

# Shared neural mechanisms for processing emotions in music and vocalizations

Alice Mado Proverbio<sup>1,2</sup>  | Francesco De Benedetto<sup>1,2</sup> | Martina Guazzone<sup>1,2</sup>

<sup>1</sup>Department of Psychology, University of Milano-Bicocca, Milan, Italy

<sup>2</sup>Milan Center for Neuroscience, Milan, Italy

## Correspondence

Alice Mado Proverbio, Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo 1, U6, 20126 Milan, Italy.  
Email: mado.proverbio@unimib.it

## Funding information

Università di Milano-Bicocca, Grant/Award Number: 16940

## Abstract

The neural mechanisms involved in the processing of vocalizations and music were compared, in order to observe possible similarities in the encoding of their emotional content. Positive and negative emotional vocalizations (e.g. laughing, crying) and violin musical stimuli digitally extracted from them were used as stimuli. They shared the melodic profile and main pitch/frequency characteristics. Participants listened to vocalizations or music while detecting rare auditory targets (bird tweeting, or piano's arpeggios). EEG was recorded from 128 sites. P2, N400 and *Late positivity* responses of ERPs were analysed. P2 peak was earlier in response to vocalizations, while P2 amplitude was larger to positive than negative stimuli. N400 was greater to negative than positive stimuli. LP was greater to vocalizations than music and to positive than negative stimuli. Source modelling using swLORETA suggested that, among N400 generators, the left middle temporal gyrus and the right uncus responded to both music and vocalizations, and more to negative than positive stimuli. The right parahippocampal region of the limbic lobe and the right cingulate cortex were active during music listening, while the left superior temporal cortex only responded to human vocalizations. Negative stimuli always activated the right middle temporal gyrus, whereas positively valenced stimuli always activated the inferior frontal cortex. The processing of emotional vocalizations and music seemed to involve common neural mechanisms. Notation obtained from acoustic signals showed how emotionally negative stimuli tended to be in Minor key, and positive stimuli in Major key, thus shedding some lights on the brain ability to understand music.

## KEYWORDS

emotions, ERP, language, minor major key, music, N400, sadness, valence

**Abbreviations:** ANOVA, Analysis of variance; APA, American Psychological Association; ASA, Advanced Source Analysis; BA, Brodmann area; BEM, Boundary element model; BMJ, British Medical Journal; DBFs, Decibels relative to full scale; EOG, Electro-oculogram; ERP, Event-related potential; IFC, inferior frontal cortex; ISI, Inter-stimulus interval; ITI, Inter-trial interval; fMRI, Functional magnetic resonance imaging; fNIRS, Functional near-infrared spectroscopy; LP, Late Positivity;  $\mu$ V, Microvolt; nA, Nanoampere; rIPS, Right intra-parietal sulcus; RMS, Root mean square; RT, Reaction time; SD, Standard deviation; SE, Standard error; SMA, Self-Assessment Manikin; swLORETA, standardized weighted Low-Resolution Electromagnetic Tomography.

Edited by Sophie Molholm.

The peer review history for this article is available at <https://publons.com/publon/10.1111/ejn.14650>

## 1 | INTRODUCTION

Music is able to communicate emotions and greatly affect the listener's mood. Despite differences in compositional styles, cultures and aesthetic preferences, there are certain pieces that almost invariably evoke similar emotions in the listeners (Proverbio et al., 2015). For example lullabies have relaxing properties and lower EEG frequency while rhythmic music has exciting properties and stimulates the motor cortex and basal ganglia (Hernandez-Reif, Diego, & Field, 2006; McAuley, Henry, & Tkach, 2012; Tan, Ozdemir, Temiz, & Celik, 2015). Neuroscientists widely investigated the neural bases of aesthetic sensations in music (Brattico & Pearce, 2013; Hopyan, Manno, Papsin, & Gordon, 2015; Huron, 2013; Kawakami, Furukawa, Katahira, & Okanoya, 2013; Koelsch, 2014; Liégeois-Chauvel et al., 2014; Peretz & Zatorre, 2003; Trost, Ethofer, Zentner, & Vuilleumier, 2012; Vuilleumier & Trost, 2015; Zentner, Grandjean, & Scherer, 2008; Zhao & Chen, 2009). It is known, for example, that fearful music activates the amygdala (e.g. Baumgartner, Lutz, Schmidt, & Jäncke, 2006; Koelsch et al., 2013), that atonal music induces *fear bradycardia* (Proverbio et al., 2015), that listening to heroic and triumphant music activates the ventral striatum (Vuilleumier & Trost, 2015) while listening to sad, nostalgic music activates the insula and the cingulate cortex (Huron, 2013; Janata, 2009; Trost et al., 2012) among other regions. Generally speaking, music characterized by major mode relatively fast tempo and high frequencies is perceived as positive, happy. Minor mode, slower tempo and lower frequencies are instead associated with sadness and negativity (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001; Eerola, Friberg, & Bresin, 2013; Juslin, Liljeström, Västfjäll, & Lundqvist, 2010; Rigg, 1964; Zentner & Eerola, 2009).

Another important structural parameter of music is harmonic dissonance, which makes musical stimuli more or less aesthetically pleasant. It has been demonstrated a physiologically based innate preference for harmonic than inharmonic intervals or chords in 3 days newborns (Perani et al., 2010; Virtala, Huotilainen, Putkinen, Makkonen, & Tervaniemi, 2012), in human adults (Bidelman & Heinz, 2011; Bidelman & Krishnan, 2009; Minati et al., 2009; Proverbio, Orlandi, & Pisanu, 2016), and in primates such as macaques (Fishman et al., 2001), or chimpanzees (Sugimoto et al., 2010). If major chords are generally perceived as positive and minor chords as negative sounding, the specific mechanisms that give music its emotional meaning remain poorly understood (e.g. Bakker & Martin, 2015; Juslin et al., 2010).

It has been hypothesized by some authors that minor chords share the same harmonic relationships that are found in sad speech prosody (Cook, Fujisawa, & Takami, 2006). Paquette, Takerkart, Saget, Peretz, and Belin (2018)

proposed that music could tap into neuronal circuits that have evolved primarily for the processing of biologically important emotional vocalizations and a few studies have indeed found common brain activations (including middle and superior temporal cortex, medial frontal cortex and cingulate cortex) during listening to both non-linguistic vocalizations and musical excerpts (Aubé, Angulo-Perkins, Peretz, Concha, & Armony, 2015; Belin et al., 2008; Escoffier, Zhong, Schirmer, & Qiu, 2013). Although from these studies it can be surely argued that similar mechanisms support the comprehension of emotional content of vocalizations and music, there is not a precise indication about specific similarities/differences in neural processing and its time course, because of the different structure of stimuli usually compared. Non-verbal emotional vocalizations tend to activate the inferior frontal cortex (IFC) (Belin et al., 2008; Warren et al., 2006), also in newborns (Cheng, Lee, Chen, Wang, & Decety, 2012; Grossmann, Oberecker, Koch, & Friederici, 2010). The sensitivity of IFC to vocalizations is due to the numerous connections with the superior temporal gyrus, from which it receives different types of auditory and affective information. Whitehead and Armony (2018) compared human vocalization with music and found that the latter (vocal or instrumental), preferentially activated regions in the superior temporal gyrus, while the human voice, either spoken or sung, mostly activated more strongly a large area along the superior temporal sulcus.

In order to directly investigate whether, to extract the emotional meaning, music processing engaged neural regions devoted to non-linguistic vocalizations, in this study musical stimuli were perfectly identical to vocal stimuli under many perceptual aspects, except timbre. In fact, differences in perceptual structure may result in different signals. Musical stimuli were obtained by converting in music real positive and emotional non-verbal vocalizations of female and males speakers. Although deriving from the vocalizations, music appeared as completely novel to the listener (as ascertained through an open question posed at the end of validation session), slightly dodecaphonic in style.

To investigate (somewhat similarly) the relation between music structure and emotion, Curtis and Bharucha (2010) examined whether the pitch contours of affective speech exhibited patterns similar to those used in the musical communication of emotion. They had male and female actors reciting bisyllabic utterances with four different emotions (anger, happiness, pleasantness and sadness) and then performed pitch contour analysis on the stimuli. It was found that sad utterances presented a peak at 300 cents (a descending minor third), while angry utterances showed ascending minor second intervals (100 cents) and ascending perfect fifth ones (700 cents intervals). These findings corroborate the hypothesis that the ability to attribute a sad character to sad music (featuring a minor modality) may

depend on their tapping at innate neural mechanism for processing emotional vocalizations.

This hypothesis was directly tested in the present study. Audio files of spontaneous emotional vocalizations (e.g. crying, laughing, expressing appreciation or pain) were converted in music melodies with a violin timbre. ERPs to vocalizations and musical stimuli randomly mixed with fictitious targets were then recorded in participants unaware of study's purpose. Brain potentials to stimuli were analysed by considering their emotional valence and type: it was expected to find similarities based on valence and regardless of stimulus type, in the hypothesis that music and vocalization shared a common neural mechanism for the extraction of their emotional content.

The available literature on speech processing, considering both semantic content and emotional prosody, have found evidence that P200, N400 and Late Positivity (LP) components of ERPs are typically modulated by stimulus emotional value. P2 is the earliest response being sensitive to the emotional content of linguistic stimuli (Paulmann & Kotz, 2008; Paulmann, Seifert, & Kotz, 2010; Schimer, Chen, Ching, Tan, & Hong, 2013; Schirmer & Kotz, 2006) with positive words typically eliciting larger P200s than neutral words (e.g. Kanske & Kotz, 2007; Paulmann, Bleichner, & Kotz, 2013). Later in time, N400 response has been found larger to angry than happy speech prosody (Grossmann, Striano, & Friederici, 2005). Finally, long latency LPs are known to be modulated by the emotional valence of verbal stimuli (Carretiè et al., 2008; Hinojosa, Méndez-Bértolo, & Pozo, 2010; Kanske & Kotz, 2007), with positively valenced words typically eliciting enhanced LP responses (Bayer & Schacht, 2014; Herbert, Junghofer, & Kissler, 2008; Herbert, Kissler, Junghöfer, Peyk, & Rockstroh, 2006). In this context, we hypothesized that the processing of affective vocalizations and music shared some similarities with the processing of speech prosody (Proverbio & Santoni, 2019). Therefore, the amplitude of P200, N400 and Late Positivity was analysed to observe whether the known effects were recognizable on ERPs.

Source reconstructions of N400 responses were also performed in the 450–600 ms time window, wishing to find common neural circuits for similar emotional correlates, regardless of stimulus type (human voice or violin music). According to the literature (e.g. Belin et al., 2008), it was expected a right middle/superior temporal activation for negative stimuli and an inferior frontal activation for positive stimuli.

In addition, it was expected to find a pattern of hemispheric asymmetry in brain activation with a stronger left hemisphere activity during listening to positive music and right hemisphere activity during listening to negative music, as shown in previous fMRI/EEG study (e.g. Flores-Gutiérrez et al., 2007).

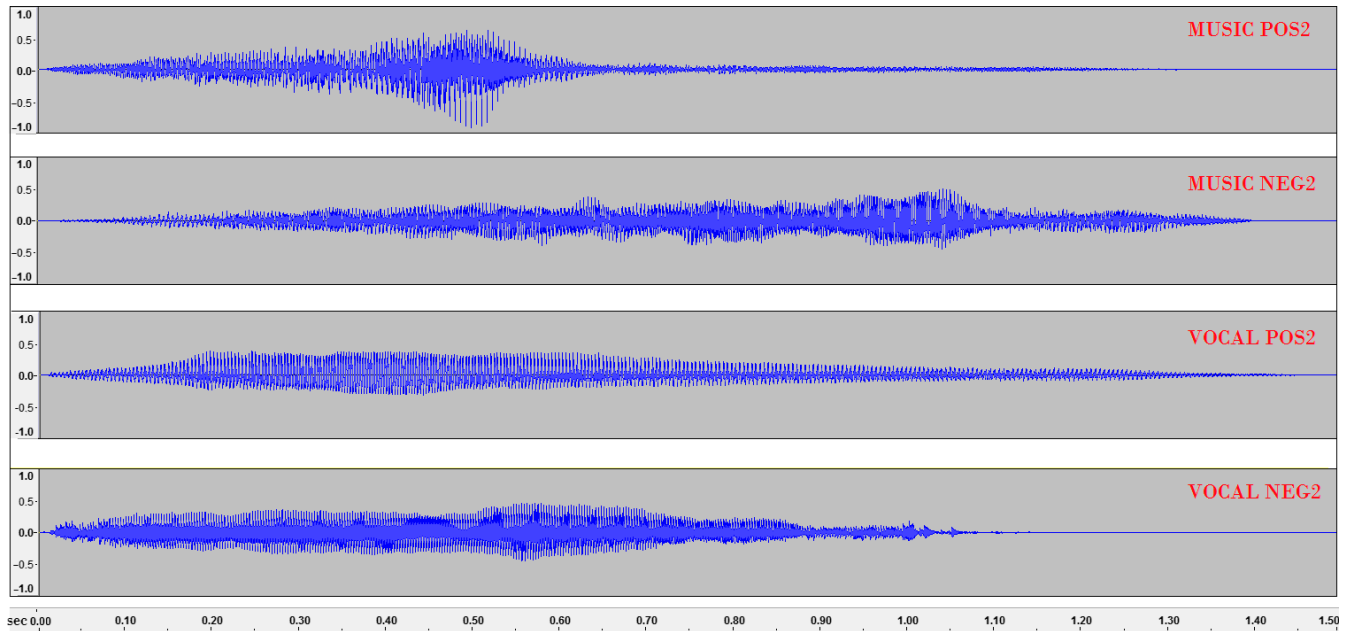
## 2 | MATERIAL AND METHODS

### 2.1 | Participants

Twenty University students (five men and 15 women) with a mean age of 22 years ( $SD = 4.68$ ) took part in the study. All participants had normal or corrected-to-normal vision and were right-handed as assessed by the Edinburgh Inventory and reported an average 0.77 score of lateral preference (scale  $-1$  left,  $+1$  right), and none had any left-handed relatives. Experiments were conducted with the understanding and written consent of each participant according to the Declaration of Helsinki (BMJ 1991; 302:1194) with approval from the local Ethical Committee (Prot. 0000273/14) and in compliance with APA ethical standards for the treatment of human volunteers (1992, American Psychological Association). All participants received academic credit for their participation. Data from two men and three women were subsequently discarded because of excessive eye movements and electroencephalogram (EEG) artefacts. The average of the final sample of participants was 23.07 ( $DS = 5.19$ ).

