What Is Faster than *Where* in Vocal Emotional Perception

Sara Temudo and Ana P. Pinheiro

Abstract

■ Voices carry a vast amount of information about speakers (e.g., emotional state; spatial location). Neuroimaging studies postulate that spatial ("where") and emotional ("what") cues are processed by partially independent processing streams. Although behavioral evidence reveals interactions between emotion and space, the temporal dynamics of these processes in the brain and its modulation by attention remain unknown. We investigated whether and how spatial and emotional features interact during voice processing as a function of attention focus. Spatialized nonverbal vocalizations differing in valence (neutral, amusement, anger) were presented at different locations around the head, whereas listeners discriminated either the spatial location or emotional quality of the voice. Neural activity was measured with ERPs of the EEG. Affective ratings were collected at the end of the EEG session. Emotional vocalizations elicited decreased N1 but increased P2 and late positive potential amplitudes. Interactions of space and emotion occurred at the salience detection stage: neutral vocalizations presented at right (vs. left) locations elicited increased P2 amplitudes, but no such differences were observed for emotional vocalizations. When task instructions involved emotion categorization, the P2 was increased for vocalizations presented at front (vs. back) locations. Behaviorally, only valence and arousal ratings showed emotion–space interactions. These findings suggest that emotional representations are activated earlier than spatial representations in voice processing. The perceptual prioritization of emotional cues occurred irrespective of task instructions but was not paralleled by an augmented stimulus representation in space. These findings support the differential responding to emotional information by auditory processing pathways. ■

INTRODUCTION

The human voice is one of the most salient sounds in our acoustic environment. Voices carry not only speech information but also relevant cues about the speaker (Belin et al., 2004). Among the vast amount of information communicated by voices, emotional cues are particularly relevant, allowing the listener to make inferences about the speaker's intentions and current mood (Pinheiro, Lima, Albuquerque, Anikin, & Lima, 2019; Ceravolo et al., 2016a). It is therefore not surprising that emotional vocal cues are prioritized during voice perception (e.g., Pinheiro, Barros, & Pedrosa, 2016): They are automatically detected even when task irrelevant (e.g., Pinheiro, Barros, Dias, & Kotz, 2017) and grab more attention than neutral cues (e.g., Pinheiro et al., 2017).

Current neurobiological models postulate that vocal emotional comprehension is a multistage process involving the sensory analysis of acoustic information, the detection of the emotional salience of the voice, and the cognitive evaluation of its emotional significance (Frühholz, Trost, & Kotz, 2016; Schirmer & Kotz, 2006). However, these models have mainly focused on the mechanisms involved in emotional meaning decoding ("what"), failing to consider the role of other voice dimensions that contribute to perceived sound relevance. For example, the spatial source ("where") of an auditory object may confer emotional salience depending on its location with respect to the listener's body (Asutay & Västfjäll, 2015). However, very little is known about the impact of spatial location on emotional decoding or about the impact of emotion on sound localization.

Imagine walking down the street and suddenly hearing a scream. In this situation, it is unarguably relevant to quickly identify whether the speaker is expressing fear but detecting the spatial location of the speaker might determine how far we are from a potential danger from which we might want to get away from. Auditory spatial information plays a critical role in our everyday lives, for example, by guiding attention toward salient events in the environment (Derey, Rauschecker, Formisano, Valente, & de Gelder, 2017). Several lines of neurophysiological, anatomical, and neuropsychological evidence suggest that auditory cortices can be separated into dissociable and parallel cortical pathways processing identity (anterolateral "what") and spatial (caudolateral "where") cues (Zündorf, Lewald, & Karnath, 2016; Kryklywy, Macpherson, Greening, & Mitchell, 2013; Ahveninen et al., 2006; Clarke, Bellmann, Meuli, Assal, & Steck, 2000), which share similarities with the visual system.

Although the processing of emotional and spatial features appears to rely on distinct brain pathways, the extent to which these streams interact is still a matter of debate. Behavioral evidence suggests that spatial location may enhance sound salience, affecting its perceived arousal (Tajadura-Jiménez, Larsson, Väljamäe, Västfjäll, & Kleiner,

---

University of Lisbon

Journal of Cognitive Neuroscience 37:2, pp. 239–265
https://doi.org/10.1162/jocn_a_02251

2010) and valence (Asutay & Västfjäll, 2015), and modulating how attentional resources are engaged (Ceravolo et al., 2016a). This may explain the facilitated detection of emotional sounds presented in the rear (vs. front) space (Asutay & Västfjäll, 2015). More recent findings suggest that emotion and space interactions in voice perception are modulated by task demands (Pinheiro et al., 2019): Spatial discrimination of vocalizations was affected by the emotional quality of the voice even when emotion was task irrelevant (Pinheiro et al., 2019), but emotion categorization was unaffected by spatial manipulations (Pinheiro et al., 2019). However, it remains to be clarified how spatial and emotional properties of the voice interact at the brain level. The interaction between "what" and "where" pathways may facilitate the perception of relevant auditory information, particularly when visual cues are not available. For instance, considering the example described above, is the processing of the speaker's location facilitated when we hear a scream compared with a laugh or, in turn, is emotional decoding easier if the speaker is located behind versus in front of us?

ERPs of the EEG are particularly informative in the quest to clarify whether and when emotional and spatial features of vocalizations interact at a subsecond time resolution (Figure 1). Studies probing separate effects of emotion and space confirm the sensitivity of the N1, P2, and late positive potential (LPP) to modulations of these vocal features. The N1 component occurs within the first 100 msec poststimulus onset and reflects the initial sensory analysis of the acoustic properties of the voice (Liu et al., 2012; Kotz & Paulmann, 2011). The P2 component, peaking approximately at 200 msec after stimulus onset, is thought to reflect the automatic detection of emotional salience (Liu et al., 2012). Later evoked potentials such as the LPP are typically observed 400 msec poststimulus onset and reflect later cognitive operations of stimulus evaluation, including its emotional significance (Pell et al., 2015; Kotz & Paulmann, 2011). When comparing emotional and neutral vocal stimuli, differences were found on the N1 (Liu et al., 2012; Kotz & Paulmann, 2011), P2, and LPP amplitudes (Pell et al., 2015; Paulmann, Bleichner, & Kotz, 2013; Kotz & Paulmann, 2011). These ERP components are also modulated by spatial manipulations (N1: Valdés-Conroy, Sebastián, Hinojosa, Román, & Santaniello, 2014; Lewald & Getzmann, 2011; P2: Xie et al., 2014; Lewald & Getzmann, 2011; LPP: Valdés-Conroy et al., 2014). Earlier components (N1, P2) were found to be sensitive to changes in the location of nonvocal sounds (Lewald & Getzmann, 2011). In later processing stages, the LPP was interactively modulated by emotion and space in the visual modality (Valdés-Conroy et al., 2014). Nonetheless, the temporal dynamics of emotion and spatial cues in voice perception remains unknown. It is also unclear if selective attention to either spatial or emotional features of the voice facilitates sound localization and emotion recognition by modulating the "what" and "where" pathways in a feature-specific fashion.

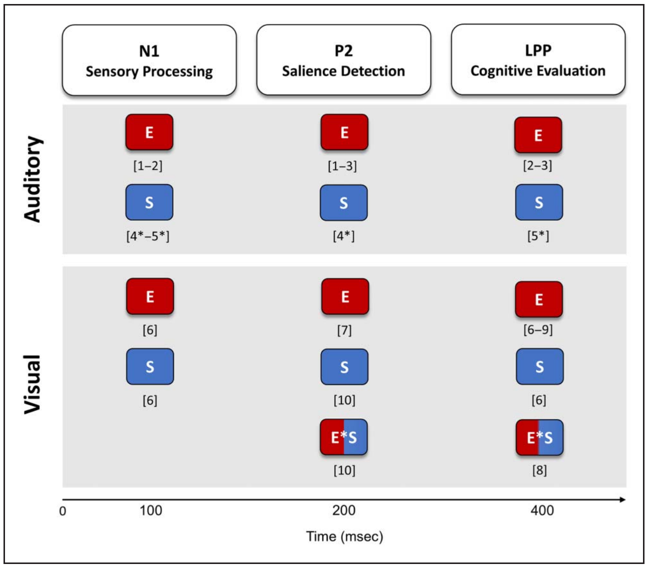

Figure 1. Schematic illustration of emotion (E), space (S), and Emotion × Space interaction (E × S) effects on the N1, P2, and LPP components reported in the auditory and visual modalities. [1] Liu et al. (2012); [2] Pell et al. (2015); [3] Paulmann et al. (2013); [4*] Lewald and Getzmann (2011); [5*] Koiwa, Masaoka, Kusumi, and Homma (2010)*Slow Wave (SW); [6] Valdés-Conroy et al. (2014); [7] Kanske and Kotz (2007); [8] Du et al. (2017); [9] Van Strien et al. (2010); [10] Xie, Wang, and Chang (2014). *Studies in the auditory modality using nonvocal stimuli.

To address these questions, the current study examined the independent and interactive effects of the emotional quality and spatial source of vocalizations under distinct task requirements, combining ERP and behavioral methods in a between-subjects design. Spatialized nonverbal vocalizations portraying either positive (amusement), negative (anger), or neutral states were presented in different locations around the head (left vs. right; front vs. back), while participants performed one of two tasks: spatial location discrimination or emotion recognition. The same vocalizations were presented in both tasks.

The ERP analysis was focused on the N1, P2, and LPP components. On the basis of previous ERP studies, amplitude modulations were expected as a function of emotional salience (Liu et al., 2012; Kotz & Paulmann, 2011). Specifically, spatialized emotional vocalizations should elicit reduced N1 (Liu et al., 2012) and increased P2 and LPP amplitudes (Pell et al., 2015; Paulmann et al., 2013; Liu et al., 2012) compared with neutral ones.

Previous studies revealed that the spatial source of a sound also confers information on the salience of an auditory object depending on its location with respect to the listener's body (Asutay & Västfjäll, 2015). Specifically, sounds in the rear space are perceived as more arousing than sounds in the front space (Pinheiro et al., 2019; Tajadura-Jiménez et al., 2010). In addition, the P2 component has been proposed to reflect automatic salience detection (Paulmann & Kotz, 2008), and both the P2 and the LPP are modulated by arousal (Pell et al., 2015;

Paulmann et al., 2013; Kotz & Paulmann, 2011). The spatial source of a sound was also found to modulate P2 amplitudes (Lewald & Getzmann, 2011) and studies in the visual modality reported interaction effects of emotional and spatial features on both the P2 (Xie et al., 2014) and LPP components (Du, Wang, Abrams, & Zhang, 2017). Considering this evidence, interactive effects of emotion and space were expected on the P2 and LPP. Specifically, increased P2 and LPP amplitudes were expected in response to emotional vocalizations presented from the back (vs. front). As prior studies failed to provide evidence for interactive effects of emotion and space on earlier processing stages (Du et al., 2017; Xie et al., 2014), we did not expect such interactions on the N1.

