–287.

https://doi.org/10.1080/23279095.2017.1290636

Farmer, E., Jicol, C., & Petrini, K. (2020). Musicianship Enhances Perception But Not Feeling of Emotion From Others' Social Interaction Through Speech Prosody. *Music Perception*, *37*(4), 323–338. https://doi.org/10.1525/mp.2020.37.4.323

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal Recognition of Three Basic Emotions in Music. *Current Biology*, *19*(7), 573–576. https://doi.org/https://doi.org/10.1016/j.cub.2009.02.058

Fruhholz, S., van der Zwaag, W., Saenz, M., Belin, P., Schobert, A. K., Vuilleumier, P., & Grandjean, D. (2016). Neural decoding of discriminative auditory object features depends on their socio-affective valence. *Social Cognitive and Affective Neuroscience*, *11*(10), 1638–1649. https://doi.org/10.1093/scan/nsw066

Frühholz, S, Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, *68*, 1–15. https://doi.org/10.1016/j.neubiorev.2016.05.002

Frühholz, Sascha, Trost, W., & Kotz, S. A. (2016). The sound of emotions-Towards a unifying neural network perspective of affective sound processing. *Neuroscience and Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002

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. https://doi.org/10.1093/brain/awl247

Fuller, C. D., Galvin, J. J., Maat, B., Free, R. H., & Başkent, D. (2014). The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations?

Frontiers in Neuroscience, *8*, 179. https://www.frontiersin.org/article/10.3389/fnins.2014.00179

Galinha, I. C., & Pais-Ribeiro, J. L. (2005). Contribuição para o estudo da versão portuguesa da Positive and Negative Affect Schedule (PANAS): II-Estudo psicométrico [Contribution to the study of the Portuguese version of the Positive and Negative Affect Schedule (PANAS): II-Psychometric study]. *Análise Psicológica*, *23*(2), 219–227.

Giordano, B. L., Whiting, C., Kriegeskorte, N., Kotz, S. A., Gross, J., & Belin, P. (2021). The representational dynamics of perceived voice emotions evolve from categories to dimensions. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-021-01073-0

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

Grau-Sánchez, J., Münte, T. F., Altenmüller, E., Duarte, E., & Rodríguez-Fornells, A. (2020). Potential benefits of music playing in stroke upper limb motor rehabilitation. *Neuroscience & Biobehavioral Reviews*, *112*, 585–599. https://doi.org/https://doi.org/10.1016/j.neubiorev.2020.02.027

Halwani, G., Loui, P., Rueber, T., & Schlaug, G. (2011). Effects of Practice and Experience on the Arcuate Fasciculus: Comparing Singers, Instrumentalists, and Nonmusicians. *Frontiers in Psychology*, *2*, 156. https://www.frontiersin.org/article/10.3389/fpsyg.2011.00156

Herholz, S. C., & Zatorre, R. J. (2012). Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron*, *76*(3), 486–502. https://doi.org/https://doi.org/10.1016/j.neuron.2012.10.011

Ilie, G., & Thompson, W. F. (2006). A Comparison of Acoustic Cues in Music and Speech for Three Dimensions of Affect. *Music Perception*, *23*(4), 319–330.

https://doi.org/10.1525/mp.2006.23.4.319

Jeon, J., & Fricke, F. R. (1997). Duration of Perceived and Performed Sounds. *Psychology of Music*, *25*(1), 70–83. https://doi.org/10.1177/0305735697251006

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage*, *58*(2), 665–674. https://doi.org/10.1016/j.neuroimage.2011.06.035

Juslin, P N, & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Juslin, Patrik N, & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch Discrimination: Are Professional Musicians Better than Nonmusicians? *Journal of Basic and Clinical Physiology and Pharmacology*, *12*(2), 125–144. https://doi.org/doi:10.1515/JBCPP.2001.12.2.125