### 2.2 | Stimuli and material

The stimuli consisted of 64 audio files (32 positive and 32 negative) obtained by transforming real emotional vocalizations into music melodies. In addition, eight chirping sounds of birds and eight piano arpeggios were used as target stimuli. The vocalizations and melodies were repeated 3–4 times each in random order within the experiment, while the targets were repeated twice each. As for emotional vocalizations, 24 of these were selected from the *Oxford Vocal Sounds database* (OxVoc) and 40 were recorded ex novo specifically for the study. The *Oxford Vocal Sounds* (Parsons, Young, Craske, Stein, & Kringelbach, 2014) is a database including emotional vocalizations of adults, children and animals. The vocal sounds are natural, not recited, in fact they were extracted from video-diaries published on *YouTube* or on blogs by American and British subjects between 18 and 30 years. For the present study, 12 laughter and 12 crying of adult speakers were selected as experimental stimuli, and a few crying of children and neutral vocalizations of adults were used as stimuli for the training sequences, presented before the experimental sequences. The 12 selected crying were voiced by female speakers, while among the laughs, eight were voiced by women and four by men. This female prevalence was determined (according to the authors, namely Parsons et al., 2014), by a difficulty in accessing recordings of spontaneous affective male vocalizations. The stimuli were modified using *Audacity* software and were cut out so that they



**FIGURE 1** Frequency spectrum analysis relative to 4 examples of stimuli belonging to the various categories, obtained via *Audacity* software. The spectrogram is a spectro-temporal representation of the auditory sounds. The horizontal direction of the spectrogram represents time (from 0 to 1.5 s.); the vertical direction represents normalized frequency (ranging from 0 to 44,100 Hz). Darker parts of the spectrogram indicate higher energy densities, lighter parts indicate mean lower energy densities [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

lasted 1500 ms. The sound intensity was modulated so that it reached its maximum value in the first 150 ms and gradually decreased in the last 150 ms ("Volume Envelope" function). Figure 1 shows the energy distribution (spectrogram) of some example of stimuli, taken from each of the stimulus categories.

The other emotional vocalizations used for the study were recorded with *Apple's iPhone 4s "Recorder"* application and consisted in: 10 surprise exclamations, 10 appreciation sounds, 10 screams of fear and 10 complaints of psychological or physical pain. The vocalizations were produced by professional and amateur actors. For the exclamations of surprise and appreciation, actors were asked to think about an exclamatory phrase that involved surprise or appreciation (e.g. WOW, congratulations, good for you!) while for the screams and complaints of pain it was simply asked to think of something scary or to feel pain or desperation. The exclamations of surprise and those of appreciation were voiced by three women and three men; the screams of fear by eight women and two men; the laments of pain by nine women and one men. All the vocalizations lasted 1,500 ms and were attenuated exactly as the stimuli taken from *Oxford Vocal Sounds database*. Eight exemplars of auditory clips of music and vocalizations (two for each of the four categories) are provided as Supplementary Files and can be listened to by the readers.

Musical stimuli were created by means of *Apple's Logic Pro X* program (version 10.3.1). The human vocalizations were imported into the software and transformed in music

through the "create MIDI from flex data" function, able to produce MIDI instrumental tracks via *Musical Instrument Digital Interface*. The selected timbre was violin. Figure 1 shows the notation of some of the negative and positive music stimuli obtained by means of *AnthemScore 3.6.0* software. It can be noted that positive music excerpts contained intervals typical of major tonality, such as octaves, perfect fifths and major thirds.

As for target stimuli, the piano arpeggios were all of higher key and were recorded directly from a real piano. The chirps of birds were taken from *YouTube* videos, by means of the "Recorder" application of an *iPhone 4s*. Musical stimuli lasted 1,500 ms and were attenuated at the beginning and at the end as all the other stimuli, as described above.

The mean square value of the amplitude (RMS) was uniformed for all stimuli (vocalizations, melodies and targets) at  $-20$  dBFS, so that the intensity was the same for all stimuli (meter resolution = 0.1db). Auditory stimuli were reproduced in a pair of *Sennheiser HD202* headphones. Sound intensity was measured through a sound level meter in HIGH-FAST mode. Intensity values underwent a two-ways ANOVA comparing vocalizations and melodies, and their emotional content. The results showed a lack of difference across classes [ $F_{1,37} = 0.66$ ;  $p = .420$ ] and valences, respectively [ $F_{1,37} = 0.42$ ;  $p = .657$ ].

To avoid possible basic differences in spectrogram between positive and negative stimuli, eventually affecting early ERP responses, the frequency spectrum of each individual musical

stimulus was analysed from 0 to 250 ms, by means of *Audacity* software (parameters were as follows: sample rate: 44,100 Hz. length processed: 1,103 samples; 0.025000 s). *Audacity* is an audio editor software performing spectrogram visualization and detailed frequency analysis. The mean values so obtained underwent an ANOVA comparing (non-target) negative and positive stimuli whose results showed a lack of difference across classes [ $F_{1,31} = 0.001$ ;  $p = .9813$ ]. Although the average value represented only the average frequency spectrum of the signals belonging to the two classes of stimuli (positive and negative music) in the given time window, this analysis has provided an evaluation of the relative uniformity of the two envelopes, which could have been different, for example, in the case of speech with different formant frequencies due to gender or age variability of the speakers. It is noteworthy that this approach does not compare the dynamic spectro-temporal features of the stimuli, and thus while overall average value did not differ, the dynamic signal may well have differed and influenced the sensory responses.

### 2.3 | Stimulus validation

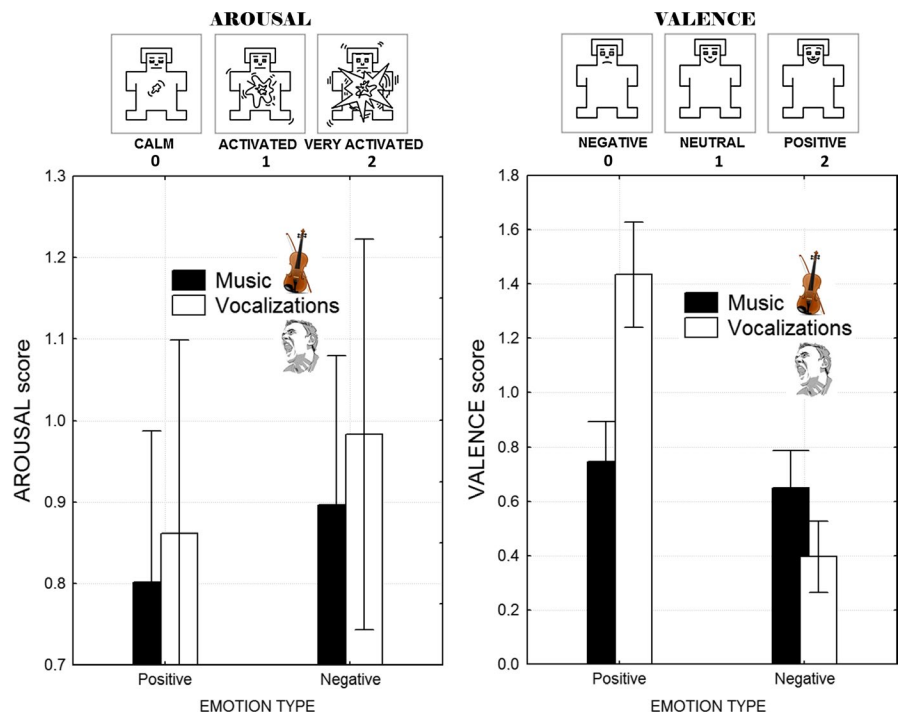
Twelve judges took part in stimuli validation session. They were men and women ageing on average 24.75 years ( $SD = 2.77$ ). Judges received visual instructions and provided a modified version of the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994), a non-verbal visual assessment technique, used to assess the emotional reaction to stimuli, considering the valence, arousal and associated dominance of the latter. In this study, only valence and

arousal were considered (see Figure 2, top). Judges were auditorily presented with two sequences of (randomly mixed) 64 vocalizations and 64 music stimuli. They were required to score each stimulus for the two dimensions. The individual scores underwent two repeated-measures analyses of variance (ANOVAs), one for arousal and one for valence, whose factors were as follows: Timbre (vocalization vs. music) and Emotion (positive vs. negative).

The ANOVA performed on Arousal scores showed the significance of Emotion [ $F_{1,11} = 6.33$ ,  $p < .03$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.37$ ], with higher arousal levels induced by negative (e.g. cries) (0.94,  $SE = 0.09$ ) than positive stimuli (e.g. laughter) (0.83,  $SE = 0.10$ ), as displayed in Figure 2 (left).

The results of the ANOVA carried out on the average scores of valence showed the effect of Timbre [ $F_{1,11} = 11.79$ ,  $p < .0005$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.512$ ], with human vocalizations (0.92,  $SE = 0.05$ ) being judged more positively than the melodies (0.698,  $SE = 0.06$ ). Also significant was the factor Emotion [ $F_{1,11} = 82.08$ ,  $p < .001$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.88$ ] with positive stimuli being indeed evaluated as more positive in valence (1.09,  $SE = 0.06$ ) than negative stimuli (0.52,  $SE = 0.05$ ). Mean values of stimulus valence are displayed in Figure 2 (right).

Finally, the interaction of Timbre  $\times$  Emotion [ $F_{1,11} = 62.32$ ,  $p < .001$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.85$ ] showed that the auditory stimuli evaluated more positively in valence were positive vocalizations (1.44,  $SE = 0.09$ ), followed by positive melodies ( $M = 0.75$ ,  $SE = 0.07$ ), while the less positive were negative vocalizations (0.40,  $SE = 0.06$ ), followed by negative music ( $M = 0.65$ ,  $SE = 0.06$ ).



**FIGURE 2** (Top) Self-Assessment Manikin (SAM), non-verbal pictorial assessment technique used to measure the arousal and valence induced by stimulus perception. Taken and modified from the original figure devised by Bradley and Lang (1994). (Bottom): Left. Mean scores (along with standard deviations) relative to perceived arousal as a function of stimulus type and emotional valence. Right. Mean scores relative to valence as a function of stimulus type and emotional type [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

## 2.4 | Procedure

Participants were seated comfortably in a dark, acoustically and electrically shielded test area, wearing a set of *Sennheiser HD202* headphones. They faced a high-resolution screen located 100 cm from their eyes and were instructed to gaze at the centre of the screen, where a small red circle served as a fixation point, and to avoid any eye or body movements during the recording session. At all time during EEG recording, the same picture of an urban landscape was projected on the screen so that visual processing occurred, and parieto/occipital alpha artefacts were avoided. Stimuli were presented in a random order at the centre of the screen in 10 blocks of 36–40 trials featuring the same number of positive and negative stimuli, and including an unpredictable number of targets (from 3 to 6). Stimuli lasted 1,500 ms and the inter-stimulus interval (ISI) ranged between 1300–400 ms.

The experimental session was preceded by two training sessions that included two runs of 20 non-target stimuli and 4/5 targets each, one for each response hand. Vocalizations and music presented during training were not presented again in the experimental sessions (they were children vocalizations, adult neutral voices and relative music melodies). Ten experimental runs were presented, 5 featuring vocalizations, 5 featuring music. The sequence presentation order varied across subjects. The task consisted in responding as accurately and quickly as possible to the sound of birds in vocalization runs, or the sound of piano, in music runs. Vocalization and music runs were alternated. The beginning of each run was preceded by the presentation of 3 warning signals: Attention, Set, Go lasting 1 s. The hand used was alternated during the recording session. Hand order and the task conditions were counterbalanced across subjects. At the beginning of each session, participants were reminded which response hand would be used.

## 2.5 | EEG recording and analysis

The EEG was continuously recorded from 128 scalp sites arranged according to the international 10–5 system (Oostenfeld & Praamstra, 2001) at a sampling rate of 512 Hz. Horizontal and vertical eye movements were also recorded. Average mastoids served as the reference lead. The EEG and electro-oculogram (EOG) were amplified with a half-amplitude band pass of 0.016–70 Hz. An off-line low-pass filter was set at 30 Hz. Electrode impedance was maintained below 5 k $\Omega$ . EEG epochs were synchronized with the onset of stimulus presentation. The EEG was recorded and analysed using *EEProbe* recording software (*ANT Software*, Enschede, The Netherlands). A computerized artefact rejection criterion was applied before averaging to discard epochs

in which eye movements, blinks, excessive muscle potential or amplifier blocking occurred. The artefact rejection criterion was a peak-to-peak amplitude exceeding 50  $\mu$ V, and the rejection rate was  $\sim$ 5%. ERPs were averaged off-line from  $-100$  ms to 1,500 ms after stimulus onset. ERP components were identified and measured with reference to the average baseline voltage over the interval of  $-100$  to 0 ms at the sites and latencies at which they reached their maximum amplitudes. EEG epochs associated with an incorrect behavioural response were excluded.