In the current study, task instructions aimed to promote the allocation of attentional resources toward distinct vocal features, leading to differences in the implementation of perceptual and cognitive strategies. On the basis of previous studies documenting modulatory effects of task instructions (i.e., implicit vs. explicit processing of emotional information) on the N1 (Ho, Schröger, & Kotz, 2015), P2 (Spreckelmeyer, Kutas, Urbach, Altenmüller, & Münte, 2009), and LPP (Van Strien, De Sonneville, & Franken, 2010) amplitudes, we expected amplitude differences in these ERP components as a function of attention focus. Notwithstanding, this hypothesis is exploratory because there is no direct evidence of ERP modulations by attentional focus on emotional versus spatial features of the voice.

METHODS

Participants

An a priori power analysis was conducted using MorePower software (Campbell & Thompson, 2012) to estimate the sample size. Considering a medium-sized effect of .25, α at .05, and power set to .80, 40 participants should be tested.

Forty-nine participants were recruited. Seven participants were excluded: Two of them did not meet all the inclusion criteria, and the remaining five were excluded because of excessive EEG motion artifacts. The final sample consisted of 42 participants (21 male), with ages between 19 and 29 years (M = 22.40, SD = 2.60). A between-subjects design was used. Twenty participants completed the location discrimination task (10 male; M = 22.30, SD = 2.55, age range = 19–28 years), and 22 individuals performed the emotion recognition task (11 male; M = 22.50, SD = 2.64, age range = 19–29 years). Inclusion criteria were: European Portuguese as native language, age between 18 and 31 years, right handedness (Edinburgh Handedness Inventory; Oldfield, 1971), no self-reported history of psychiatric disorder (Brief Symptoms Inventory; Portuguese adaptation by Canavarro, 1999) or substance abuse (Alcohol, Smoking and Substance Involvement Screening Test; validated for the Portuguese population by Mostardinha, Bártolo, Bonifácio, & Pereira, 2019), no use of medication for psychiatric disorders, and normal or corrected-to-normal vision and hearing. Written informed consent was obtained from all participants, who received course credits for their participation in the study. The study was approved by the Ethics Committee of the Faculty of Psychology, University of Lisbon.

Stimuli

Spatialized nonverbal vocalizations portraying positive (amusement), negative (anger), or neutral states were presented in six different locations around the head: 45°, 90°, 135°, 225°, 270°, and 315° (see Figure 2). The stimuli used in the present study were the same as in Pinheiro and colleagues (2019). The two emotional categories were chosen as they represent opposites of the valence continuum and share similar acoustic profiles (e.g., high intensity and variable fundamental frequency (F0); Juslin & Laukka, 2003). Emotional vocalizations were selected from the

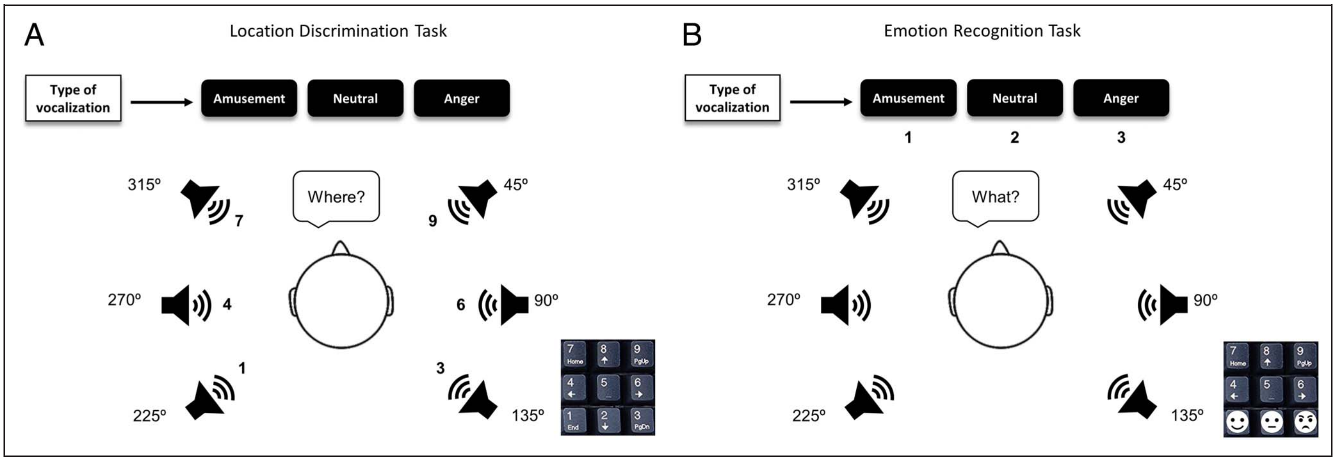

Figure 2. Schematic illustration of the experimental setup in the spatial location discrimination (A) and emotion recognition (B) tasks. Adapted from Pinheiro and colleagues (2019).

Table 1. Acoustic Properties of the Vocalizations.

Emotional	*Space*	*Duration (msec)*	*f0 Mean (Hz)*	*f0 Min (Hz)*	*f0 Max (Hz)*
Neutral	Front	929	174.925	163.521	228.857
	Back	929	178.802	163.478	228.683
	Left	929	177.479	163.614	228.756
	Right	929	177.479	163.614	228.756
Amusement	Front	984	295.029	205.863	378.368
	Back	984	281.910	182.879	366.334
	Left	984	289.78	197.520	374.314
	Right	984	289.78	197.520	374.314
Anger	Front	934	187.571	87.251	342.943
	Back	934	189.422	87.383	363.043
	Left	934	189.231	86.959	351.424
	Right	934	189.231	86.959	351.424

corpus of nonverbal vocalizations by Lima, Castro, and Scott (2013). Because the corpus of nonverbal vocalizations by Lima and colleagues (2013) did not include neutral vocalizations, neutral stimuli were selected from the Montreal Affective Voices battery (Belin, Fillion-Bilodeau, & Gosselin, 2008), validated for the Portuguese population (Vasconcelos, Dias, Soares, & Pinheiro, 2017). The same set of stimuli was used in both tasks.[1] The acoustic properties of the stimuli are described in Table 1.

The spatialization of vocalizations was achieved using the MIT Head-Related Transfer Function database (https://sound.media.mit.edu/resources/KEMAR.html). Binaural sounds were generated with 45°, 90°, 135°, 225°, 270°, and 315° azimuth degrees. The impulse responses of each head-related transfer function were then convoluted with the original sound samples, originating a spatially located sound, by adding inter-aural time difference and inter-aural level difference cues to correspond to a new spatial location (see Pinheiro et al., 2019). This virtual auditory environment preserves spatial integrity and perceptual realism (Kryklywy et al., 2013) during sound presentation via headphones (HD 202, Sennheiser). For each type of vocalization (i.e., amusement, neutral, and anger; total = 144), eight different stimuli were selected (four female and four male participants). Sounds were repeated to achieve 35 stimuli per condition, leading to 630 vocalizations in each task. Following prior studies (e.g., Liu et al., 2012), each condition included 35 trials to ensure a good signal-to-noise ratio while minimizing task duration.

Procedure

Each participant participated in one of two tasks (spatial location discrimination, emotion recognition) while the EEG was recorded. Participants were randomly allocated to each task while ensuring that the proportion of male and female participants was equivalent. After completing the EEG task, participants rated the valence and arousal of each vocalization using a 9-point scale (Bradley & Lang, 1994).

EEG sessions were conducted in an electrically shielded and acoustically isolated booth. Participants were comfortably seated 100 cm away from the computer monitor, and responses were provided using the number pad on the keyboard. Stimulus presentation, timing of events, and recording of participants' responses were controlled with Presentation software (Neurobehavioral Systems, Inc.). Both tasks involved a similar trial structure (see Figure 3). Each trial began with a fixation cross, presented 1000 msec before the sound and which remained up until 1000 msec after sound presentation. Then, a question mark and a picture of the number pad were presented to prompt participant's response and minimize working

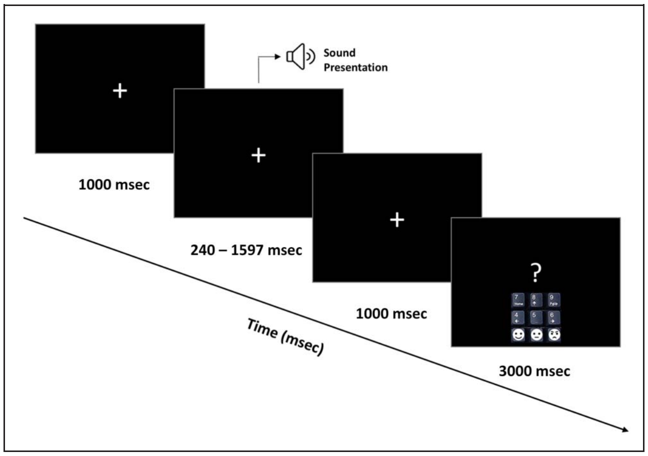

Figure 3. Illustration of an experimental trial in the emotion recognition task. The trial structure is similar for both tasks.

memory demands. Participants had a maximum of 3000 msec to respond before the beginning of the next trial. Each task consisted of 630 trials, distributed over four blocks. Participants were allowed short pauses of 20 sec every 32 stimuli, as well as longer breaks between blocks. Stimuli were pseudorandomized to ensure that the same emotion or spatial location was not presented more than three consecutive times (Bertels, Kolinsky, Coucke, & Morais, 2013). Each task lasted approximately 60 min. During the EEG tasks, no feedback was provided.

Location Discrimination Task

Participants were asked to discriminate the spatial location of the vocalizations (see Figure 2A). Before the EEG task, a training session with 70 trials allowed participants to get familiarized with the task instructions and response keys. Feedback was provided during the training session of the spatial location discrimination task. Vocalizations presented in the training session were not included in the EEG task.

Emotion Recognition Task

Participants were asked to identify the emotional category associated with each vocalization in a forced-choice emotion recognition task (see Figure 2B). Before the EEG task, a training session with six trials allowed participants to get familiarized with the task instructions and response keys. No feedback was provided during the training session of the emotion recognition task. Vocalizations presented in the training session were not included in the EEG task. Participants responded using the three bottom keys of the number pad, which had stickers of cartoon faces portraying the three emotional categories to minimize working memory demands (see Figure 2B and Figure 3). The keys assigned to each emotion were counterbalanced to avoid potential biases because of an implicit association of emotion to space (Amorim & Pinheiro, 2019).

Posttask Affective Ratings

After completing the EEG task, participants rated the affective properties of the stimulus in two dimensions (i.e., valence and arousal), using a 9-point scale (Bradley & Lang, 1994). Following stimulus presentation, participants indicated (a) how pleasant (1 = *very unpleasant*, 9 = *very pleasant*) and (b) how arousing they considered each vocalization (1 = *very calm*, 9 = *very aroused*). These ratings were collected for all vocalizations. The task lasted approximately 30 min.