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The Brain of Opera Singers: Experience-Dependent Changes in Functional Activation. *Cerebral Cortex*, *20*(5), 1144–1152. https://doi.org/10.1093/cercor/bhp177

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., Manly, T., & Scott, S. K. (2018). Beatboxers and Guitarists Engage Sensorimotor Regions Selectively When Listening to the Instruments They can Play. *Cerebral Cortex*, *28*(11), 4063–4079. https://doi.org/10.1093/cercor/bhy208

Lappe, C., Herholz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical Plasticity Induced by Short-Term Unimodal and Multimodal Musical Training. *The Journal of Neuroscience*, *28*(39), 9632–9639. https://doi.org/10.1523/JNEUROSCI.2254-08.2008

Lima, C. F., Anikin, A., Monteiro, A. C., Scott, S. K., & Castro, S. L. (2019). Automaticity in the recognition of nonverbal emotional vocalizations. *Emotion*, *19*(2), 219–233. https://doi.org/10.1037/emo0000429

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. https://doi.org/10.1037/a0024521

Lima, C. F., Correia, A. I., Müllensiefen, D., & Castro, S. L. (2018). Goldsmiths Musical Sophistication Index (Gold-MSI): Portuguese version and associations with socio-demographic factors, personality and music preferences. *Psychology of Music*, *48*(3), 376–388. https://doi.org/10.1177/0305735618801997

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. https://doi.org/10.1097/WNR.0b013e32834ea757

Liu, Y., Huang, H., McGinnis-Deweese, M., Keil, A., & Ding, M. (2012). Neural Substrate of the Late Positive Potential in Emotional Processing. *Journal of Neuroscience*, *32*(42), 14563 LP – 14572. https://doi.org/10.1523/JNEUROSCI.3109-12.2012

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. M. (2009). Selective neurophysiologic responses to music in instrumentalists with different listening biographies. *Human Brain Mapping*, *30*(1), 267–275. https://doi.org/https://doi.org/10.1002/hbm.20503

Martins, M., Pinheiro, A. P., & Lima, C. F. (2021). Does Music Training Improve Emotion Recognition Abilities? A Critical Review. *Emotion Review*, *13*(3), 199–210. https://doi.org/10.1177/17540739211022035

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*(1), 36–47. https://doi.org/https://doi.org/10.1016/j.heares.2006.05.004

Moran, T. P., Jendrusina, A. A., & Moser, J. S. (2013). The psychometric properties of the late positive potential during emotion processing and regulation. *Brain Research*, *1516*, 66–75. https://doi.org/https://doi.org/10.1016/j.brainres.2013.04.018

Mualem, O., & Lavidor, M. (2015). Music education intervention improves vocal emotion recognition. *International Journal of Music Education*, *33*(4), 413–425. https://doi.org/10.1177/0255761415584292

Munzer, S., Berti, S., & Pechmann, T. (2002). Encoding of timbre, speech, and tones: Musicians vs. nonmusicians. *Psychologische Beiträge*, *44*(2), 187. https://search.proquest.com/scholarly-journals/encoding-timbre-speech-tones-musicians-vs-non/docview/212159059/se-2?accountid=134457

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*, *104*(40), 15894–15898. https://doi.org/10.1073/pnas.0701498104

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. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2008). Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians. *Psychophysiology*, *45*(6), 994–1007. https://doi.org/https://doi.org/10.1111/j.1469-8986.2008.00689.x

Nolden, S., Rigoulot, S., Jolicoeur, P., & Armony, J. L. (2017). Effects of musical expertise on oscillatory brain activity in response to emotional sounds. *Neuropsychologia*, *103*, 96–105. https://doi.org/10.1016/j.neuropsychologia.2017.07.014

Pantev, C., Engelien, A., Candia, V., & Elbert, T. (2001). Representational Cortex in Musicians. *Annals of the New York Academy of Sciences, 930*(1), 300–314. https://doi.org/https://doi.org/10.1111/j.1749-6632.2001.tb05740.x