Four swLORETA (*standardized weighted Low-Resolution Electromagnetic Tomography*) inverse solutions were applied to surface potentials recorded in the N400 latency range (450–600 ms) in response to the four stimulus types. LORETA is an algorithm that provides discrete linear solutions to inverse EEG problems. It is a reliable localization method considering, simultaneously, (a) magnetic flux covariance, (b) phase synchronization and (c) coherence tomography of electromagnetic signals (separately tested for statistical significance). The resulting solutions correspond to the 3D distribution of neuronal electrical activity that has the maximally similar orientation and strength between neighbouring neuronal populations (represented by adjacent voxels). In this study, an improved version of this algorithm, the standardized weighted (sw)LORETA was used (Palmero-Soler, Dolan, Hadamschek, & Tass, 2007). This version, referred to as swLORETA, incorporates a singular value decomposition-based lead field-weighting method. The source space properties included a grid spacing (the distance between two calculation points) of five points (mm) and an estimated signal-to-noise ratio, which defines the regularization where a higher value indicates less regularization and therefore less blurred results, of three. The use of a value of 3–4 for the computation of the SNR in Tikhonov's regularization produces superior accuracy of the solutions for any inverse problem that is assessed. swLORETA was performed on the grand-averaged group data to identify statistically significant electromagnetic dipoles ( $p < .05$ ) in which larger magnitudes correlated with more significant activation. Dipoles with a magnitude below 2 nA were not significantly active according to swLORETA. The data were automatically re-referenced to the average reference as part of the LORETA analysis. A realistic boundary element model (BEM) was derived from a T1-weighted 3D MRI dataset through segmentation of the brain tissue. This BEM model consisted of one homogeneous compartment comprising 3,446 vertices and 6,888 triangles. *Advanced Source Analysis* (ASA) employs a realistic head model of three layers (scalp, skull and brain) and is created using the BEM. This realistic head model comprises a set of irregularly shaped boundaries and the conductivity values for the compartments between them. Each boundary is approximated by a number of points, which are

interconnected by plane triangles. The triangulation leads to a more or less evenly distributed mesh of triangles as a function of the chosen grid value. A smaller value for the grid spacing results in finer meshes and vice versa. With the aforementioned realistic head model of three layers, the segmentation is assumed to include current generators of brain volume, including both grey and white matter. Scalp, skull and brain region conductivities were assumed to be 0.33, 0.0042 and 0.33, respectively. The source reconstruction solutions were projected onto the 3D MRI of the Collins brain, which was provided by the Montreal Neurological Institute. The probabilities of source activation based on Fisher's  $F$  test were provided for each independent EEG source, whose values are indicated in a "unit" scale (the larger the value, the more significant). It should be however considered the somewhat limited spatial resolution of swLORETA, as compared to other neuroimaging techniques such as MEG or fMRI. Both the segmentation and generation of the head model were performed using the ASA software program *Advanced Neuro Technology* (Zanow & Knosche, 2004).

Three repeated-measures ANOVAs were performed on the magnitude values of activations of the electromagnetic dipoles obtained from swLORETA. For comparing dipoles active/inactive in specific conditions, the under-threshold value of 1 nA was assigned by default to inactive dipoles. "1" was used as default value to signal zero activity so that measurements might undergo analyses of variances. It was used a conservative approach (a number approaching 0, such as 0.0001, might have been used as well). 1 nA was selected because represented an under-threshold value, indicating that a given brain area was not statistically active (the threshold value being 2 nA). One ANOVA was performed considering 1 between-groups factor of variability: stimulus type with 2 levels (music, vocalizations) and 1 within-groups factor: brain area (with 9 levels). The other ANOVA considered 1 between-groups factor of variability: emotional valence (negative, positive) and 1 within-groups factor: brain area (with 9 levels). The third ANOVA performed considered 1 between-groups factor of variability: emotional valence (negative, positive) and 1 within-groups factor: brain area (with 5 levels, excluding those areas that were strictly modality-specific according to the previous ANOVAs).

The amplitude and latency of the central P2 response was measured at C1, Cz, C2 sites between 150 and 250 ms. The anterior N400 mean area amplitude was quantified at Fp1 Afp3h, Afz, Fpz, Afp4h and Fp2 sites in the 450–600 ms time window. The mean area amplitude of the anterior/frontal Late Positivity (LP) was quantified at Aff1, Fz, Afz, Fpz, Aff2 sites in the 1200–1400 ms time window.

ERP data were subjected to multifactorial repeated-measures ANOVAs with 3 within-subjects factors: stimulus type (vocalization, music), valence (negative and positive) and

electrodes (depending on the ERP component of interest). Multiple comparisons of means were performed with Tukey's post hoc test.

The Greenhouse–Geisser correction was applied to compensate for possible violations of the sphericity assumption associated with factors which had more than two levels. The epsilon ( $\epsilon$ ) values and the corrected probability levels (in the case of epsilon < 1) are reported. Partial eta squared values were systematically provided to estimate effect sizes. Behavioural data (response times and percentages of correct responses) were also analysed and underwent repeated ANOVAs as described below.

## 3 | RESULTS

### 3.1 | Behavioural data

An ANOVA carried out on individual RTs data as a function of timbre and response hand showed the significance of stimulus type [ $F_{1,14} = 21.17$ ;  $p < .0005$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.65$ ], with faster responses to musical targets (586 ms,  $SD = 36$ ) than vocalizations (652 ms,  $SE = 33$ ). Also significant was the factor hand [ $F_{1,14} = 4.38$ ;  $p < .05$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.24$ ], with faster RTs emitted with the right (605 ms,  $SE = 36$ ) than left hand (634 ms,  $SE = 33$ ).

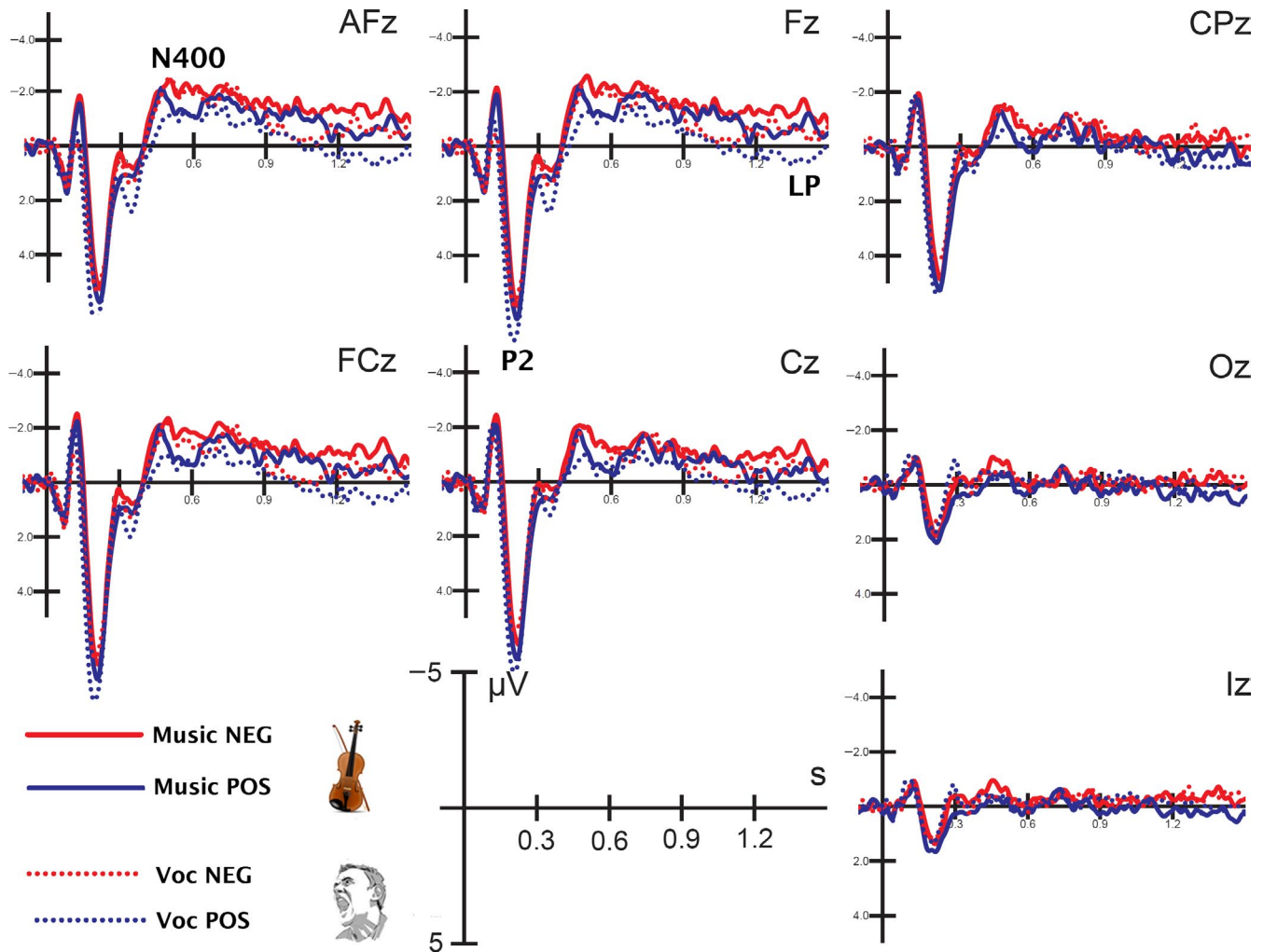
The percentage of correct target recognitions for vocalizations was 98.03% and for musical stimuli was 100%. An ANOVA carried out on individual data as a function of timbre and response hand did not show any significance.

### 3.2 | Electrophysiological data

Figure 3 shows grand-average ERP waveforms recorded at anterior, central and posterior scalp sites as a function of stimulus type and emotional valence. It can be noted that both factors modulated ERP responses at anterior sites after about 200 ms, while the sensory analysis indexed by earlier responses did not show any perceptual difference across stimuli.

The ANOVA applied to P2 latency values (150–250 ms) showed the significance of stimulus type [ $F_{1,14} = 5.24$ ;  $p < .04$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.27$ ] with earlier responses to vocalizations (203 ms,  $SE = 0.004$ ) than music (210 ms,  $SE = 0.003$ ), as visible in Figure 4a.

Repeated-measures ANOVA applied to P2 amplitude values showed the significance of valence factor [ $F_{1,14} = 23.93$ ;  $p < .001$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.63$ ] with greater P2 responses to positive (7.14  $\mu\text{V}$ ,  $SE = 0.85$ ) than negative stimuli (6.02  $\mu\text{V}$ ,  $SE = 0.68$ ), regardless of stimulus type (Figure 4b and waveforms of Figure 5). Also significant was the factor electrode [ $F_{1,14} = 6.41$ ;  $p < .005$ ;  $\epsilon = 0.966$ ,



**FIGURE 3** Grand-average ERPs recorded at anterior/frontal, fronto/central, central, central/parietal, occipital inion midline sites as a function of stimulus valence (positive vs. negative) and type (human voice vs. violin) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

corrected  $p$  value = .005;  $\eta_p^2 = 0.31$ ] with larger P2 responses over the left (C1 = 6.94  $\mu\text{V}$ , SE = 0.71) than the right hemisphere (C2 = 6.68  $\mu\text{V}$ , SE = 0.84), as proved by post hoc comparisons.

Repeated-measures ANOVA applied to N400 amplitude values between 450–600 ms showed the significance of valence factor [ $F_{1,14} = 12.29$ ;  $p < .002$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.51$ ], with greater amplitude N400 to negative stimuli ( $-2.14 \mu\text{V}$ ,  $SD = 0.33$ ) than positive stimuli ( $-1.38 \mu\text{V}$ , SE = 0.33), as visible in Figure 4c and waveforms of Figure 5. Topographical maps of Figure 6 show the scalp distribution of N400 response, much greater to negative stimuli, regardless of stimulus type.