EEG Data Recording

EEG data were recorded using a 64-channel Active Two Biosemi system in a continuous mode at a digitization rate of 512 Hz and stored on a disk for later analysis. Eye blinks and movements were monitored through electrodes placed on both temples (horizontal electrooculogram) and another one below the left eye (vertical electrooculogram). Two additional electrodes were placed on both mastoids to be used as offline reference channels later.

EEG Data Analysis

EEG data were analyzed using BrainVision Analyzer 2 software (Brain Products). The EEG channels were referenced offline to the average of the left and right mastoids and filtered using a 0.1- to 30-Hz bandpass filter. Individual epochs were created with a −200-msec prestimulus baseline and 1000 msec poststimulus for each condition (i.e., Neutral 45°; Neutral 90°; Neutral 135°; Neutral 225°; Neutral 270°; Neutral 315°; Amusement 45°; Amusement 90°; Amusement 135°; Amusement 225°; Amusement 270°; Amusement 315°; Anger 45°; Anger 90°; Anger 135°; Anger 225°; Anger 270°; Anger 315°). Epochs were baseline-corrected using the −200- to 0-msec prestimulus interval. Vertical and horizontal eye movements were corrected using the method of Gratton, Coles, and Donchin (1983). EEG segments were semi-automatically screened for eye movements, muscle artifacts, electrode drifting, and amplifier blocking: EEG epochs exceeding ± 100 μV were rejected. After artifact rejection, at least 75% of segments per condition per participant entered the analyses. On average, 33 trials were entered per condition (M = 33.19, SD = 0.93). There were no differences between conditions in the number of segments included in the analyses ($p > .05$).

Only EEG epochs associated with correct behavioral responses were included in the analysis of ERP components in the emotion recognition task. In the spatial location discrimination task, the analysis considered trials associated with both correct and incorrect responses because of task difficulty, aiming to ensure an adequate signal-to-noise ratio. Indeed, spatial acuity of hearing is reduced compared with vision. Unlike visual and somatosensory areas, the auditory cortex lacks a topographical space representation, which is reflected in poor performance in sound source localization (Kong et al., 2014).

The ERP analysis focused on the N1, P2, and LPP components. Mean amplitudes were analyzed in time windows selected according to previous studies (Pinheiro, Rezaii, et al., 2016; Pell et al., 2015; Paulmann et al., 2013; Liu et al., 2012): 110–190 msec (N1), 200–280 msec (P2), and 450–1000 msec (LPP). Long-lasting late positive effects (LPP) were analyzed in two latency windows to track differences between the conditions more accurately (Pinheiro, Sarzedas, Roberto, & Kotz, 2023; Conde et al., 2022): early LPP phase (450–700 msec); late LPP phase (701–1000 msec). Mean amplitude was chosen over peak amplitude as it is less sensitive to high frequency noise and variability in component latency (Luck, 2005). On the basis of previous studies (Pinheiro et al., 2023; Pinheiro, Rezaii, et al., 2016; Liu et al., 2012), as well as on inspection of grand average waveforms, the following ROIs were selected for the statistical analyses: frontocentral (FCz, FC3, FC4), central (Cz, C3, C4), and centroparietal (CPz, CP3, CP4).

Statistical Analyses

The SPSS statistical software package (Version 29.0, IBM Corp.) was used in the statistical analyses. The alpha level was set at .05. To reduce the number of levels in the ROI factor and to simplify the statistical models, locations were grouped into four spaces (front: 45° and 315°; back: 135° and 225°; right: 45°, 90°, and 135°; and left: 225°, 270°, and 315°) following previous studies (Pinheiro et al., 2019). Data were separately analyzed considering two spatial axes: front–back and left–right.

ERP Data

Repeated-measures ANOVAs were separately computed for the N1, P2, and LPP mean amplitudes, with emotion (three levels: neutral, amusement, anger), space (two levels: front vs. back; left vs. right), and ROI (three levels: frontocentral, central, and centroparietal) as within-subjects factors, and task (two levels: spatial location discrimination and emotion recognition) as between-subjects factor. Main effects and interactions were followed up with pairwise comparisons using Bonferroni correction for multiple comparisons. The analyses were corrected for nonsphericity using the Greenhouse–Geisser method, when the Mauchly's test indicated that the assumption of sphericity had been violated.

Behavioral Data

Behavioral measures included recognition accuracy data in each task (correct responses in the spatial location discrimination and emotion recognition tasks) and affective ratings (valence and arousal of each vocalization, rated on a 9-point scale; Bradley & Lang, 1994). These behavioral measures were separately analyzed with a repeated-measures ANOVA with Emotion (three levels: neutral, amusement, and anger) and Space (2 levels: front vs. back; left vs. right) as within-subject factors. Main effects and interactions were followed up with pairwise comparisons using Bonferroni correction for multiple comparisons. The analyses were corrected for nonsphericity using the Greenhouse–Geisser method, when the Mauchly's test indicated that the assumption of sphericity had been violated.

RESULTS

We first report independent effects of Emotion and Space, and then interactive effects of the two properties.

ERP Results

Mean amplitudes of the N1, P2, and LPP components per condition are shown in Table 2. Figures depicting grand average waveforms over the nine electrode sites included in the statistical analyses can be found in the Appendix.

Table 2. Mean Amplitudes of the N1, P2, and Early and Late LPP Components Per Emotional Condition

		N1	*P2*	*eLPP*	*lLPP*
Emotion	*Space*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Neutral	Front	−1.645 (0.337)	2.993 (0.374)	−2.808 (0.382)	−3.633 (0.401)
	Back	−1.502 (0.319)	2.746 (0.357)	−2.752 (0.305)	−3.417 (0.328)
	Left	−1.909 (0.345)	2.519 (0.359)	−2.900 (0.330)	−3.660 (0.350)
	Right	−1.734 (0.340)	3.197 (0.391)	−2.889 (0.347)	−3.734 (0.350)
Amusement	Front	−1.212 (0.304)	3.765 (0.342)	−1.306 (0.332)	−2.176 (0.404)
	Back	−1.118 (0.320)	3.604 (0.362)	−1.452 (0.346)	−2.319 (0.394)
	Left	−1.322 (0.321)	3.725 (0.325)	−1.326 (0.338)	−2.296 (0.400)
	Right	−1.216 (0.306)	3.925 (0.364)	−1.246 (0.352)	−2.122 (0.388)
Anger	Front	−1.101 (0.314)	4.063 (0.371)	−0.810 (0.405)	−1.420 (0.410)
	Back	−1.083 (0.331)	4.009 (0.385)	−0.705 (0.359)	−1.372 (0.364)
	Left	−1.209 (0.344)	4.001 (0.364)	−0.733 (0.383)	−1.520 (0.361)
	Right	−1.301 (0.307)	4.060 (0.364)	−0.709 (0.369)	−1.399 (0.415)

M = mean; *SD* = standard deviation.

Table 3. Summary of the Repeated-measures ANOVA Results for the N1 Component

	Front–Back				*Left–Right*			
Effects	*F*	*df*	*p*	η^2	*F*	*df*	*p*	η^2
Emotion	6.627	2, 80	.002	.142	11.533	2, 80	<.001	.224
Space	.518	1, 40	.476	.013	.262	1, 40	.612	.007
Emotion × Space	.133	2, 80	.875	.003	.672	1.711, 68.445	.492	.017
Task	<.001	1, 40	.991	<.001	<.001	1, 40	.984	<.001
Emotion × Task	.393	2, 80	.677	.010	1.122	2, 80	.331	.027
Space × Task	2.323	1, 40	.135	.055	.141	1, 40	.710	.004
Emotion × Space × Task	.485	2, 80	.618	.012	1.249	2, 80	.292	.030

Statistically significant effects ($p < .05$) are shaded in gray. *df* = degrees of freedom.

Emotion Effects

Independent effects of Emotion were observed on the N1, P2, and LPP amplitude. These effects were robustly detected in both spatial axes (see Tables 3–6).

N1

Emotional vocalizations were associated with reduced N1 amplitudes compared with neutral vocalizations (front–back: all *p*s ≤ .038; left–right: all *p*s ≤ .002;

Table 4. Summary of the Repeated-measures ANOVA Results for the P2 Component

	Front–Back				*Left–Right*			
Effects	*F*	*df*	*p*	η^2	*F*	*df*	*p*	η^2
Emotion	23.575	2, 80	<.001	.371	30.360	2, 80	<.001	.431
Space	1.928	1, 40	.173	.046	8.534	1, 40	.006	.176
Emotion × Space	.183	2, 80	.833	.005	3.489	2, 80	.035	.080
Task	.205	1, 40	.653	.005	.220	1, 40	.642	.005
Emotion × Task	.996	2, 80	.374	.024	.345	2, 80	.709	.009
Space × Task	4.535	1, 40	.039	.102	1.267	1, 40	.267	.031
Emotion × Space × Task	.467	2, 80	.628	.012	.839	2, 80	.436	.021

Statistically significant effects ($p < .05$) are shaded in gray.

Table 5. Summary of the Repeated-measures ANOVA Results for the eLPP Component

	Front–Back				*Left–Right*			
Effects	*F*	*df*	*p*	η^2	*F*	*df*	*p*	η^2
Emotion	67.197	2, 80	<.001	.627	73.447	1.598, 63.939	<.001	.647
Space	.002	1, 40	.968	<.001	.098	1, 40	.756	.002
Emotion × Space	.384	2, 80	.682	.010	.038	2, 80	.963	.001
Task	.661	1, 40	.421	.016	.389	1, 40	.537	.010
Emotion × Task	.163	2, 80	.850	.004	.259	2, 80	.722	.006
Space × Task	.699	1, 40	.408	.017	.239	1, 40	.628	.006
Emotion × Space × Task	.676	2, 80	.512	.017	1,112	2, 80	.334	.027

Statistically significant effects ($p < .05$) are shaded in gray.

Table 6. Summary of the Repeated-measures ANOVA Results for the lLPP Component

	Front–Back				*Left–Right*			
Effects	*F*	*df*	*p*	η^2	*F*	*df*	*p*	η^2
Emotion	59.254	2, 80	<.001	.597	69.822	2, 80	<.001	.636
Space	.061	1, 40	.806	.002	.324	1, 40	.572	.008
Emotion × Space	.475	1.686, 67.437	.591	.012	.392	2, 80	.677	.010
Task	.504	1, 40	.482	.012	.261	1, 40	.612	.006
Emotion × Task	2.143	2, 80	.124	.051	1.135	2, 80	.327	.028
Space × Task	.979	1, 40	.328	.024	.044	1, 40	.834	.001
Emotion × Space × Task	1.886	2, 80	.158	.045	1.222	2, 80	.300	.030

Statistically significant effects ($p < .05$) are shaded in gray.

Table 3), with no differences between amused and angry (front–back: $p = 1$; left–right: $p = 1$) voices (see Figures 4 and 5).