Pantev, C., & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience & Biobehavioral Reviews*, *35*(10), 2140–2154. https://doi.org/https://doi.org/10.1016/j.neubiorev.2011.06.010

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport*, *12*(1). https://journals.lww.com/neuroreport/Fulltext/2001/01220/Timbre_specific_enhancement_of_auditory_cortical.41.aspx

Paquette, S., Peretz, I., & Belin, P. (2013). The "Musical Emotional Bursts": a validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology*, *4,* 509. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00509

Paquette, S., Rigoulot, S., Grunewald, K., & Lehmann, A. (2020). Temporal decoding of vocal and musical emotions: Same code, different timecourse? *Brain Research*, *1741*, 146887. https://doi.org/https://doi.org/10.1016/j.brainres.2020.146887

Paquette, S., Takerkart, S., Saget, S., Peretz, I., & Belin, P. (2018). Cross-classification of musical and vocal emotions in the auditory cortex. *Annals of the New York Academy of Sciences*, *1423*(1), 329–337. https://doi.org/https://doi.org/10.1111/nyas.13666

Park, M., Gutyrchik, E., Welker, L., Carl, P., Pöppel, E., Zaytseva, Y., Meindl, T., Blautzik, J., Reiser, M., & Bao, Y. (2015). Sadness is unique: neural processing of emotions in speech prosody in musicians and nonmusicians. *Frontiers in Human Neuroscience*, *8*, 1049. https://www.frontiersin.org/article/10.3389/fnhum.2014.01049

Parsons, C. E., Young, K. S., Jegindø, E.-M. E., Vuust, P., Stein, A., & Kringelbach, M. L. (2014). Music training and empathy positively impact adults' sensitivity to infant distress. *Frontiers in Psychology*, *5*, 1440. https://www.frontiersin.org/article/10.3389/fpsyg.2014.01440

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108. https://doi.org/https://doi.org/10.1016/j.heares.2013.08.011

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, *19*(2), 209–213. https://doi.org/10.1097/WNR.0b013e3282f454db

Paulmann, S, Bleichner, M., & Kotz, S. (2013a). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, 345. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00345

Paulmann, S, Bleichner, M., & Kotz, S. A. (2013b). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, 345. https://doi.org/10.3389/fpsyg.2013.00345

Paulmann, S, Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, *50*(7), 1609–1620. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2012.03.014

Paulmann, Silke, Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, 345. https://doi.org/10.3389/fpsyg.2013.00345

Pinheiro, A. P., Barros, C., Dias, M., & Kotz, S. A. (2017). Laughter catches attention! *Biological Psychology*, *130*, 11–21. https://doi.org/10.1016/j.biopsycho.2017.09.012

Pinheiro, A. P., Barros, C., & Pedrosa, J. (2016). Salience in a social landscape: Electrophysiological effects of task-irrelevant and infrequent vocal change. *Social*

Cognitive and Affective Neuroscience, *11*(1), 127–139. https://doi.org/10.1093/scan/nsv103

Pinheiro, A. P., Barros, C., Vasconcelos, M., Obermeier, C., & Kotz, S. A. (2017). Is laughter a better vocal change detector than a growl? *Cortex*, *92*, 233–248. https://doi.org/10.1016/j.cortex.2017.03.018

Pinheiro, A. P., del Re, E., Mezin, J., Nestor, P. G., Rauber, A., McCarley, R. W., Gonçalves, Ó. F., & Niznikiewicz, M. A. (2013). Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: an electrophysiological investigation. *Psychological Medicine*, *43*(3), 603–618. https://doi.org/10.1017/S003329171200133X

Pinheiro, A. P., Rezaii, N., Nestor, P. G., Rauber, A., Spencer, K. M., & Niznikiewicz, M. (2016). Did you or I say pretty, rude or brief? An ERP study of the effects of speaker's identity on emotional word processing. *Brain and Language*, *153–154*, 38–49. https://doi.org/10.1016/j.bandl.2015.12.003