The ANOVA performed on *late positivity* mean area values recorded in the 1,200–1,400 ms time window gave rise to the significance of stimulus type [ $F_{1,14} = 6.83$ ;  $p < .020$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.33$ ] with greater LP responses to vocalizations ( $-0.24 \mu\text{V}$ ,  $SD = 0.36$ ) than music ( $-1.03 \mu\text{V}$ ,  $SD = 0.30$ ). This effect is visible in Figure 7. Also significant was the stimulus emotional valence [ $F_{1,14} = 26.36$ ;  $p < 0.001$ ;  $\epsilon = 1$ ;  $\eta_p^2 = 0.65$ ] with greater responses to positive ( $-0.2 \mu\text{V}$ ,

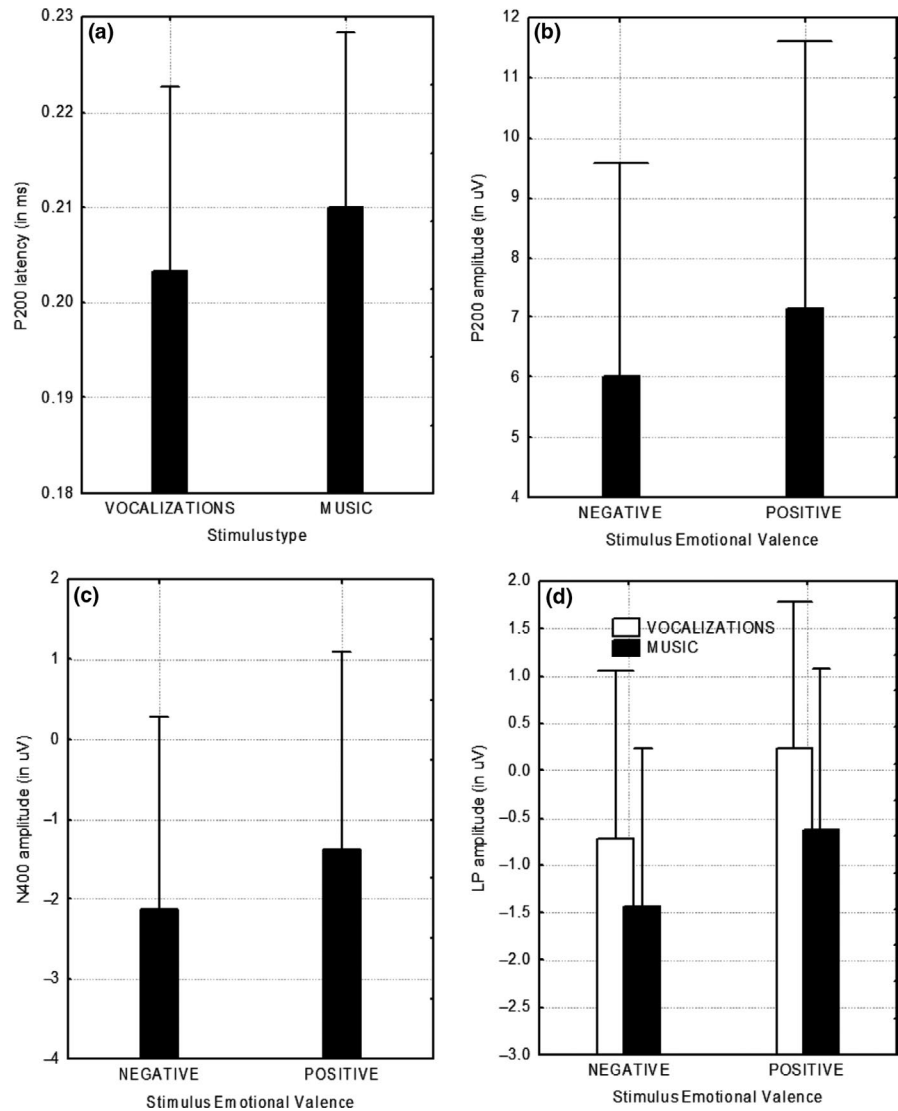
$SD = 0.28$ ) than negative stimuli ( $-1.08 \mu\text{V}$ ,  $SD = 0.31$ ). Both effects can be appreciated in Figure 4d.

### 3.3 | Source reconstruction

Four swLORETA (standardized weighted Low-Resolution Electromagnetic Tomography) inverse solutions were applied to surface potentials recorded in the N400 latency range (450–600 ms) in response to the four stimulus types (Figure 8). The magnitude of the activation of active electromagnetic dipoles explaining the N400 surface potentials elicited by the four stimulus types underwent repeated measures ANOVAs. They showed that some of the areas were active for all conditions, while others were specific to stimulus material (vocal vs. instrumental) or to the emotional content of the stimuli (negative vs. positive in valence). In Table 1 are listed the swLORETA activations that were found in both stimulation conditions, that is music and vocalization. The ANOVA performed

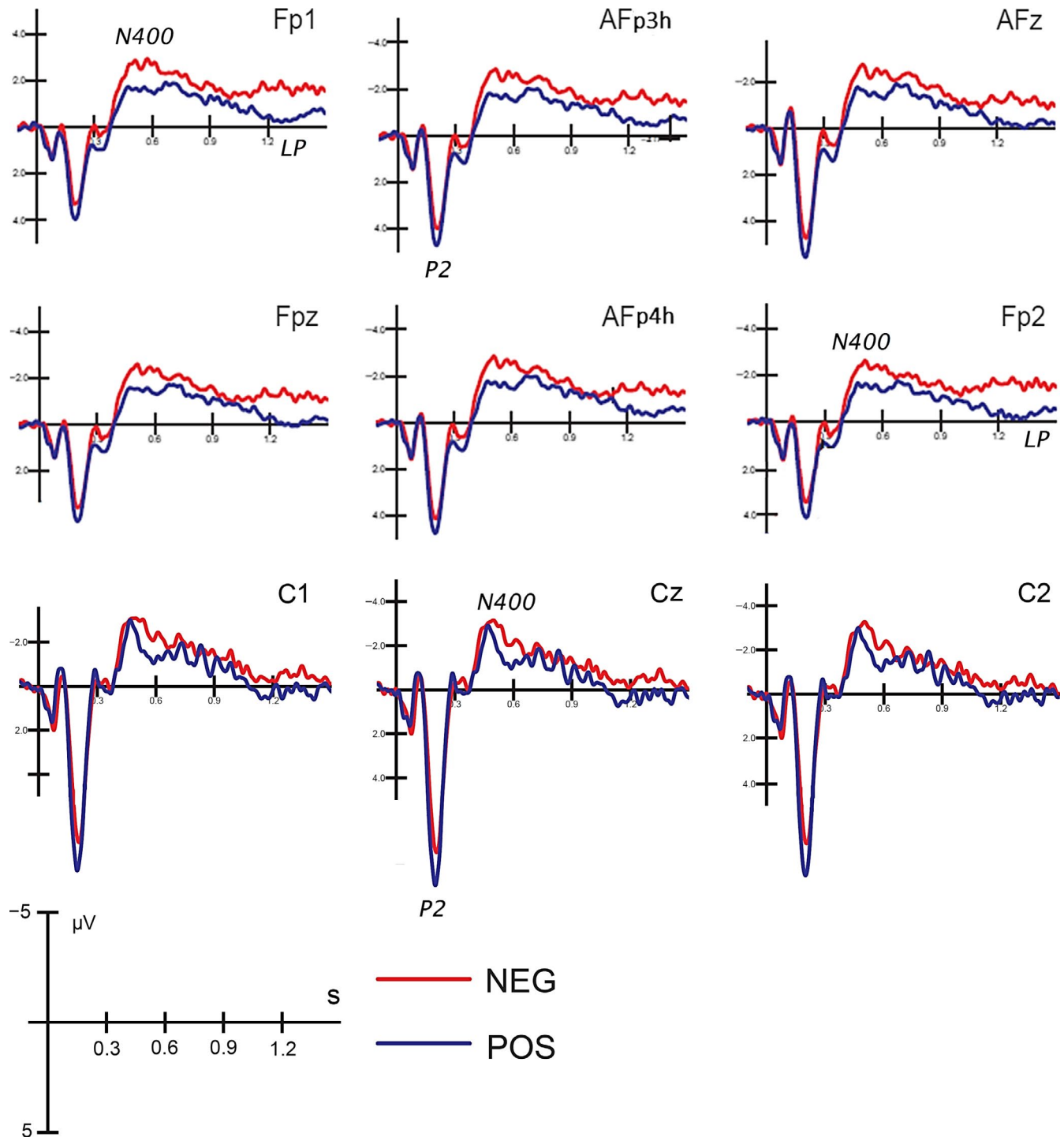


**FIGURE 4** (a) Mean latency values of P2 response (in ms) as a function of stimulus type. (b) Mean amplitude values of P2 amplitude (in  $\mu\text{V}$ ) as a function of stimulus emotional valence. (c) Mean amplitude values of N400 amplitude (in  $\mu\text{V}$ ) as a function of stimulus emotional valence. (d) Mean amplitude values of LP amplitude (in  $\mu\text{V}$ ) as a function of stimulus type and emotional valence



by considering the effect of stimulus type gave rise to the significance of area factor [ $F_{8,16} = 11.39$ ;  $p < .00005$ ;  $\eta_p^2 = 0.85$ ] with the strongest activation being found in the left superior frontal gyrus (BA8) (ISFG = 24.68 nA,  $SD = 4.74$ ) and the left middle temporal gyrus (BA21) (IMTG = 12.22 nA,  $SD = 1.93$ ) as can be appreciated in Figure 9. The further significance of area X stimulus type [ $F_{8,16} = 2.67$ ;  $p < .05$ ;  $\eta_p^2 = 0.57$ ] and relative post hoc comparisons showed that, while some areas were more active to vocalizations or to music, others did not significantly vary their activation as a function of stimulus type, namely the ISFG, the left and right MTG, the right uncus and the rIFG (see Figure 9). The right parahippocampal region of the limbic lobe (rPPH\_BA34) was specifically active for all music types, along with the right cingulate cortex (rCC\_BA23), as displayed in Figure 8a. Again, the left superior temporal cortex (ISTG\_BA39) and the left uncus (lUnc\_BA28) were strongly active during processing of human vocalizations and not musical stimuli as visible in Figure 8b.

The ANOVA performed on the magnitudes of activation of the electromagnetic sources as a function of emotional valence and brain area gave rise to the significant effect of valence [ $F_{1,2} = 74.45$ ;  $p < .01$ ;  $\eta_p^2 = 0.97$ ] with stronger signals to negative than positive stimuli regardless of stimulus type. The factor brain area also yielded significance [ $F_{8,16} = 26.34$ ;  $p < .0005$ ;  $\eta_p^2 = 0.79$ ]. The analyses of simple effects carried out on the valence X area factor proved that some areas were indeed not statistically sensitive to emotional valence (the values are listed in Table 2 for rPPH\_BA34, rCC\_BA23, lUnc\_BA28, ISTG\_BA39). The repeated-measures ANOVA performed as a function of emotional valence and brain areas (which were significantly sensitive to valence in the previous ANOVA) showed again the significance of both valence ( $p < .01$ ) and brain area ( $p < .00005$ ), but also of the interaction of valence X emotion [ $F_{4,8} = 7.75$ ;  $p < .008$ ;  $\eta_p^2 = 0.79$ ]. The results are depicted in Figure 10. Post hoc comparisons showed that while all areas were more strongly active to negative than positive stimuli, only the IFG was more responsive to positive than negative vocalizations (see Figure 10 and Table 2), as also depicted in Figure 8d.

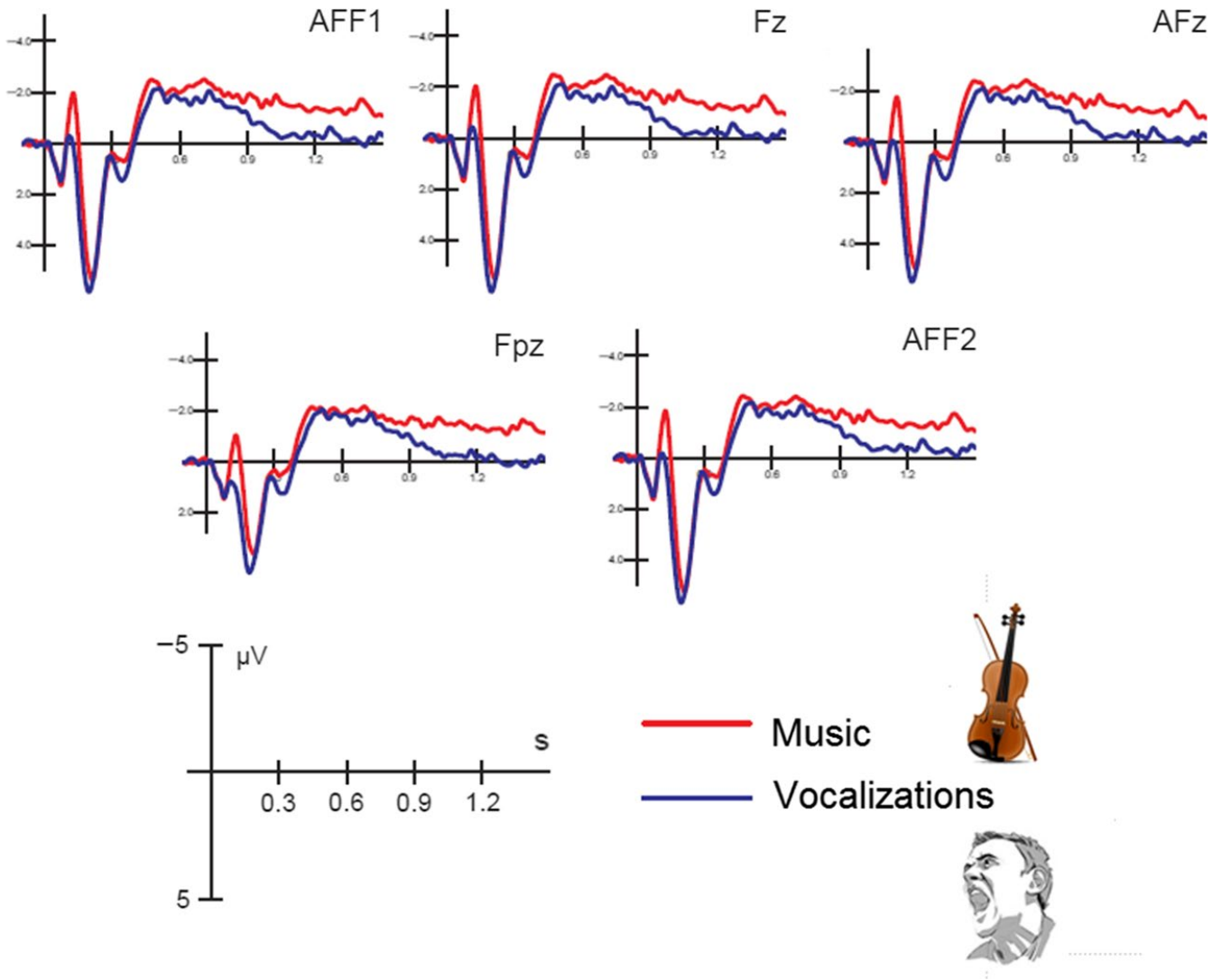
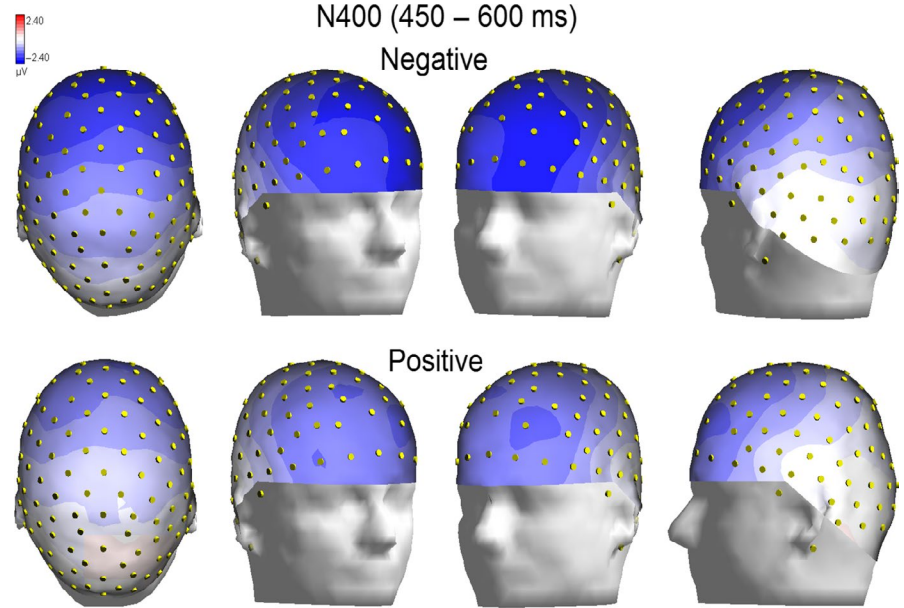