P2

P2 amplitudes were increased to emotional compared with neutral vocalizations (front–back: all *p*s < .001; left–right: all *p*s < .001; Table 4), with no differences between amusement and anger (front–back: $p = .125$; left–right: $p = .604$; see Figures 4 and 5).

Early LPP

The LPP was increased in response to emotional compared with neutral vocalizations (front–back: all *p*s <

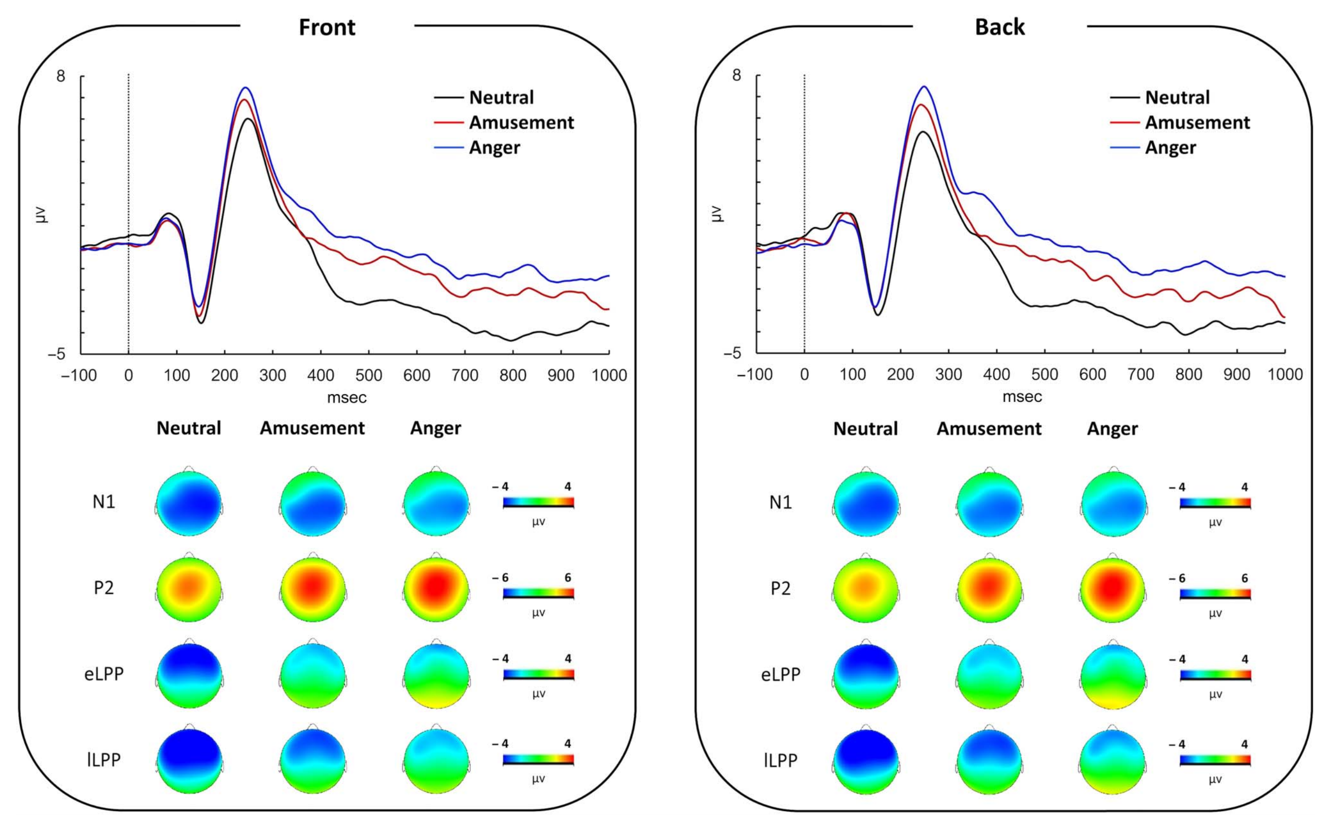

Figure 4. Grand average waveforms contrasting neutral, amused, and angry vocalizations in the front–back axis. Voltage maps show the topographic distribution of the N1 (110–190 msec), P2 (200–280 msec), early LPP (450–700 msec), and late LPP (700–1000 msec) components. Images are shown over the electrode Cz. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

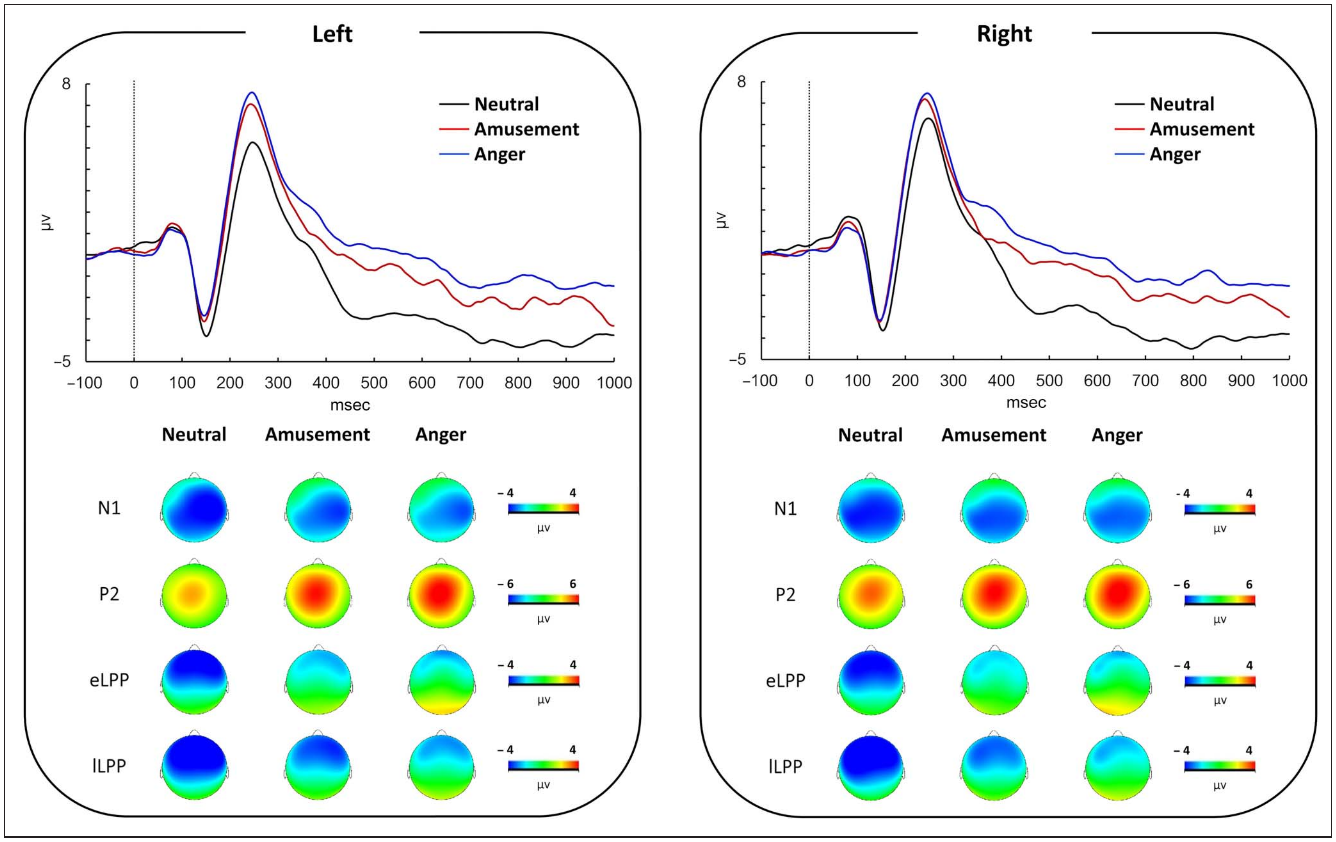

Figure 5. Grand average waveforms contrasting neutral, amused, and angry vocalizations in the left–right axis. Voltage maps show the topographic distribution of the N1 (110–190 msec), P2 (200–280 msec), early LPP (450–700 msec), and late LPP (700–1000 msec) components. Images are shown over the electrode Cz. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

.001; left–right: all ps $< .001$; Table 5). The LPP was also sensitive to valence: Amplitudes were enhanced in response to vocalizations expressing anger compared with amusement (front–back: $p = .002$; left–right: $p = .002$; see Figures 4 and 5).

Late LPP

The late phase of the LPP showed similar emotion effects (Table 6). Emotional vocalizations were associated with increased LPP amplitudes compared with neutral ones (front–back: all ps $< .001$; left–right: all ps $< .001$). A valence-specific effect showed enhanced LPP amplitudes in response to vocalizations expressing anger compared with amusement (front–back: $p < .001$; left–right: $p < .001$; see Figures 4 and 5).

Space Effects

Independent effects of spatial location were noted on the P2 component (see Table 4). Specifically, spatial location modulated the P2 amplitude in the left–right axis, $F(1, 40) = 8.534$; $p = .006$, partial $\eta^2 = .176$: Vocalizations presented on the right side elicited increased P2 amplitudes compared with those presented on the left (see Figure 6). Spatial location did not significantly modulate N1 and LPP amplitudes ($p > .05$; see Tables 3, 5 and 6).

Emotion and Space Interactions

Interactive effects of Emotion and Space were observed on the P2. Specifically, the interaction between emotion and space reached significance in the left–right axis (see Table 4 and Figure 6). The P2 amplitude was increased in response to neutral vocalizations presented from the right versus left side ($p = .001$). However, the P2 did not differ between the left and right sides for emotional vocalizations (angry: $p = .732$; amused: $p = .239$). Emotion and Space did not interactively modulate N1 and LPP amplitudes in both front–back and left–right axes (see Tables 3, 5 and 6, and Figure 6).