Pinheiro, A. P., Vasconcelos, M., Dias, M., Arrais, N., & Gonçalves, O. F. (2015). The music of language: An ERP investigation of the effects of musical training on emotional prosody processing. *Brain and Language*, *140*, 24–34. https://doi.org/10.1016/j.bandl.2014.10.009

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 976–986. https://doi.org/10.1037/0096-1523.20.5.976

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Proverbio, A. M., De Benedetto, F., & Guazzone, M. (2020). Shared neural mechanisms for processing emotions in music and vocalizations. *European Journal of Neuroscience*, *51*(9), 1987–2007. https://doi.org/https://doi.org/10.1111/ejn.14650

Rammsayer, T., & Altenmüller, E. (2006). Temporal Information Processing in Musicians and Nonmusicians. *Music Perception*, *24*(1), 37–48. https://doi.org/10.1525/mp.2006.24.1.37

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience*, *290*, 175–184. https://doi.org/https://doi.org/10.1016/j.neuroscience.2015.01.033

Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., & Münte, T. F. (2001). Event-related brain potentials to sound omissions differ in musicians and nonmusicians. *Neuroscience Letters*, *308*(1), 33–36. https://doi.org/https://doi.org/10.1016/S0304-3940(01)01977-2

Sauter, D. A., & Eimer, M. (2010). Rapid Detection of Emotion from Human Vocalizations. *Journal of Cognitive Neuroscience*, *22*(3), 474–481. https://doi.org/10.1162/jocn.2009.21215

Schellenberg, E. G. (2020). Correlation = causation? Music training, psychology, and neuroscience. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(4), 475–480. https://doi.org/10.1037/aca0000263

Schirmer, A. (2018). Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing. *Social Cognitive and Affective Neuroscience*, *13*(1), 1–13. https://doi.org/10.1093/scan/nsx142

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, *16*(6), 635–639. https://doi.org/00001756-200504250-00024 [pii]

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., & Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage*, *41*(1), 113–122. https://doi.org/https://doi.org/10.1016/j.neuroimage.2008.01.067

Sharp, A., Houde, M.-S., Bacon, B.-A., & Champoux, F. (2019). Musicians Show Better Auditory and Tactile Identification of Emotions in Music. *Frontiers in Psychology*, *10*, 1976. https://www.frontiersin.org/article/10.3389/fpsyg.2019.01976

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *76*(6), 1690–1695. https://doi.org/10.1121/1.391605

Strong, J. V, & Mast, B. T. (2019). The cognitive functioning of older adult instrumental musicians and nonmusicians. *Aging, Neuropsychology, and Cognition*, *26*(3), 367–386. https://doi.org/10.1080/13825585.2018.1448356

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, *161*(1), 1–10. https://doi.org/10.1007/s00221-004-2044-5

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*(1), 46–64. https://doi.org/10.1037/1528-3542.4.1.46

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, *8*(6), 838–849. https://doi.org/10.1037/a0014080

Vasconcelos, M., Dias, M., Soares, A. P., & Pinheiro, A. P. (2017). What is the Melody of That Voice? Probing Unbiased Recognition Accuracy with the Montreal Affective Voices. *Journal of Nonverbal Behavior*, *41*(3), 239–267. https://doi.org/10.1007/s10919-017-0253-4

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28. https://doi.org/10.1007/BF00987006

Weijkamp, J., & Sadakata, M. (2016). Attention to affective audio-visual information: Comparison between musicians and nonmusicians. *Psychology of Music*, *45*(2), 204–215. https://doi.org/10.1177/0305735616654216

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. https://doi.org/10.1016/S0079-6123(06)56013-3

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage*, *40*(4), 1871–1887. https://doi.org/https://doi.org/10.1016/j.neuroimage.2008.01.026