**FIGURE 5** Grand-average ERPs recorded at left and right anterior/frontal, prefrontal and central sites to negative or positive stimuli, showing the much larger N400 response to the former, regardless of stimulus type [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

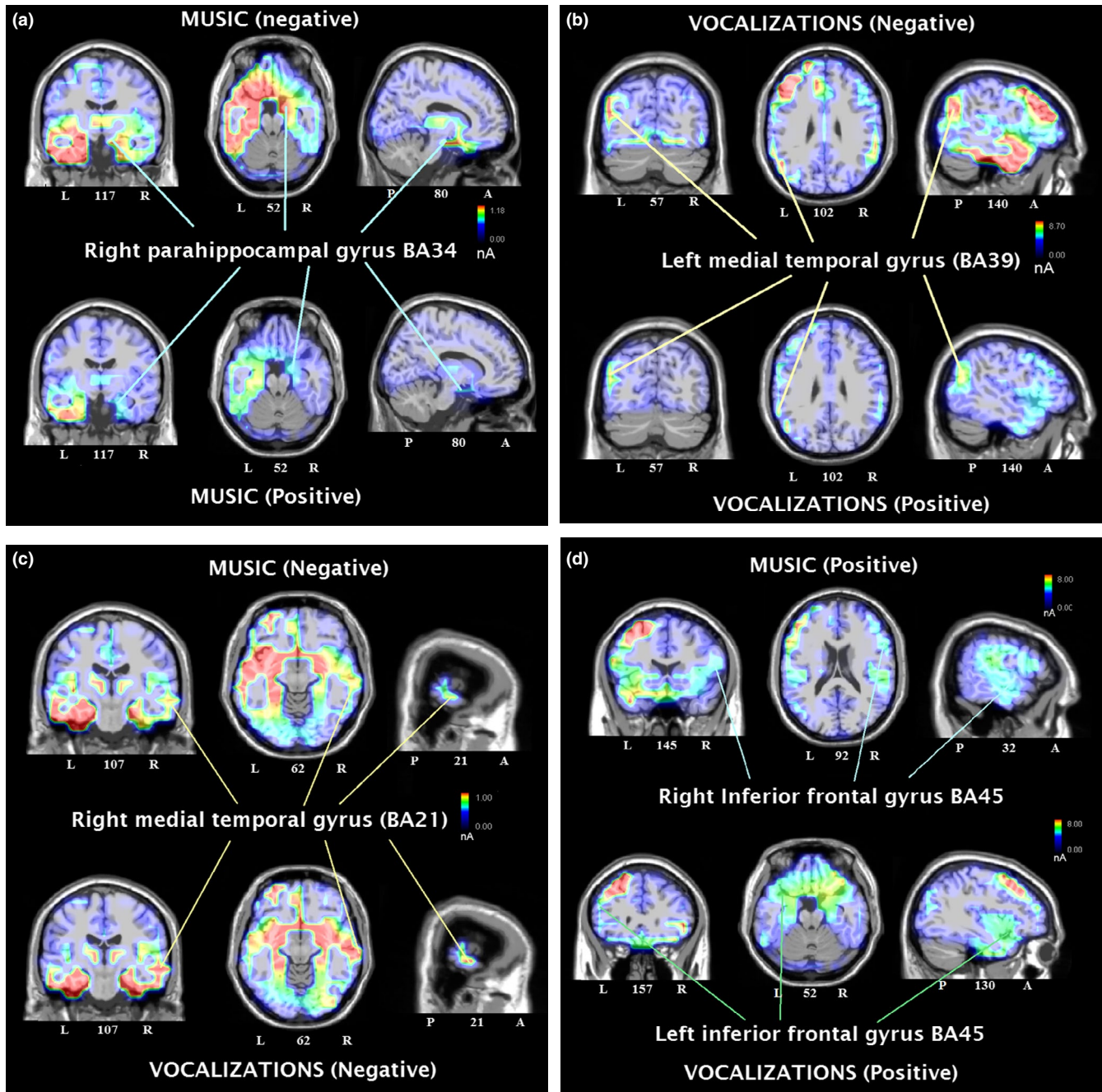
Overall, the ISFG was the strongest generator in any listening condition, with a much greater magnitude (in nA) for negative than positive emotional content. In addition, the IMTG was always active during listening, and with greater amplitude for music than human voice, and for negative than positive stimuli. Lastly, the activation of the right uncus (BA36) of the limbic cortex was also found in common with all listening conditions, and again with a much greater magnitude (in

nA) for negative than positive stimuli. Conversely, a series of regions (reported in Table 2) were specifically active for selected stimulus types and emotional contents. As for the emotional content of stimuli, negatively valenced stimuli always activated the right middle temporal gyrus (BA21), where positively valenced stimuli always activated the inferior frontal cortex (BA45), over the right for the music and over the left for the human voice, as displayed in Figure 8c,d, respectively.

**FIGURE 6** Topographical distribution of surface voltage recorded in the 450–600 time window corresponding to N400 response. The negative response was visibly of greater amplitude to negatively than positively valenced stimuli, regardless of stimulus type (human voice or violin) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 7** Grand-average ERPs recorded at anterior sites to vocalizations or music, showing the much larger LP response to the former, regardless of stimulus valence [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 8** (a) Coronal, axial and sagittal brain images of swLORETA applied to N400 potential and highlighting the active common electromagnetic dipoles for the two types of music stimuli (negative and positive). (b) Coronal, axial and sagittal brain images of swLORETA highlighting the active common electromagnetic dipoles for the two types of vocal stimuli (negative and positive). (c) Coronal, axial and sagittal brain images of swLORETA highlighting the active common electromagnetic dipoles for the two types of negative stimuli (vocalizations and music). (d) Coronal, axial and sagittal brain images of swLORETA highlighting the active common electromagnetic dipoles for the two types of positive stimuli (vocalizations and music). The different colours represent differences in the magnitude of the electromagnetic signal (in nA) recorded in the specific time window. Numbers refer to the displayed brain slice in sagittal view: the left section belongs to the right hemisphere and the right one to the left hemisphere. A = anterior, P = posterior [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

## 4 | DISCUSSION

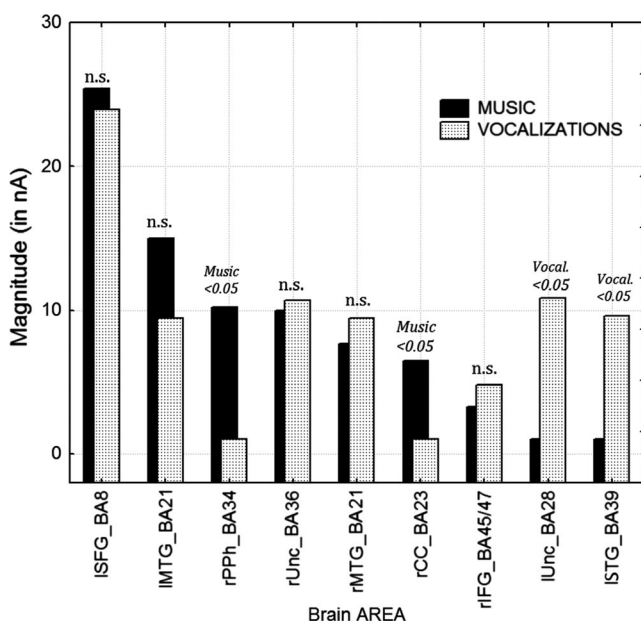
The aim of this research was to compare the neural processes involved in the coding of musical and non-verbal vocal stimuli, and to observe any commonality or difference

that concerned the coding of the emotional content of information. To this end, violin melodies were created from spontaneous emotional vocalizations, so that they shared the same characteristics in terms of frequency, intensity, melodic and rhythmic profile. Although extracted from

**TABLE 1** The table lists the 3 strongest electromagnetic dipoles explaining the surface voltage recorded in the N400 latency range (450–600 ms) in response to the four stimulus types and ordered as a function of signal magnitude (in nA), according to swLORETA. For these areas, the ANOVA performed on magnitude values did not yield a significant effect of stimulus type (music vs. vocalizations)

Magn.	Hem.	Lobe	Gyrus	BA	Stimulus type	Presumed function
31.97	Left	Frontal	SFG	8	Vocalizations (negative)	Encoding, working memory, selective attention
30.47	Left	Frontal	SFG	8	Music (negative)	
20.38	Left	Frontal	SFG	8	Music (positive)	
15.29	Left	Frontal	SFG	8	Vocalizations (positive)	
17.88	Left	Temporal	MTG	21	Music (negative)	Music analysis (Brown et al., 2004; Proverbio et al., 2016); speech and semantic processing (Friederici, Opitz, & von Cramon, 2000)
12.08	Left	Temporal	MTG	21	Music (positive)	
12.01	Left	Temporal	MTG	21	Vocalizations (negative)	
6.91	Left	Temporal	MTG	21	Vocalizations (positive)	
14.26	Right	Limbic	Uncus	36	Vocalizations (negative)	Affective processing (Zeidman & Maguire, 2016); Affective relevance (Austin, 2012; Proverbio et al., 2016)
12.08	Right	Limbic	Uncus	36	Music (negative)	
7.90	Right	Limbic	Uncus	36	Music (positive)	
7.09	Right	Limbic	Uncus	36	Vocalizations (positive)	

Note: H., hemisphere; L, lobe; Magn., magnitude. In reviewing dipole localizations and approximate Brodmann areas (the estimated spatial accuracy of swLORETA being 5 mm) it should be considered the limited spatial resolution of the technique, combining EEG data with MRI images.



**FIGURE 9** Mean magnitude values (in nA) of swLORETA active dipoles, as a function of stimulus type (regardless of the emotional valence of stimuli). “Music” on the top of bar indicates a significantly greater magnitude for musical than vocal stimuli, and vice versa for “Vocal”; n.s.= not significant difference across conditions (music vs. vocalizations). ISFG\_BA9 = left superior frontal gyrus; lMTG\_BA21 = left middle temporal gyrus; rPPh\_BA34 = right parahippocampal gyrus; rUnc\_BA36 = right uncus; rMTG\_BA21 = right middle temporal gyrus; rCC\_BA23 = right cingulate cortex; rIFG\_BA45/47 = right inferior frontal gyrus; lUnc\_BA28 = left uncus; lSTG\_BA39 = left superior temporal cortex

real vocalizations violin stimuli were all in all musical fragments, with a violin playing a given melodic contour. No subject recognized a similarity between the two classes

of stimuli, which appear novel and original during listening. Due to the implicit nature of the task (participants paid attention to other targets and ignored both human vocalizations and violin music), any difference in ERP waveforms related to the emotional content of these stimuli was the result of an automatic extraction of emotional cues from unattended auditory information.