Attention Effects

Attentional focus modulated the processing of spatial cues in voice perception but only in the P2 timeframe. Specifically, the P2 was modulated by both task and

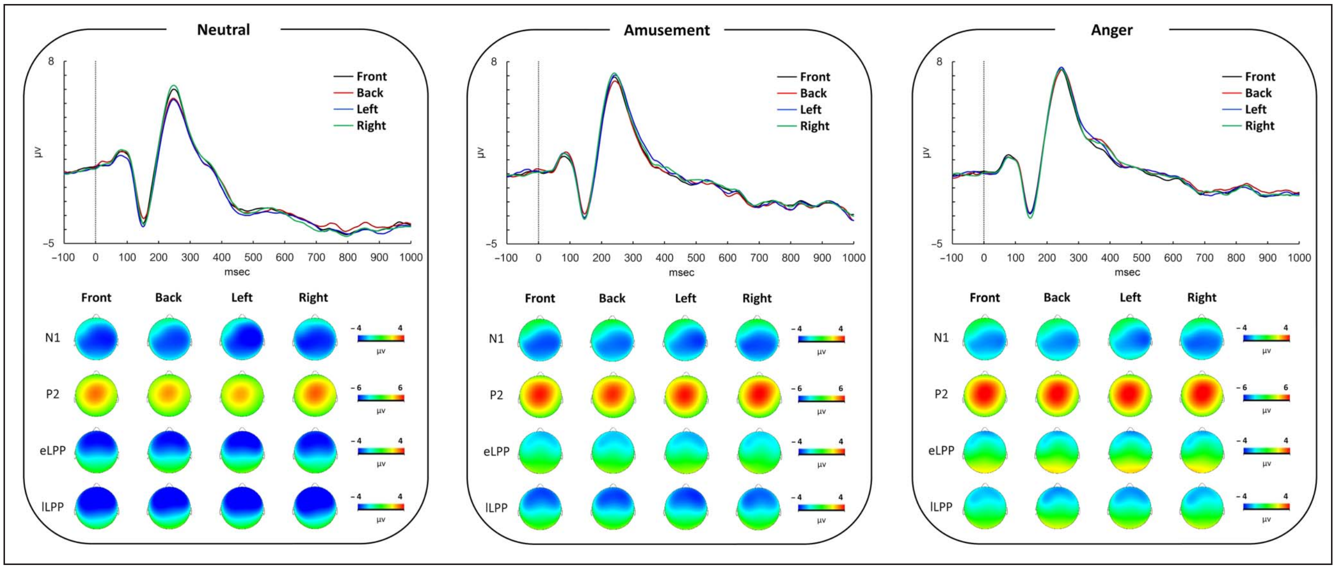

Figure 6. Grand average waveforms contrasting vocalizations presented at the front, back, left, and right locations per emotional category. Voltage maps show the topographic distribution of the N1 (110–190 msec), P2 (200–280 msec), early LPP (450–700 msec), and late LPP (700–1000 msec) components. Images are shown over the electrode Cz. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

spatial location in the front–back axis (see Table 4 and Figure 7): The P2 was increased in response to vocalizations presented at the front compared with back locations but only when listeners were instructed to evaluate the emotional quality of the voice (p = .015). In the spatial location discrimination task, there were no differences between front and back locations (p = .611). In the left–right axis, the interaction between Space and Task did not reach significance (see Table 4 and Figure 8). Task instructions did not significantly modulate N1 and LPP amplitudes ($p > .05$; see Tables 3, 5, and 6, and Figures 7 and 8).

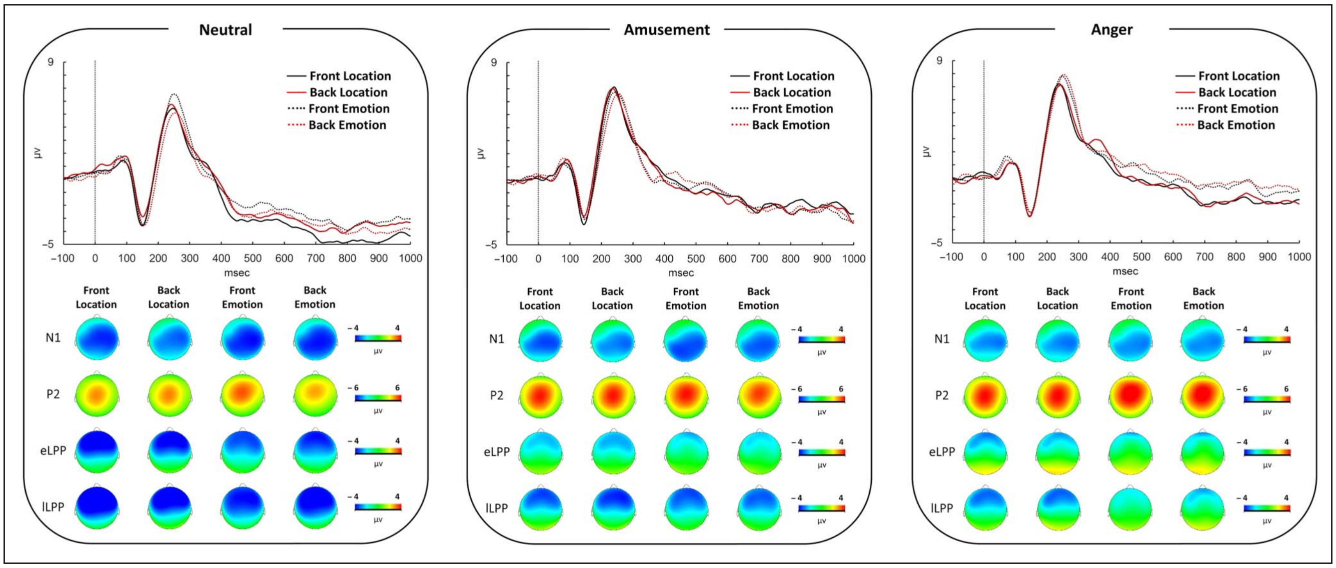

Figure 7. Grand average waveforms contrasting vocalizations presented at the front and back locations in the spatial location discrimination and emotion recognition tasks per emotional category. Voltage maps show the topographic distribution of the N1 (110–190 msec), P2 (200–280 msec), early LPP (450–700 msec), and late LPP (700–1000 msec) components. Images are shown over the electrode Cz. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

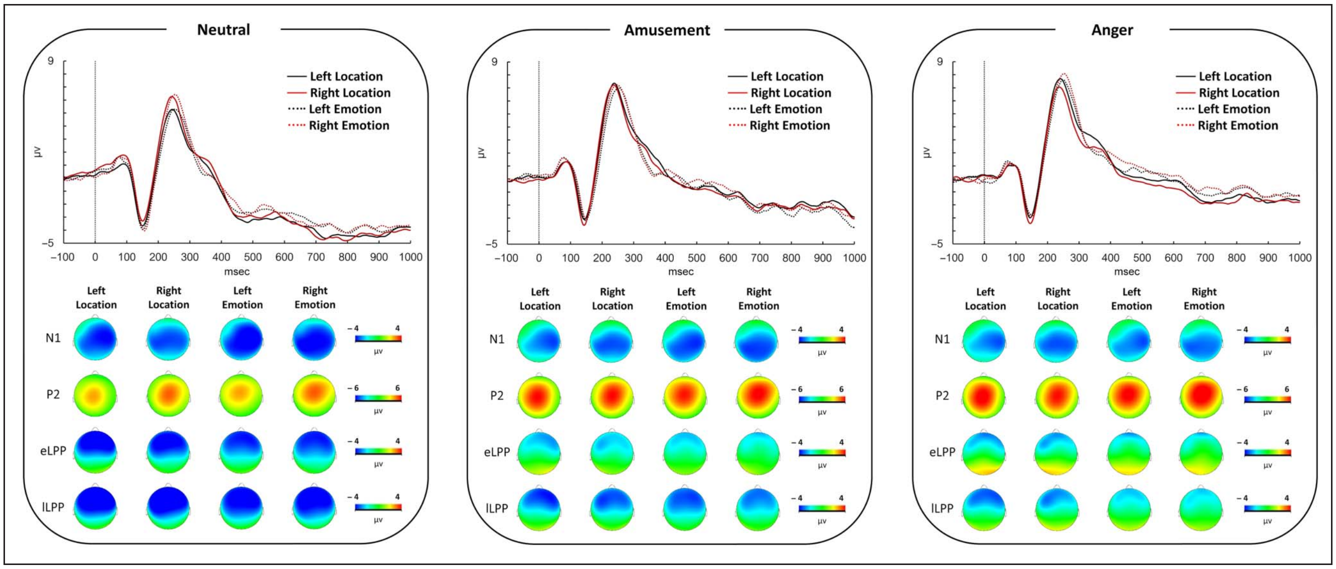

Figure 8. Grand average waveforms contrasting vocalizations presented at the left and right locations in the spatial location discrimination and emotion recognition tasks per emotional category. Voltage maps show the topographic distribution of the N1 (110–190 msec), P2 (200–280 msec), early LPP (450–700 msec), and late LPP (700–1000 msec) components. Images are shown over the electrode Cz. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

Behavioral Results

Accuracy in the spatial location discrimination and emotion recognition tasks is shown in Table 7; mean arousal and valence ratings are shown in Table 9.

Emotion Effects

Spatial Location Discrimination

The emotional relevance of the voice modulated location accuracy in both front–back ($p = .005$) and left–right axes ($p < .001$; see Table 8). Amused vocalizations were more accurately located than both neutral (front–back: $p = .022$; left–right: $p = .013$) and angry (front–back: $p = .019$; left–right: $p = .001$) vocalizations, with no differences between the latter two (front–back: $p = 1$; left–right: $p = .213$; see Table 8).

Table 7. Mean Percentage of Correct Responses in the Spatial Location Discrimination and Emotion Recognition Tasks

		Location Task	*Emotion Task*
Emotion	*Space*	*M* (*SD*)	*M* (*SD*)
Neutral	Front	27.5 (4.2)	98.0 (0.6)
	Back	40.9 (3.9)	98.6 (0.5)
	Left	46.6 (2.5)	97.7 (0.6)
	Right	41.6 (2.2)	98.0 (0.5)
Amusement	Front	27.5 (3.9)	99.0 (0.3)
	Back	47.4 (5.1)	99.0 (0.4)
	Left	48.3 (3.3)	99.2 (0.2)
	Right	45.5 (2.6)	99.0 (0.3)
Anger	Front	22.5 (3.3)	99.0 (0.5)
	Back	43.9 (4.5)	98.5 (0.6)
	Left	44.4 (2.7)	99.1 (0.3)
	Right	41.0 (2.6)	98.3 (0.7)

Emotion Recognition

Recognition accuracy was not affected by the emotional quality of the vocalizations (front–back: $p = .550$; left–right: $p = .074$; see Table 8).

Space Effects

Spatial Location Discrimination

Discrimination accuracy differed as a function of the spatial source of vocalizations in both axes (front–back: $p = .003$; left–right: $p = .018$). Vocalizations coming from the back and left locations were more accurately located than those presented at the front and right locations, respectively (see Table 8).

Emotion Recognition

Recognition accuracy was not affected by the spatial source of the vocalizations (front–back: $p = .806$; left–right: $p = .241$; see Table 8).

Emotion and Space Interactions

Emotion and Space did not interactively modulate spatial location discrimination or emotion recognition accuracy

Table 8. Summary of the Repeated-measures ANOVA Results for the Behavioral Measures

		Front–Back				*Left–Right*			
Measure	*Effects*	*F*	*df*	*p*	η^2	*F*	*df*	*p*	η^2
Location discrimination accuracy	Emotion	6.055	2, 38	.005	.242	12.827	2, 38	<.001	.403
	Space	11.687	1, 19	.003	.381	6.717	1, 19	.018	.261
	Emotion × Space	2.778	1.452, 27.584	.094	.128	.651	2, 38	.527	.033
Emotion recognition accuracy	Emotion	.606	2, 42	.550	.028	2.778	2, 42	.074	.117
	Space	.062	1, 21	.806	.003	1.456	1, 21	.241	.065
	Emotion × Space	2.132	2, 42	.131	.092	1.740	1.564, 32.850	.196	.077
Valence ratings	Emotion	441.884	1.274, 42.039	<.001	.931	440.917	1.282, 42.314	<.001	.930
	Space	3.894	1, 33	.057	.106	2.884	1, 33	.099	.080
	Emotion × Space	8.829	1.537, 50.732	.001	.211	3.086	1.659, 54.757	.063	.086
Arousal ratings	Emotion	64.965	1.520, 50.165	<.001	.663	66.251	1.521, 50.199	<.001	.668
	Space	28.908	1, 33	<.001	.467	.354	1, 33	.556	.011
	Emotion × Space	.952	2, 66	.391	.028	3.849	1.666, 54.992	.034	.104

Statistically significant effects ($p < .05$) are shaded in gray.

(see Table 8). However, significant interactions were found on mean valence and arousal ratings.

Valence Ratings

The interaction was significant in the front–back axis ($p = .001$; see Table 8): Angry vocalizations were rated as more negative when presented at front compared with rear locations ($p = .001$); however, no such differences were observed for amused ($p = .092$) or neutral ($p = .178$) vocalizations.

Arousal Ratings

The interaction was significant in both axes. In the front–back axis ($p < .001$; see Table 8), vocalizations coming from the front were considered more arousing than those coming from the back (see Table 8). In the left–right axis ($p = .034$; see Table 8), angry vocalizations were considered more arousing when presented on the left than on the right side ($p = .044$). However, no such differences were observed for amused ($p = .292$) and neutral ($p = .438$) vocalizations (Table 9).

In summary, at the brain level, independent effects of Emotion were observed on the N1, P2, and LPP components. Emotion and Space interactively modulated the P2 amplitude but only on the left–right axis. Attentional focus interacted with space within the P2 timeframe but only on the front–back axis. Behaviorally, independent effects of Emotion and Space were observed on location discrimination accuracy, whereas interactive effects were

Table 9. Ratings of Valence and Arousal Per Condition

		Valence	*Arousal*
Emotion	*Space*	*M (SD)*	*M (SD)*
Neutral	Front	5.00 (0.03)	3.39 (0.26)
	Back	5.02 (0.02)	3.31 (0.27)
	Left	5.01 (0.02)	3.36 (0.26)
	Right	5.00 (0.03)	3.33 (0.26)
Amusement	Front	7.21 (0.16)	5.96 (0.31)
	Back	7.14 (0.16)	5.79 (0.32)
	Left	7.15 (0.16)	5.86 (0.32)
	Right	7.22 (0.16)	5.92 (0.32)
Anger	Front	1.91 (0.11)	7.25 (0.17)
	Back	2.06 (0.13)	7.14 (0.18)
	Left	1.94 (0.12)	7.27 (0.18)
	Right	1.98 (0.12)	7.18 (0.18)

Valence and arousal ratings range between 1 and 9.

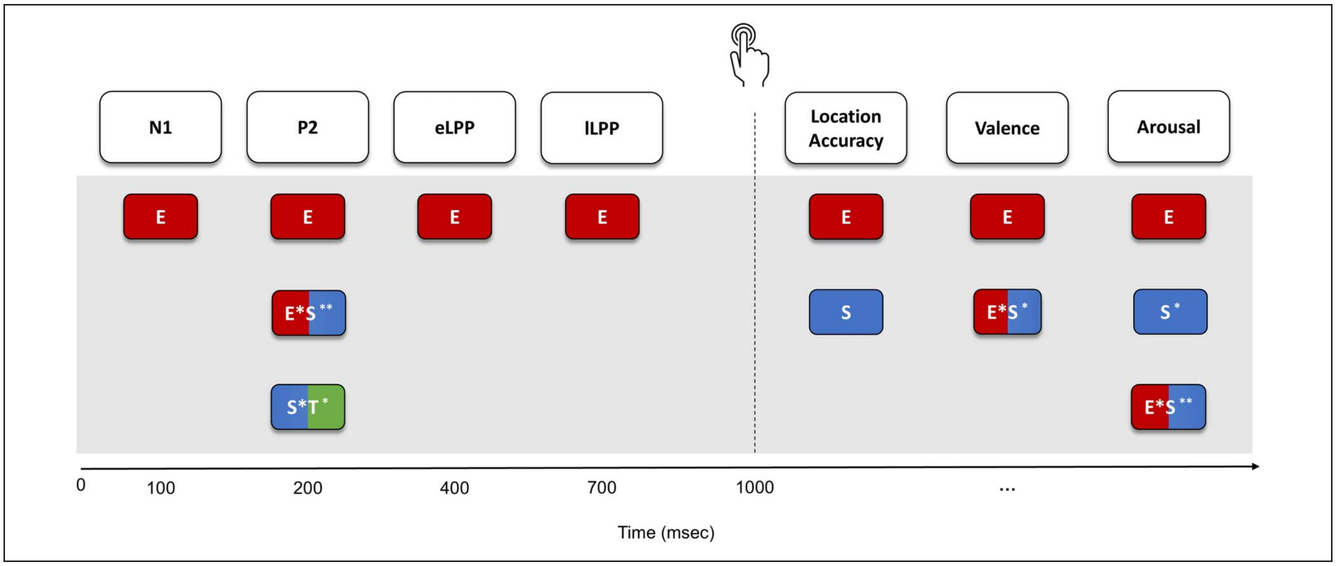

Figure 9. Schematic illustration of the main ERP and behavioral results. Electrophysiological data were analyzed until 1000 msec after stimulus onset. Behavioral data, signaled by the keypress image, were collected from 1000 msec onward. With regard to emotion effects, they first emerge in the N1 timeframe and persist throughout the P2 and LPP timeframes. Space effects only emerge in the P2 timeframe and are also observed after 1000 msec in arousal ratings. Interactions between Space and Task also occur in the P2 timeframe. Finally, interactions between Emotion and Space are also observed behaviorally, from 1000 msec onward, in affective ratings. E = main effect of Emotion; S = main effect of Space; S × T = interaction between Space and Task; E × S = interaction between Emotion and Space; *front–back axis; **right–left axis.

observed on affective ratings (see Figure 9 for a schematic illustration of the results).

DISCUSSION

Combining EEG and behavioral methods, we examined the independent and interactive effects of emotion and space in voice perception, and their modulation by task instructions. The results clarify the role of emotion with respect to its impact in dual pathway models of sensory processing. First, the EEG findings revealed that emotion modulated the early (i.e., N1 and P2) and late (i.e., LPP) stages of voice processing, whereas effects of spatial location were only observed on the P2. Spatial location also interacted with either emotion or attention focus approximately at 200 msec post-voice onset, coinciding with the timing of processes that abstract emotional salience from a set of auditory cues (Schirmer & Kotz, 2006). Second, the behavioral findings replicated interactive effects of emotion and spatial location on ratings of stimulus valence and arousal, but only for vocalizations of negative valence. In the next sections, we first discuss independent effects of emotion and spatial location and then discuss their interactions.

Independent Effects of Emotion and Space

EEG findings revealed that the analysis of auditory emotional relevance occurs rapidly, with differences in neural activity between emotional and nonemotional sounds starting within 100 msec poststimulus onset. Confirming our first hypothesis, spatialized emotional vocalizations were associated with reduced N1 and increased P2 and LPP amplitudes compared with neutral stimuli, extending prior findings with nonspatialized voices (Pell et al., 2015; Paulmann et al., 2013; Liu et al., 2012; Kotz & Paulmann, 2011). These findings confirm the facilitated sensory analysis (N1), enhanced salience decoding (P2), and sustained elaborative processing (LPP) of emotional relative to neutral vocalizations, irrespective of attention focus. Whereas early processing stages (N1 and P2) were more sensitive to stimulus arousal, later processing stages (LPP) revealed valence-specific effects: Angry vocalizations were more thoroughly processed and engaged heightened sustained attention compared with vocalizations expressing amusement.

Whereas emotion modulated voice perception within the first 100 msec poststimulus onset, the effects of space occurred approximately 100 msec later. In early sensory areas of the "dorsal" and "ventral" pathways, neurons respond to both spatial and nonspatial sound attributes but may show selectivity for one of these two attributes. Previous studies showed faster activation of the auditory "where" versus "what" pathway (Ahveninen et al., 2006; Anourova et al., 2001). However, these studies manipulated nonspatial sound features such as pitch (Anourova et al., 2001) and phonetic information (Ahveninen et al., 2006) but did not account for the emotional quality of the vocal stimuli. This critical difference may explain the discrepant results as emotional cues seem to be prioritized over other types of vocal features during voice processing (e.g., Pinheiro, Barros, et al., 2016). Other studies that manipulated both emotional and spatial information also failed to observe earlier effects for spatial than emotional vocal cues (Burra, Kerzel, Munoz Tord, Grandjean, &

Ceravolo, 2019; Gädeke, Föcker, & Röder, 2013). Taken together, these effects suggest that emotional representations are activated earlier than spatial representations and reinforce the prioritization of emotional cues in voice perception (e.g., Pinheiro, Barros, et al., 2016).

The present findings also indicate that auditory emotional content does not directly augment stimulus representation in space, in line with fMRI studies showing that emotion leads to increased activity in the anterior-lateral "what" pathway but does not modulate regions of auditory cortex sensitive to location (Grisendi, Clarke, & Da Costa, 2023; Kryklywy et al., 2013). This suggests that auditory processing pathways show differential responding to emotional information. Differences in how the neurocognitive system responds to emotional attributes of voices could nevertheless still affect explicit spatial discrimination at a later postdecisional stage. Here, we found that spatial localization was more accurate for positive than neutral and negative vocalizations. Affiliative vocal signals such as laughter are particularly salient cues (Pinheiro et al., 2017; Vasconcelos et al., 2017; Johnstone, van Reekum, Oakes, & Davidson, 2006) that may engage more attentional resources than negative and neutral cues (e.g., Pinheiro et al., 2017). Previous fMRI studies—characterized by lower temporal resolution than the EEG—showed stronger spatial effects for positive than negative stimuli (Grisendi et al., 2023; Kryklywy et al., 2013). The mechanisms by which explicit spatial decoding is facilitated by positive vocal cues should be addressed in future studies.

Interactions between Emotion and Space

Emotion and space interactively modulated P2 amplitudes. This finding indicates that the spatial and nonspatial auditory pathways may not be strictly parallel. The present interaction implies that "what" and "where" vocal cues were integrated around 200 msec poststimulus onset, when listeners putatively tag the relevance of vocal input. We observed that the P2 was increased for neutral vocalizations presented at right (vs. left) locations. No such differences were observed for positive and negative vocalizations.