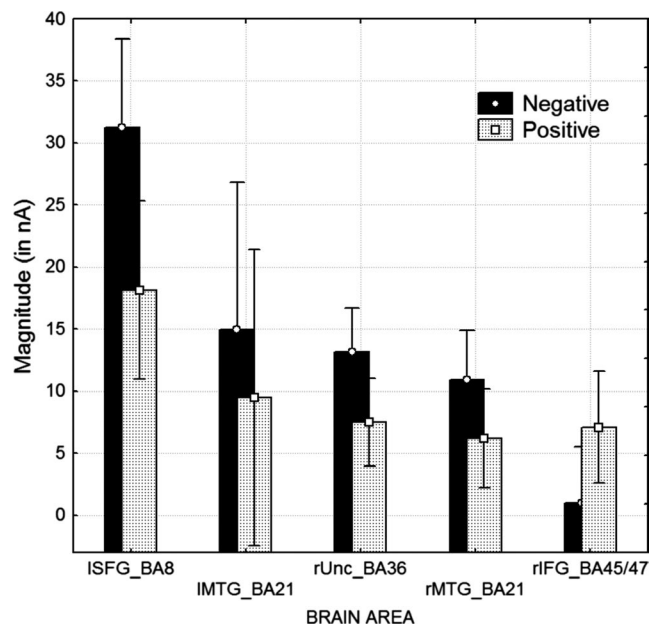
While human vocalizations were clearly distinguishable in positive and negative ones (in the validation test), violin excerpts were more difficult to be interpreted, although still significantly discriminated on the basis of their valence. On the other hand, positive and negative stimuli were clearly discriminable in term of the arousal induced in the listener, indifferently whether vocal or musical in nature. Musical stimuli were analysed according to the same categories of vocal stimuli, in the aim of ascertain whether ERP responses were modulated in a similar fashion. We hypothesized that music and vocalizations used patterns of similar acoustic cues to express emotions (Juslin & Laukka, 2003; Pasquette et al., 2018), and these cues were automatically and inattentively processed by the listeners.

The earlier ERP response modulated by stimulus content was P2 (150–250 ms), recorded over central sites (where auditory stimuli normally elicit the larger evoked sensorial response) and generally linked to attention (Kisley & Cornwell, 2006) and perceptual recognition. According to the available literature, auditory P2 would not be only sensitive to perceptual parameters, such as intensity or pitch sounds, but also to stimulus cognitive properties, attentional valence (Crowley & Colrain, 2004), sound familiarity, exposure and auditory learning (Tremblay et al., 2014). In this vein, it is conceivable that auditory P2 might be modulated by stimulus emotional valence and relevance.

**TABLE 2** Distinctive swLORETA activations obtained in the N400 latency range (450–600 ms) for the different stimulus types and emotional content, according to ANOVAs

Magn.	Hem.	Lobe	Gyrus	BA	Stimulus type	Presumed function
<b>Music</b>						
12.41	Right	Limbic	Parahippocampal	34	Music (negative)	Affective imaginations and episodic memories (Zeidman & Maguire, 2016); Music listening (Leaver et al., 2009)
8.01	Right	Limbic	Parahippocampal	34	Music (positive)	
7.41	Right	Limbic	Cingulate cortex	23	Music (negative)	Music listening (Brown et al., 2004); sad music (Khalfa, Schon, Anton, & Liégeois-Chauvel, 2005)
5.55	Right	Limbic	Cingulate cortex	23	Music (positive)	
<b>Vocalizations</b>						
10.65	Left	Temporal	Wernicke area	39	Vocalizations (negative)	Speech comprehension (e.g. Turken & Dronkers, 2011); Takeichi et al., 2010)
8.52	Left	Temporal	Wernicke area	39	Vocalizations (positive)	
7.95	Left	Limbic	Uncus	28	Vocalizations (negative)	Language (Vassal et al., 2016)
3.88	Left	Limbic	Uncus	28	Vocalizations (positive)	
<b>Negative emotion</b>						
9.82	Right	Temporal	Middle Temporal	21	Music (negative)	Negative affect, negative vocalizations (human and animal, Belin et al., 2008); sad music with lyrics (Brattico et al., 2011)
12.48	Right	Temporal	Middle Temporal	21	Vocalizations (negative)	
<b>Positive emotion</b>						
5.61	Right	Frontal	Inferior Frontal	45	Music (positive)	Positive vocalizations (also animal ones, Belin et al., 2008). Positive music (Warren et al., 2006)
6.70	Left	Frontal	Inferior Frontal	45	Vocalizations (positive)	

Note: Magn., magnitude; Hem., hemisphere.



**FIGURE 10** Mean magnitude values (in nA) of swLORETA active dipoles, as a function of the emotional valence of auditory stimuli (regardless of stimulus type). The differences across emotional categories were all statistical significant in this ANOVA, with all areas showing enhanced activity to negative stimuli, and the right IFG showing enhanced activity for positive stimuli

In our study, positive stimuli (regardless of timbre: human voice or violin) elicited larger P2s than negative stimuli, thus possibly indexing a common response for the recognition of

emotional valence of auditory stimuli. A similar pattern of results has been provided by studies involving verbal language, with positively valenced words eliciting significantly larger P200s than neutral words (Kanske & Kotz, 2007). Again, in a study by Paulmann et al. (2013) investigating the effects of emotional prosody on the neural processing of sentences spoken in six different emotions (anger, disgust, fear, happiness, surprise, sadness) or in a neutral tone of voice it was found that high arousing stimuli elicited a greater amplitude P200 than low arousing stimuli. This similarity hints at a common mechanism for extracting the positive valence of vocal and, indeed, auditory sounds (music) on the basis of stimulus melodic structure (see Figure 11).

In our study, vocal stimuli also elicited earlier P2 responses than musical stimuli, possibly because of their greater biological relevance. An interesting paper by Gall and Wilczkynski (2015) has shown that P2 to auditory stimuli (conspecifics' vocalizations) was indeed modulated by social relevance. However, this study was performed in frogs, their P2 occurring in less than 10 ms, which appears as a rather different electrical response.

In our study, the privileged status of vocalizations with respect to music also emerged from validation data. The judges, in fact, reported to be feeling more activated during listening to vocalizations than to music. In addition, in the judgement of valence, positive vocalizations were clearly discriminated from the negative ones and evaluated as more positive than positive music melodies. This sharp separation demonstrates

**FIGURE 11** Notation relative to some examples of musical stimuli, both positive (P) and negative (N), extracted by the type of vocalization indicated (e.g. appreciation or fear). Notation was automatically obtained from acoustic signals by means of AnthemScore 3.6.0 software. It can be noted the in the positive stimuli the melodic or harmonic profile contained intervals typical of major tonality and more consonant chords such as octaves, perfect fifths and major thirds

POSITIVE		NEGATIVE	
<p><b>P1. Appreciation</b></p>	<p><b>N1. Weeping</b></p>		
<p><b>P2. Happiness</b></p>	<p><b>N2. Crying</b></p>		
<p><b>P3. Happiness</b></p>	<p><b>N3. Fear</b></p>		
<p><b>P4. Surprise</b></p>	<p><b>N4. Fear</b></p>		
<p><b>P5. Laughter</b></p>	<p><b>N5. Pain</b></p>		
<p><b>P6. Appreciation</b></p>	<p><b>N6. Sadness</b></p>		

the precision with which vocalizations were analysed and categorized with respect to music melodies, although they shared most of their perceptual characteristics, except for the overall timbre envelope. However, the similar P2 behaviour for music and voice stimuli hints at a common biological mechanism for processing music and voice prosody (Juslin, 1998; Patel & Peretz, 1997).

The N400 recorded between 450 and 600 milliseconds on frontal electrodes was greater to negative than positive stimuli, regardless of stimulus type. The greater amplitude of the N400 for negative vocalizations fits with previous literature showing a N400 modulation for the emotional prosody, meant as patterns of changes in the voice pitch depending on the emotional state of the speaker (Banse & Scherer, 1996). Schirmer, Kotz, and Friederici (2005) presented semantically neutral sentences read with a happy or sad prosody followed by congruous or incongruous emotional words. The results showed larger N400 responses to emotionally incongruent compared to congruent words. In another study (Grossmann et al., 2005), it was shown

how the N400 was more pronounced to linguistic stimuli with a negative rather than a positive prosody. The authors presented auditorily a series of words read with an angry, happy or neutral prosody to children of 7 months and found that the angry prosody evoked larger negative responses (at about 450 ms) than the other conditions. In the present study, the finding that N400 component was greater to negative stimuli regardless of stimulus type (Vocalizations and Melodies) likely suggests that the emotional connotation of music was extracted by means of the same neural mechanism engaged in the processing of speech emotional prosody (Proverbio & Santoni, 2019). This fits with the hypotheses advanced by other authors that music and vocalizations used patterns of similar acoustic cues to communicate emotions (Juslin & Laukka, 2003; Paquette et al., 2018). The long latency ERP potential LP recorded between 1,200 and 1,400 ms over frontal electrodes was sensitive to both stimulus type and emotional valence, thus paralleling with behavioural pattern and conscious experience. LP was of greater amplitude to human vocalizations than to music,

and to positive than negative stimuli. The latter piece of data is similar to what found with speech prosody and semantic content (Carretiè et al., 2008; Hinojosa et al., 2010; Kanske & Kotz, 2007), with positively valenced words typically eliciting enhanced LP responses of ERPs (Bayer & Schacht, 2014; Herbert et al., 2008, 2006) than neutral or negative stimuli. This similarity hints at a common mechanism for extracting the emotional content of speech, vocal stimuli and music.

As for inner brain mechanisms, the analyses of neural sources of N400 brain potentials carried out by means of swLORETA showed interesting commonalities and distinctions in the patterns of brain activation correlated to the processing and recognition of timbre and emotional content of auditory stimuli. In all condition of listening, it was found most active the left superior frontal gyrus (BA8), with a much greater magnitude for negative than positive stimulus emotional content. This area has been described as being active during working memory tasks (du Boisgueheneuc et al., 2006), supporting stimulus encoding (especially the left area, Tulving, Kapur, Craik, Moscovitch, & Houle, 1994) and featuring semantic and verbal comprehension (Gabrieli, Poldrack, & Desmond, 1998) as well as music listening (Flores-Gutiérrez et al., 2007; Platel, Baron, Desgranges, Bernard, & Eustache, 2003). Also active for all stimulus types was the left middle temporal gyrus (BA21) with greater amplitude for music than the human voice, and for negative than positive stimuli. Similarly, the right uncus (BA36) of the limbic cortex was more active to negative than positive stimuli. According to fMRI studies, the left middle temporal gyrus (BA21) is typically one of the strongest sources of activity during musical perception (Ohnishi et al., 2001) especially in musicians, indicating analytic processing. Furthermore, it seems strongly sensitive to the vocal expressions of anger and happiness (Johnstone, Reekum, Oakes, & Davidson, 2006), exactly as in the present study, thus indicating its role in the processing of auditory expression of emotions, conveyed by vocal and musical prosody. As for the uncus activation, it might index stimulus affective relevance, since it is involved in the expression and comprehension of emotional speech (Leitman et al., 2010). As part of the limbic system, the right uncus (BA36) is also more active during perception of emotionally relevant than irrelevant visual stimuli (e.g. opposite versus. own-sex faces, own child versus. others babies faces (Austin, 2012). As for the stronger brain activation of most brain areas (BA8, BA21, BA36) during processing of negative than positive stimuli, this has been consistently found in a variety for stimulation conditions: visual images, sounds, and visual images paired with sounds (Brown & Cavanagh, 2017), emotional words (Wang, Shangquan, & Lu, 2019), affective IAPS scenes (e.g. Proverbio et al., 2009) emotional facial expressions (e.g. Güntekin & Basar, 2007; Proverbio, Brignone, Matarazzo, Zotto, & Zani, 2006); music (Pallesen

et al., 2005). It has been postulated that negative events recruit greater attentional resources than positive events, because of their biological relevance (Delplanque, Lavoie, Hot, Silvert, & Sequeira, 2004; Yuan et al., 2007), which will explain the deeper processing and related enhanced neurometabolic signal.

For all types of music stimuli, the parahippocampal (BA 34) and cingulate gyri (BA 23), of the right hemisphere, were always found to be active, similarly to other studies (Brown, Martinez, & Parsons, 2004; Leaver, Lare, Zielinski, Halpern, & Rauschecker, 2009). These structures are both part of the limbic system, which supports the hypothesis that music in general evokes emotions in listeners. Several studies have shown the involvement of the limbic system in the processing and perception of music: studies on healthy subjects without any musical expertise (Liégeois-Chauvel et al., 2014; Mitterschiffthaler, Fu, Dalton, Andrew, & Williams, 2007), but also studies on patients with lesions of the right hemisphere (Khalifa et al., 2008). The parahippocampal gyrus, in particular, has been associated with the perception of musical unpleasantness, that is dissonance (Blood, Zatorre, Bermudez, & Evans, 1999; Khalifa et al., 2008), and with affective imaginations, episodic memories and mental imagery (Mitterschiffthaler et al., 2007; Mizrak, Singmann, & Öztekin, 2017; Sheldon & Levine, 2018; Zeidman & Maguire, 2016). On the other hand, all vocalizations (either negative or positive) activated the left medial temporal gyrus (BA 39), also known as Wernicke area. It is well known that BA39 area plays an important role in speech processing (e.g. Ardila, Bernal, & Rosselli, 2016; Turken & Dronkers, 2011).

As for the emotional content of information and regardless of timbre, negative stimuli always activated the right middle temporal area (BA21), which the neuroscientific literature has related to the processing of negative affect, negative vocalizations (human and animal, Belin et al., 2008), sad music with lyrics (Brattico et al., 2011), sad versus. happy voices (Ohnishi et al., 2001). This commonality hints at a shared mechanism for extracting negative emotional cues from auditory stimuli. As visible in Figure 11, intervals characteristic of minor modalities typical of sad music (e.g. minor thirds, diminished seventh, etc.) were actually found in the conversion from negative vocalization to music, while intervals typical of major tonality (octaves, perfect fifths, major thirds) were found in music stimuli obtained by converting positive vocalizations into music (see also Curtis & Bharucha, 2010).