The P2 is thought to involve contributions of the anterior superior temporal gyrus (Paulmann, Ott, & Kotz, 2011). This brain region is a plausible site for interactions between emotional and spatial features (Kryklywy et al., 2013), both known to contribute to perceived sound arousal. Functionally, the P2 has been linked to the integration of acoustic cues that allow rapid—most likely involuntary—stimulus appraisal (e.g., Paulmann & Kotz, 2008). This stage highlights relevant vocal features, including emotionality, high arousal, and other characteristics (Pell & Kotz, 2021) including spatial cues (see also Asutay & Västfjäll, 2015), and that will be available for higher-order cognitive processes. In the present study, the salience conferred by spatial location seems to have benefitted neutral vocalizations, which may have been processed more similarly to emotional vocalizations when presented at the right side (i.e., processed by the left hemisphere: neutral = 3.197 μV, amusement = 3.925 μV, anger = 4.060 μV; less than 1 μV difference between neutral and emotional vocalizations).

However, there was no additional salience increment conferred by spatial location to vocalizations with an emotional quality, that is, emotional salience may have overridden any potential benefit ascribed by spatial location. Besides, emotion may have augmented the representation of object features at the expense of spatial cues at this stage, in line with the observation that emotional cues lead to increased activity in the "what" but not in the "where" pathway (Grisendi et al., 2023; Kryklywy et al., 2013). This might explain why no additive effects of emotion and space were observed for emotional vocalizations.

Although sound identity and location are processed bilaterally at the cortical level, a predominantly left hemispheric network has been proposed to be involved in the encoding of the identity of a neutral auditory object and its tracking across space (Bourquin, Murray, & Clarke, 2013). This could explain the enhanced P2 for neutral vocalizations presented at the right side.

We expected emotional vocalizations to elicit an additional increase in P2 and LPP amplitudes when presented from back versus front locations. Contrary to our prediction, we observed no such differences. Spatial cues modulated emotional processing in a postdecisional stage of voice perception, reflected in behavioral indices. These findings suggest that the effects of space and emotion do not merely sum up but rather form an interdependent relationship at later processing stages that are better captured with behavioral measures. Specifically, angry vocalizations presented at the front were perceived as more negative than at the back, and more arousing when presented at the left than at the right side. The selective effects observed for angry vocalizations suggest that the salience of negative vocalizations, which are putatively more relevant for survival (Vaish, Grossmann, & Woodward, 2008), is particularly enhanced by spatial cues when an explicit evaluation of the emotional properties of the voice is demanded. However, the discrimination of vocalizations from front versus back locations was particularly difficult, and therefore these findings should be interpreted with caution.

Task Effects

Selective attention supports both sound localization and recognition (e.g., Ahveninen et al., 2006). We did not confirm our hypothesis that selective attention to sound identity versus location would modulate spatial and nonspatial processing in a task-dependent fashion. Instead, we observed that task requirements mainly affected the processing of spatial cues: Irrespective of valence, the P2 was increased in response to vocalizations presented at front (vs. back) locations only when the task involved emotion categorization. This finding is consistent with evidence from human neuropsychological studies showing that task demands may result in a differential activation of the "what" and "where" pathways (Hart, Palmer, & Hall, 2004; Alain, Arnott, Hevenor, Graham, & Grady, 2001; Maeder et al., 2001).

Early appraisal processes indexed by the P2 are driven by factors that signal salience or motivational relevance (Pell & Kotz, 2021), including spatial cues. Spatial location may enhance sound salience through, for example, its effects on perceived arousal (Tajadura-Jiménez et al., 2010). Given the sensitivity of the P2 to stimulus arousal (Pell et al., 2015; Paulmann et al., 2013; Kotz & Paulmann, 2011), the increased P2 to front vocalizations aligns with behavioral ratings of vocalizations presented at front locations as more arousing than those presented at the rear. Representations of auditory space may therefore have been enhanced by top–down attention to nonspatial sound properties.

Limitations and Future Directions

One limitation of the present study was the difficulty of the spatial location task compared with the emotion recognition task (see Table 7). Future studies should employ paradigms with comparable task demands to control for the effects of task difficulty. Furthermore, a within-subjects design would have allowed a more direct comparison of spatial and emotional voice processing under different attentional requirements, thereby accounting for interindividual differences. Future studies should probe the precise relationship between emotion and the putative "what"/"where" pathways of the auditory system using a within-subjects design and tasks with comparable difficulty. Furthermore, future electrophysiological investigations should also assess emotion and space interactions employing different spatial manipulations (i.e., looming or distance), as these cues are also perceived as highly salient and relevant for survival (Ceravolo et al., 2016b; Bach, Neuhoff, Perrig, & Seifritz, 2009).

Conclusions

The current findings reveal that emotional cues are prioritized over spatial cues during voice perception. Emotion effects were robustly observed across the two spatial axes within 100 msec post-voice onset, but space effects only emerged 100 msec later. This suggests that *what* is faster than *where* in vocal emotional perception. An interaction between spatial and nonspatial vocal cues was observed at the salience detection stage (~200 msec poststimulus onset), confirming the sensitivity of the P2 to distinct salience cues, with emotion arguably having the greatest weight.

Our findings also show that auditory emotional information does not directly result in augmented stimulus representation in space. At the brain level, spatialized emotional vocalizations were associated with enhanced sensory processing, salience detection, and cognitive analysis, irrespective of their spatial location and task instructions. These findings extend prior evidence showing that auditory processing pathways show differential responding to emotional information and, specifically, that emotion modulates activity in the "what" but not "where" auditory processing pathway.

Differences in how the neurocognitive system responds to emotional attributes of voices could nevertheless affect explicit spatial discrimination and affective ratings of voices. Returning to the example described in the Introduction section, the present findings suggest that a scream will likely be prioritized in perception compared with a neutral sound, its perceived negativity and arousal will be amplified as a function of speaker location, but it may not necessarily be associated with a facilitated location of the speaker. Shedding light on the "what" and "where" of vocal emotion, these results inform current models of vocal emotional communication.

APPENDIX

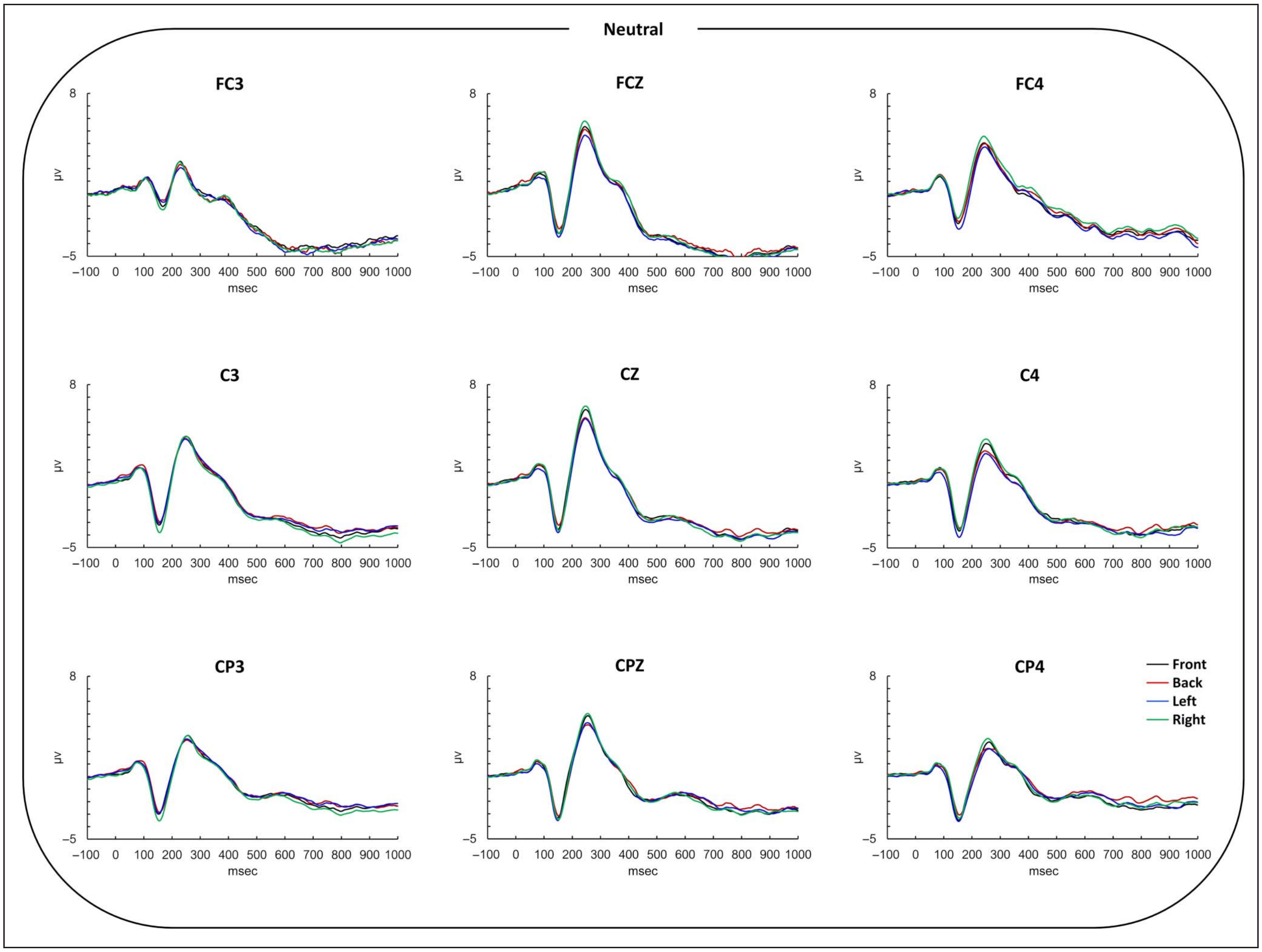

Figure A1. Grand average waveforms contrasting neutral vocalizations presented at the front, back, left, and right locations over nine electrode sites. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