Conversely, positive stimuli (regardless of timbre) activated the inferior frontal gyrus (similarly to Warren et al., 2006), and the activation was left sided for vocalizations and right-sided for music.

The present pattern of results also strongly agrees with the fMRI finding of Belin et al. (2008) who compared



animal and human vocalizations, finding a right middle temporal activation for processing negative vocalizations and an inferior frontal activity for processing positive vocalizations. Similarly, an EEG study by Davidson, Schwartz, Saron, Bennett, and Goleman (1979) showed different cortical activation patterns in response to listening to positive or negative melodies: the first showed greater fronto/temporal lateralization on the left; the latter a more lateralized activation on the right. Again, similar patterns of asymmetry have been found for visual emotional stimuli (Canli, Desmond, Zhao, Glover, & Gabrieli, 1998; Davidson, 1992). These findings also agree with the neuropsychological evidence supporting a role of the left hemisphere for positive emotions and of the right hemisphere for negative emotions (Root, Wong, & Kinsbourne, 2006; Rosadini & Rossi, 1967). Neuropsychological findings on brain damaged patients also demonstrated that the right hemisphere is associated with the expression of depressive, catastrophic emotions, whereas the left hemisphere is associated with the expression of positive and euphoric emotions (Ley & Bryden, 1979; Tucker, 1981).

Overall, the data suggest that music and vocalizations use similar patterns of acoustic cues to express emotions (Juslin & Laukka, 2003; Pasquette et al., 2018), which may explain why music is able to convey emotions, in such a universal manner, regardless of cultural factors such as, education, familiarity or aesthetic taste. It can be assumed that the processing of music harmonic and melodic profiles is somewhat innate in that it is sub-served by innate neural mechanisms for comprehending human vocalizations in specialized regions of the middle and superior temporal gyrus.

One possible limit of this study naturally concerns the somewhat limited spatial resolution of swLORETA, as compared to other neuroimaging techniques such as MEG or fMRI. However, the striking similarity of findings with those of previous literature (e.g. the fMRI study by Belin et al., 2006) further supports the validity of the present methodological approach.

## CONFLICT OF INTEREST STATEMENT

The authors declare no competing financial interests.

## ACKNOWLEDGEMENTS

This study was funded by 2017-ATE-0058 grant n° 16940 by University of Milano-Bicocca to AMP, entitled “The processing of emotional valence in language and music: an ERP investigation.” We are extremely grateful to sound engineer Roberto Oldani of *Claudio Abbado Civic School of Music* of Milan for transforming the stimuli in music, and to Christine Parsons and Morten Kringelbach for sharing the vocalization stimuli. We also thank Andrea Orlandi for his technical support.


## AUTHOR CONTRIBUTIONS

AMP conceived and planned the experiment. MG and FDB prepared the stimuli and carried out the EEG recordings. MG and FDB performed statistical analyses and contributed to the interpretation of the results. AMP interpreted the data and took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript.

## DATA AVAILABILITY STATEMENT

Anonymized data and details about preprocessing/analyses will be made available to colleagues if requested.

## ORCID

Alice Mado Proverbio  <https://orcid.org/0000-0002-5138-1523>

## REFERENCES

- Ardila, A., Bernal, B., & Rosselli, M. (2016). How Localized are language brain areas? A review of brodmann areas involvement in oral language. *Archives of Clinical Neuropsychology*, *31*(1), 112–122. <https://doi.org/10.1093/arclin/acv081>
- Aubé, W., Angulo-Perkins, A., Peretz, I., Concha, L., & Armony, J. L. (2015). Fear across the senses: Brain responses to music, vocalizations and facial expressions. *Social Cognitive and Affective Neuroscience*, *10*, 399–407. <https://doi.org/10.1093/scan/nsu067>
- Austin, J. I. (2012). The brain basis of maternal responsiveness: Systematic review and meta-analysis of neural pathways. Doctoral thesis, University of Manchester.
- Bakker, D. R., & Martin, F. H. (2015). Musical chords and emotion: Major and minor triads are processed for emotion. *Cognitive Affective and Behavioral Neuroscience*, *15*(1), 15–31. <https://doi.org/10.3758/s13415-014-0309-4>
- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. <https://doi.org/10.1037/0022-3514.70.3.614>
- Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures. *Brain Research*, *1075*, 151–164. <https://doi.org/10.1016/j.brainres.2005.12.065>
- Bayer, M., & Schacht, A. (2014). Event-related brain responses to emotional words, pictures, and faces—a cross-domain comparison. *Frontiers in Psychology*, *5*, 1106.
- Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1634), 473–481.
- Bidelman, G. M., & Heinz, M. G. (2011). Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing. *The Journal of the Acoustical Society of America*, *130*(3), 1488–1502. <https://doi.org/10.1121/1.3605559>
- Bidelman, G. M., & Krishnan, A. (2009). Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the Human Brainstem. *Journal of Neuroscience*, *29*(42), 13165–13171. <https://doi.org/10.1523/JNEUROSCI.3900-09.2009>
- Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate

- with activity in paralimbic brain regions. *Nature Neuroscience*, 2(4), 382–387. <https://doi.org/10.1038/7299>
- Boisgueheneuc, F. D., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., ... Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: A lesion study. *Brain*, 129(12), 3315–3328. <https://doi.org/10.1093/brain/awl244>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1), 49–59.
- Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. A. (2011). Functional MRI Study of Happy and Sad Emotions in Music with and without Lyrics. *Frontiers in Psychology*, 2, 308. <https://doi.org/10.3389/fpsyg.2011.00308>. eCollection 2011
- Brattico, E., & Pearce, M. T. (2013). The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Art*, 7, 48–61. <https://doi.org/10.1037/a0031624>
- Brown, D. R., & Cavanagh, J. F. (2017). The sound and the fury: Late positive potential is sensitive to sound affect. *Psychophysiology*, 54(12), 1812–1825. <https://doi.org/10.1111/psyp.12959>
- Brown, S., Martinez, M. J., & Parsons, L. M. (2004). Passive music listening spontaneously engages limbic and paralimbic systems. *NeuroReport*, 15(13), 2033–2037. <https://doi.org/10.1097/00001756-200409150-00008>
- Canli, T., Desmond, J. E., Zhao, Z., Glover, G., & Gabrieli, J. D. E. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI. *NeuroReport*, 9(14), 3233–3239. <https://doi.org/10.1097/00001756-199810050-00019>
- Carretié, L., Hinojosa, J. A., Albert, J., López-Martín, S., De La Gándara, B. S., Igoa, J. M., & Sotillo, M. (2008). Modulation of ongoing cognitive processes by emotionally intense words. *Psychophysiology*, 45(2), 188–196. <https://doi.org/10.1111/j.1469-8986.2007.00617.x>
- Cheng, Y., Lee, S.-Y., Chen, H. Y., Wang, P. Y., & Decety, J. (2012). Voice and emotion processing in the human neonatal brain. *Journal of Cognitive Neuroscience*, 24(6), 1411–1419. [https://doi.org/10.1162/jocn\\_a\\_00214](https://doi.org/10.1162/jocn_a_00214)
- Cook, N. D., Fujisawa, T. X., & Takami, K. (2006). Evaluation of the affective valence of speech using pitch substructure. *IEEE Transactions on Audio, Speech, and Language Processing*, 14, 142–151. <https://doi.org/10.1109/TSA.2005.854115>
- 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. 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>
- Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, 80(3), 1–10. [https://doi.org/10.1016/S0010-0277\(00\)00136-0](https://doi.org/10.1016/S0010-0277(00)00136-0)
- Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition*, 20(1), 125–151. [https://doi.org/10.1016/0278-2626\(92\)90065-T](https://doi.org/10.1016/0278-2626(92)90065-T)
- Davidson, R. J., Schwartz, G. E., Saron, C., Bennett, J., & Goleman, D. J. (1979). Frontal versus parietal EEG asymmetry during positive and negative affect. *Psychophysiology*, 16, 202–203.
- Delplanque, S., Lavoie, M. E., Hot, P., Silvert, L., & Sequeira, H. (2004). Modulation of cognitive processing by emotional valence studied through event-related potentials in humans. *Neuroscience Letters*, 356, 1–4. <https://doi.org/10.1016/j.neulet.2003.10.014>
- Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology*, 4, 1–12. <https://doi.org/10.3389/fpsyg.2013.00487>
- Escoffier, N., Zhong, J., Schirmer, A., & Qiu, A. (2013). Emotional expressions in voice and music: Same code, same effect? *Human Brain Mapping*, 34, 1796–1810. <https://doi.org/10.1002/hbm.22029>
- Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., ... Steinschneider, M. (2001). Consonance and dissonance of musical chords: Neural correlates in auditory cortex of monkeys and humans. *Journal of Neurophysiology*, 86, 2761–2788. <https://doi.org/10.1152/jn.2001.86.6.2761>
- Flores-Gutiérrez, E. O., Díaz, J. L., Barrios, F. A., Favila-Humara, R., Guevara, M. Á., del Río-Portilla, Y., & Corsi-Cabrera, M. (2007). Metabolic and electric brain patterns during pleasant and unpleasant emotions induced by music masterpieces. *International Journal of Psychophysiology*, 65(1), 69–84. <https://doi.org/10.1016/j.ijpsycho.2007.03.004>
- Friederici, A. D., Opitz, B., & von Cramon, D. Y. (2000). Segregating semantic and syntactic aspects of processing in the human brain: An fMRI investigation of different word types. *Cerebral Cortex*, 10(7), 698–705.
- Gabrieli, J. D., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences USA*, 95(3), 906–913. <https://doi.org/10.1073/pnas.95.3.906>
- Gall, M. D., & Wilczynski, W. (2015). Hearing conspecific vocal signals alters peripheral auditory sensitivity. *Proceedings of the Royal Society: Biological Sciences*, 282(1808), 20150749–20150749.
- Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, 65(6), 852–858. <https://doi.org/10.1016/j.neuron.2010.03.001>
- Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants? Electric brain responses to emotional prosody. *NeuroReport*, 16(16), 1825–1828. <https://doi.org/10.1097/01.wnr.0000185964.34336.b1>
- Güntekin, B., & Basar, E. (2007). Emotional face expressions are differentiated with brain oscillations. *International Journal of Psychophysiology*, 64(1), 91–100. <https://doi.org/10.1016/j.ijpsycho.2006.07.003>
- Herbert, C., Junghofer, M., & Kissler, J. (2008). Event related potentials to emotional adjectives during reading. *Psychophysiology*, 45(3), 487–498. <https://doi.org/10.1111/j.1469-8986.2007.00638.x>
- Herbert, C., Kissler, J., Junghöfer, M., Peyk, P., & Rockstroh, B. (2006). Processing of emotional adjectives: Evidence from startle EMG and ERPs. *Psychophysiology*, 43(2), 197–206. <https://doi.org/10.1111/j.1469-8986.2006.00385.x>
- Hernandez-Reif, M., Diego, M., & Field, T. (2006). Instrumental and vocal music effects on EEG and EKG in neonates of depressed and non-depressed mothers. *Infant Behavior and Development*, 29(4), 518–525. <https://doi.org/10.1016/j.infbeh.2006.07.008>
- Hinojosa, J. A., Méndez-Bértolo, C., & Pozo, M. A. (2010). Looking at emotional words is not the same as reading emotional words: Behavioral and neural correlates. *Psychophysiology*, 47(4), 748–757. <https://doi.org/10.1111/j.1469-8986.2010.00982.x>
- Hopyan, T., Manno, F. A., Papsin, B. C., & Gordon, K. A. (2015). Sad and happy emotion discrimination in music by children with cochlear implants. *Child Neuropsychology*, 6, 1–15.

- Huron, D. (2013). Why is sad music pleasurable? A possible role for prolactin. *Musicae Sciences*, *150*, 46–58.
- Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, *19*(11), 2579–2594. <https://doi.org/10.1093/cercor/bhp008>
- Johnstone, T., van Reekum, C. M., Oakes, T. R., & Davidson, R. J. (2006). The voice of emotion: An fMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience*, *1*(3), 242–249. <https://doi.org/10.1093/scan/nsl027>
- Juslin, P. N. (1998). A functionalist perspective on emotional communication in music performance (Doctoral dissertation, Uppsala University, 1998). In *Comprehensive summaries of Uppsala dissertations from the faculty of social sciences*, Vol. 78 (pp. 7–65). Uppsala, Sweden: University Library.
- 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, P. N., Liljeström, S., Västfjäll, D., & Lundqvist, L. (2010). How does music evoke emotions? Exploring the underlying mechanisms. In P. N. Juslin & J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 605–642). Oxford, UK: Oxford University Press.
- Kanske, P., & Kotz, S. A. (2007). Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research*, *1148*, 138–148. <https://doi.org/10.1016/j.brainres.2007.02.044>
- Kawakami, A., Furukawa, K., Katahira, K., & Okanoya, K. (2013). Sad music induces pleasant emotion. *Frontiers in Psychology*, *13*, 311. <https://doi.org/10.3389/fpsyg.2013.00311>
- Khalfá, S., Guye, M., Peretz, I., Chapon, F., Girard, N., Chauvel, P., & Liégeois-Chauvel, C. (2008). Evidence of lateralized anteromedial temporal structures involvement in musical emotion processing. *Neuropsychologia*, *46*(10), 2485–2493. <https://doi.org/10.1016/j.neuropsychologia.2008.04.009>
- Khalfá, S., Schon, D., Anton, J. L., & Liégeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport*, *16*, 1981–1984.
- Kisley, M. A., & Cornwell, Z. M. (2006). Gamma and beta neural activity evoked during a sensory gating paradigm: Effects of auditory, somatosensory and cross-modal stimulation. *Clinical Neurophysiology*, *117*(11), 2549–2563. <https://doi.org/10.1016/j.clinph.2006.08.003>
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*, 170–180. <https://doi.org/10.1038/nrn3666>
- Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., & Jacobs, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *NeuroImage*, *81*, 49–60. <https://doi.org/10.1016/j.neuroimage.2013.05.008>
- Leaver, A. M., Van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience*, *29*(8), 2477–2485. <https://doi.org/10.1523/JNEUROSCI.4921-08.2009>
- Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughhead, J., Valdez, J. N., ... Gur, R. C. (2010). It's Not What You Say, But How You Say it: A Reciprocal Temporo-frontal Network for Affective Prosody. *Frontiers in Human Neuroscience*, *26*(4), 19. <https://doi.org/10.3389/fnhum.2010.00019>
- Ley, R. G., & Bryden, M. P. (1979). Hemispheric differences in processing emotions and faces. *Brain and Language*, *7*, 127–138. [https://doi.org/10.1016/0093-934X\(79\)90010-5](https://doi.org/10.1016/0093-934X(79)90010-5)
- Liégeois-Chauvel, C., Bénar, C., Krieg, J., Delbé, C., Chauvel, P., Giusiano, B., & Bigand, E. (2014). How functional coupling between the auditory cortex and the amygdala induces musical emotion: A single case study. *Cortex*, *60*, 82–93. <https://doi.org/10.1016/j.cortex.2014.06.002>
- McAuley, J. D., Henry, M. J., & Tkach, J. (2012). Tempo mediates the involvement of motor areas in beat perception. *Annals of the National Academy of Sciences USA*, *1252*, 77–84. <https://doi.org/10.1111/j.1749-6632.2011.06433.x>
- Minati, L., Rosazza, C., D'Incerti, L., Pietrocini, E., Valentini, L., Scafoli, V., ... Bruzzone, M. G. (2009). Functional MRI/Event-related potential study of sensory consonance and dissonance in musicians and nonmusicians. *NeuroReport*, *20*(1), 87–92. <https://doi.org/10.1097/WNR.0b013e32831af235>
- Mitterschiffthaler, M. T., Fu, C. H. Y., Dalton, J. A., Andrew, C. M., & Williams, S. C. R. (2007). A functional MRI study of happy and sad affective states induced by classical music. *Human Brain Mapping*, *28*(11), 1150–1162. <https://doi.org/10.1002/hbm.20337>
- Mızrak, E., Singmann, H., & Öztekin, I. (2017). Forgetting Emotional Material in Working Memory. *Social Cognitive and Affective Neuroscience*, *13*(3), 331–340.
- Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., ... Imabayashi, E. (2001). Functional Anatomy of Musical Perception in Musicians. *Cerebral Cortex*, *11*, 754–760. <https://doi.org/10.1093/cercor/11.8.754>
- Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. [https://doi.org/10.1016/S1388-2457\(00\)00527-7](https://doi.org/10.1016/S1388-2457(00)00527-7)
- Pallesen, K. J., Brattico, E., Bailey, C., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2005). Emotion processing of major, minor, and dissonant chords: A functional magnetic resonance imaging study. *Annals of the National Academy of Sciences USA*, *106*(4), 450–453. <https://doi.org/10.1196/annals.1360.047>
- Palmero-Soler, E., Dolan, K., Hadamschek, V., & Tass, P. A. (2007). SwLORETA: A novel approach to robust source localization and synchronization tomography. *Physical Medicine Biology*, *52*, 1783–1800. <https://doi.org/10.1088/0031-9155/52/7/002>
- 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 National Academy of Sciences USA*, *1423*, 329–337. <https://doi.org/10.1111/nyas.13666>
- Parsons, C. E., Young, K. S., Craske, M. G., Stein, A. L., & Kringelbach, M. L. (2014). Introducing the oxford vocal (OxVoc) sounds database: A validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Frontiers in Psychology*, *5*, 1–10. <https://doi.org/10.3389/fpsyg.2014.00562>
- Patel, A. D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 191–215). Hove, UK: Psychology Press/Erlbaum (UK) Taylor & Francis.
- Paulmann, S., 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>
- 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., 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. <https://doi.org/10.1080/17470910903135668>
- Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., ... Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings National Academy of Sciences USA*, *107*, 4758–4763. <https://doi.org/10.1073/pnas.0909074107>
- Peretz, I., & R. Zatorre (Eds.) (2003). *The cognitive neuroscience of music*. Oxford, UK: Oxford University Press.
- Platel, H., Baron, J. C., Desgranges, B., Bernard, F., & Eustache, F. (2003). Semantic and episodic memory of music are subserved by distinct neural networks. *NeuroImage*, *20*(1), 244–256. [https://doi.org/10.1016/S1053-8119\(03\)00287-8](https://doi.org/10.1016/S1053-8119(03)00287-8)
- Proverbio, A. M., Brignone, V., Matarazzo, S., Del Zotto, M., & Zani, A. (2006). Gender and parental status affect the visual cortical response to infant facial expression. *Neuropsychologia*, *44*(14), 2987–2999. <https://doi.org/10.1016/j.neuropsychologia.2006.06.015>
- Proverbio, A. M., Manfrin, L., Arcari, L. A., De Benedetto, F., Gazzola, M., Guardamagna, M., ... Zani, A. (2015). Non-expert listeners show decreased heart rate and increased blood pressure (fear bradycardia) in response to atonal music. *Frontiers in Psychology*, *6*, 1–13. <https://doi.org/10.3389/fpsyg.2015.01646>
- Proverbio, A. M., Orlandi, A., & Pisanu, F. (2016). Brain processing of consonance/dissonance in musicians and controls: A hemispheric asymmetry revisited. *European Journal of Neuroscience*, *44*(6), 2340–2356. <https://doi.org/10.1111/ejn.13330>
- Proverbio, A. M., Riva, F., Martin, E., Proverbio, A. M., Adorni, R., Zani, A., & Trestianu, L. (2009). Sex differences in the brain response to affective scenes with or without humans. *Neuropsychologia*, *47*(12), 2374–2388. <https://doi.org/10.1016/j.neuropsychologia.2008.10.030>
- Proverbio, A. M., & Santoni, S. (2019). ERP makers of valence coding in emotional speech processing. (November 5, 2019). iScience, ISCIENCE-D-19-01212. Available at SSRN: doi: 10.2139/ssrn.3480697.
- Rigg, M. G. (1964). The mood effects of music: A comparison of data from earlier investigations. *Journal of Psychology*, *58*, 427–438.
- Root, J. C., Wong, P. S., & Kinsbourne, M. (2006). Left hemisphere specialization for response to positive emotional expressions: A divided output methodology. *Emotion*, *6*(3), 473–483. <https://doi.org/10.1037/1528-3542.6.3.473>
- Rosadini, G., & Rossi, G. F. (1967). On the suggested cerebral dominance for consciousness. *Brain*, *90*(1), 101–112. <https://doi.org/10.1093/brain/90.1.101>
- Schirmer, A., Chen, C. B., Ching, A., Tan, L., & Hong, R. Y. (2013). Vocal emotions influence verbal memory: Neural correlates and interindividual differences. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(1), 80–93. <https://doi.org/10.3758/s13415-012-0132-8>
- 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., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, *24*(3), 442–452. <https://doi.org/10.1016/j.cogbrainres.2005.02.022>
- Sheldon, S., & Levine, B. (2018). The medial temporal lobe functional connectivity patterns associated with forming different mental representations. *Hippocampus*, *28*(4), 269–280. <https://doi.org/10.1002/hipo.22829>
- Sugimoto, T., Kobayashi, H., Nobuyoshi, N., Kiriya, Y., Takeshita, H., Nakamura, T., & Hashiya, K. (2010). Preference for consonant music over dissonant music by an infant chimpanzee. *Primates*, *51*, 7–12. <https://doi.org/10.1007/s10329-009-0160-3>
- Takeichi, H., Koyama, S., Terao, A., Takeuchi, F., Toyosawa, Y., & Murohashi, H. (2010). Comprehension of degraded speech sounds with m-sequence modulation: An fMRI study. *NeuroImage*, *49*(3), 2697–2706. <https://doi.org/10.1016/j.neuroimage.2009.10.063>
- Tan, Y. Z., Ozdemir, S., Temiz, A., & Celik, F. (2015). The effect of relaxing music on heart rate and heart rate variability during ECG GATED-myocardial perfusion scintigraphy. *Complementary Therapies in Clinical Practice*, *21*(2), 137–140. <https://doi.org/10.1016/j.ctcp.2014.12.003>
- Tremblay, K. L., Ross, B., Inoue, K., McClannahan, K., & Collet, G. (2014). Is the auditory evoked P2 response a biomarker of learning? *Frontiers in Systems Neuroscience*, *8*, 28. <https://doi.org/10.3389/fnsys.2014.00028>
- Trost, W., Ethofer, T., Zentner, M., & Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cerebral Cortex*, *22*(12), 2769–2783. <https://doi.org/10.1093/cercor/bhr353>
- Tucker, D. M. (1981). Lateral brain function and the conceptualization of emotion. *Psychological Bulletin*, *89*, 19–46.
- Tulving, E., Kapur, S., Craik, F. I. M., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: positron emission tomography findings. *Proceedings of the National Academy of Sciences*, *91*(6), 2016–2020. <https://doi.org/10.1073/pnas.91.6.2016>
- Turken, A. U., & Dronkers, N. F. (2011). The neural architecture of the language comprehension network: Converging evidence from lesion and connectivity analyses. *Frontiers in System Neuroscience*, *5*, 1–20. <https://doi.org/10.3389/fnsys.2011.00001>
- Vassal, F., Schneider, F., Boutet, C., Jean, B., Sontheimer, A., & Lemaire, J. J. (2016). Combined DTI Tractography and Functional MRI Study of the Language Connectome in Healthy Volunteers: Extensive Mapping of White Matter Fascicles and Cortical Activations. *PLoS ONE*, *11*(3), e0152614. <https://doi.org/10.1371/journal.pone.0152614>
- Virtala, P., Huottilainen, M., Putkinen, V., Makkonen, T., & Tervaniemi, M. (2012). Musical training facilitates the neural discrimination of major versus minor chords in 13-year-old children. *Psychophysiology*, *49*, 1125–1132. <https://doi.org/10.1111/j.1469-8986.2012.01386.x>
- Vuilleumier, P., & Trost, W. (2015). Music and emotions: From enchantment to entrainment. *Annals of the National Academy of Sciences*, *1337*, 212–222.
- Wang, X., Shangquan, C., & Lu, J. (2019). Time course of emotion effects during emotion-label and emotion-laden word processing. *Neuroscience Letters*, *21*(699), 1–7. <https://doi.org/10.1016/j.neulet.2019.01.028>
- Warren, J. E., Sauter, D. A., Eisner, F., Wiland, J., Dresner, M. A., Wise, R. J. S., ... Scott, S. K. (2006). Positive Emotions Preferentially Engage an Auditory-Motor “Mirror” System. *Journal of Neuroscience*, *26*(50), 13067–13075. <https://doi.org/10.1523/JNEUROSCI.3907-06.2006>
- Whitehead, J. C., & Armony, J. L. (2018). Singing in the brain: Neural representation of music and voice as revealed by fMRI. *Human Brain Mapping*, *39*(12), 4913–4924. <https://doi.org/10.1002/hbm.24333>

- Yuan, J., Zhang, Q., Chen, A., Li, H., Wang, Q., Zhuang, Z., & Jia, S. (2007). Are we sensitive to valence differences in emotionally negative stimuli? Electrophysiological evidence from an ERP study. *Neuropsychologia*, *45*(12), 2764–2771. <https://doi.org/10.1016/j.neuropsychologia.2007.04.018>
- Zanow, F., & Knosche, T. R. (2004). ASA-Advanced Source Analysis of continuous and event-related EEG/MEG signals. *Brain Topography*, *16*, 287–290. <https://doi.org/10.1023/B:BRAT.0000032867.41555.d0>
- Zeidman, P., & Maguire, E. A. (2016). Anterior hippocampus: The anatomy of perception, imagination and episodic memory. *Nature Review Neuroscience*, *17*(3), 173–182. <https://doi.org/10.1038/nrn.2015.24>
- Zentner, M., & Eerola, T. (2009). Self-report measures and models. In P. N. Juslin & J. A. Sloboda (Eds.), *Series in affective science. Handbook of music and emotion: Theory, research, applications* (pp. 187–221). New York, NY: Oxford University Press.
- Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, *8*(4), 494–521. <https://doi.org/10.1037/1528-3542.8.4.494>
- Zhao, H., & Chen, A. C. (2009). Both happy and sad melodies modulate tonic human heat pain. *Journal of Pain*, *10*, 953–960. <https://doi.org/10.1016/j.jpain.2009.03.006>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Proverbio AM, De Benedetto F, Guazzone M. Shared neural mechanisms for processing emotions in music and vocalizations. *Eur J Neurosci*. 2020;51:1987–2007. <https://doi.org/10.1111/ejn.14650>