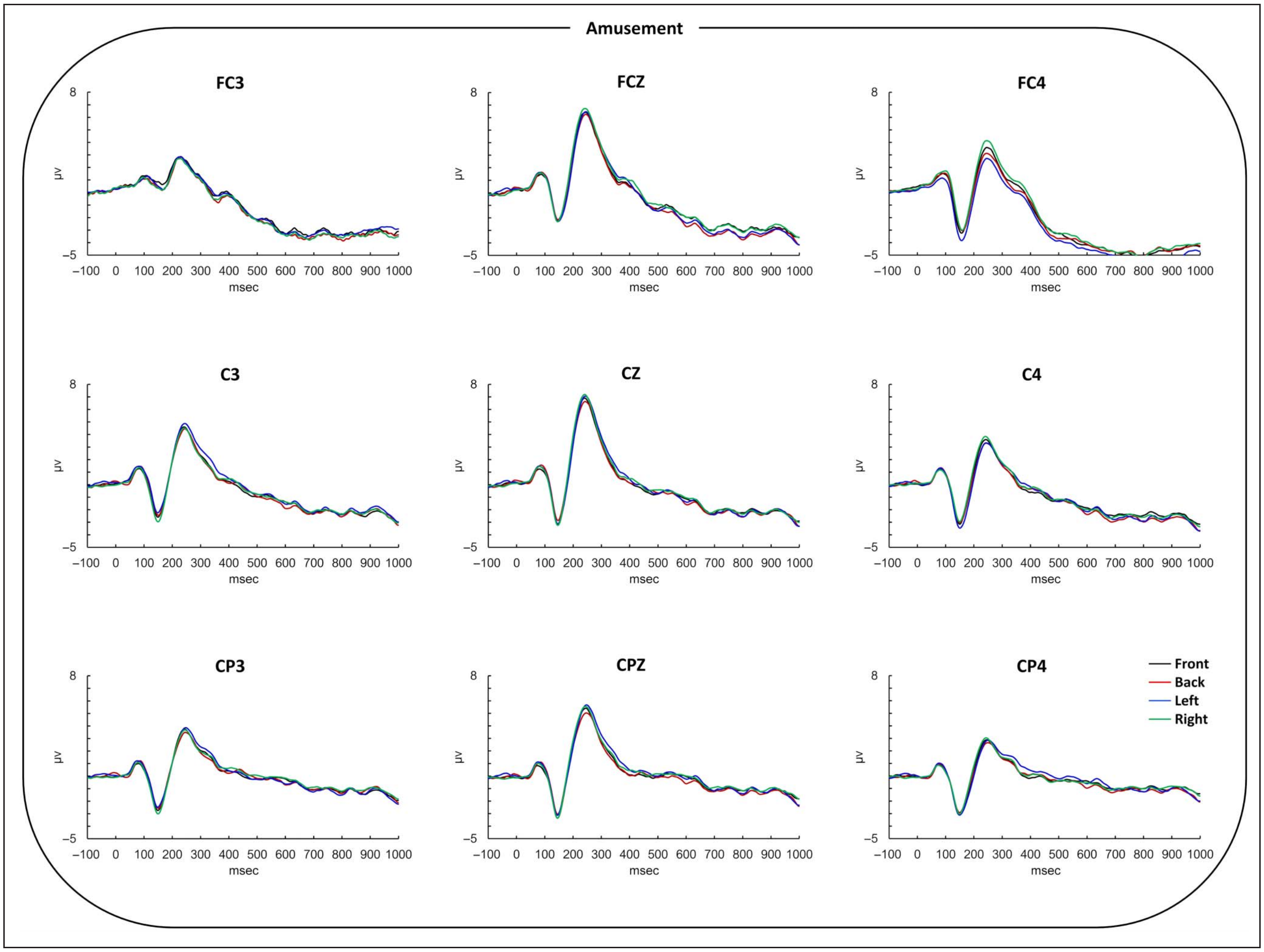

Figure A2. Grand average waveforms contrasting amused vocalizations presented at the front, back, left, and right locations over nine electrode sites. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

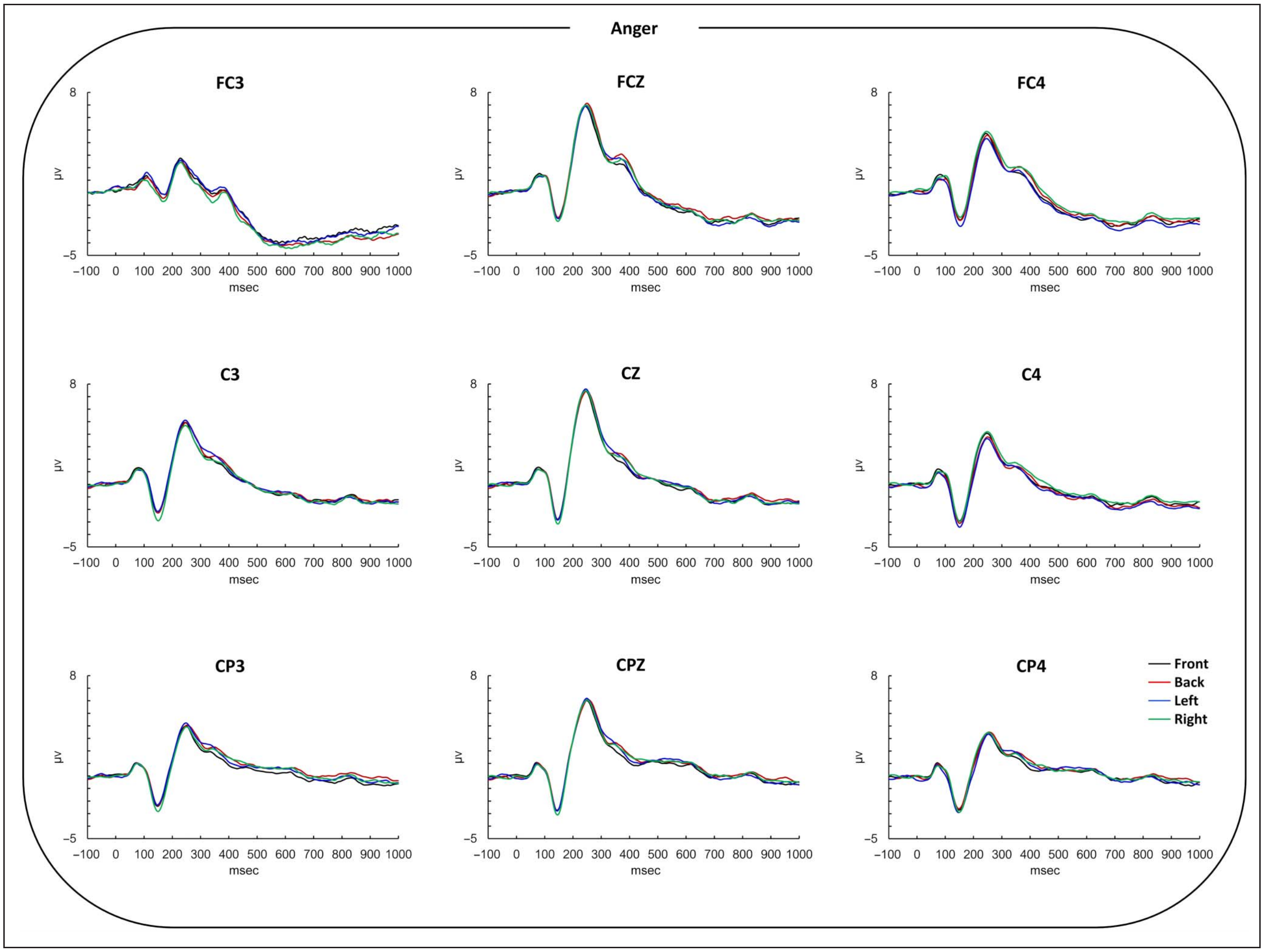

Figure A3. Grand average waveforms contrasting angry vocalizations presented at the front, back, left, and right locations over nine electrode sites. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

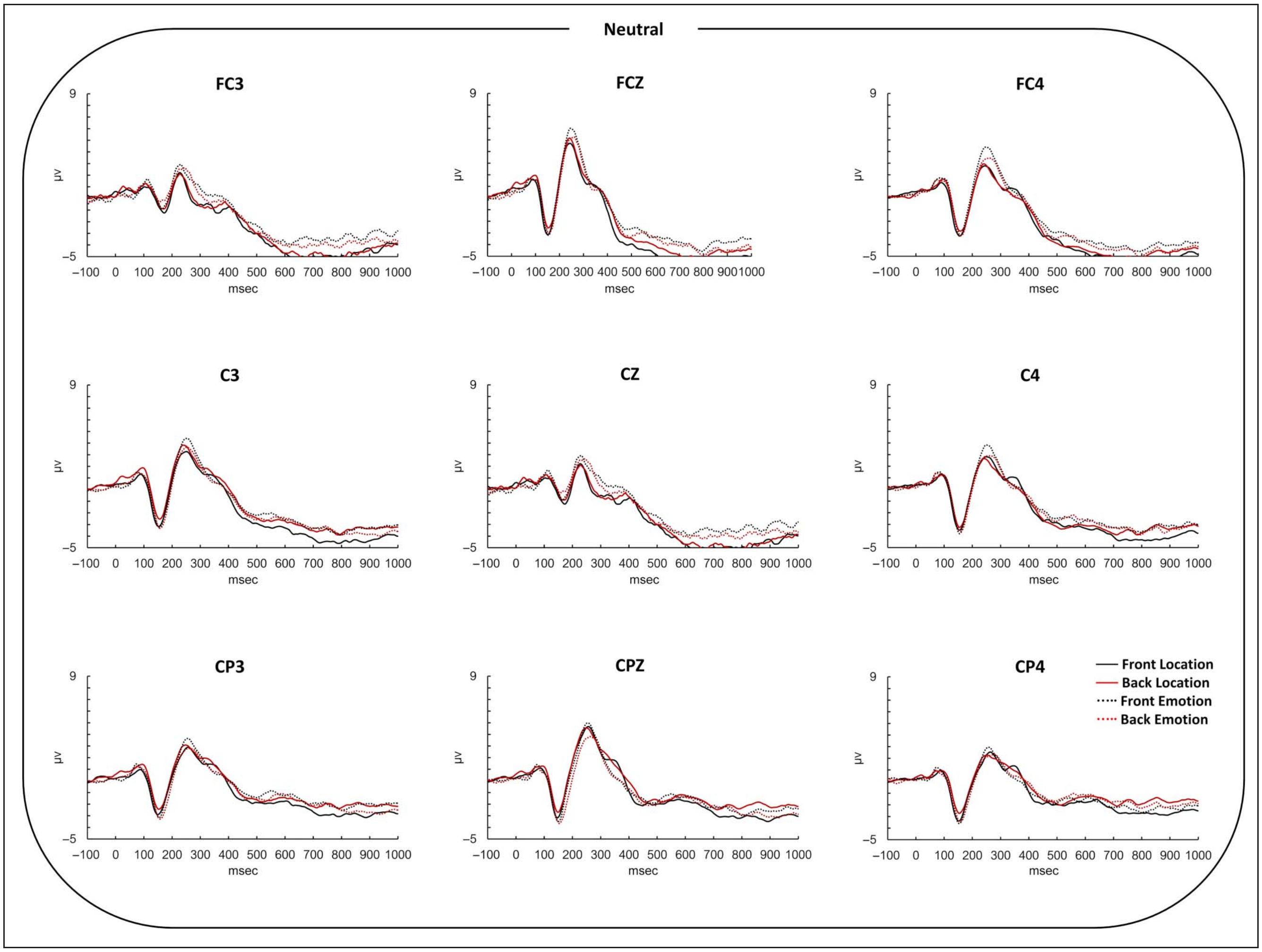

Figure A4. Grand average waveforms contrasting neutral vocalizations presented at the front and back locations in the spatial location discrimination and emotion recognition tasks. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

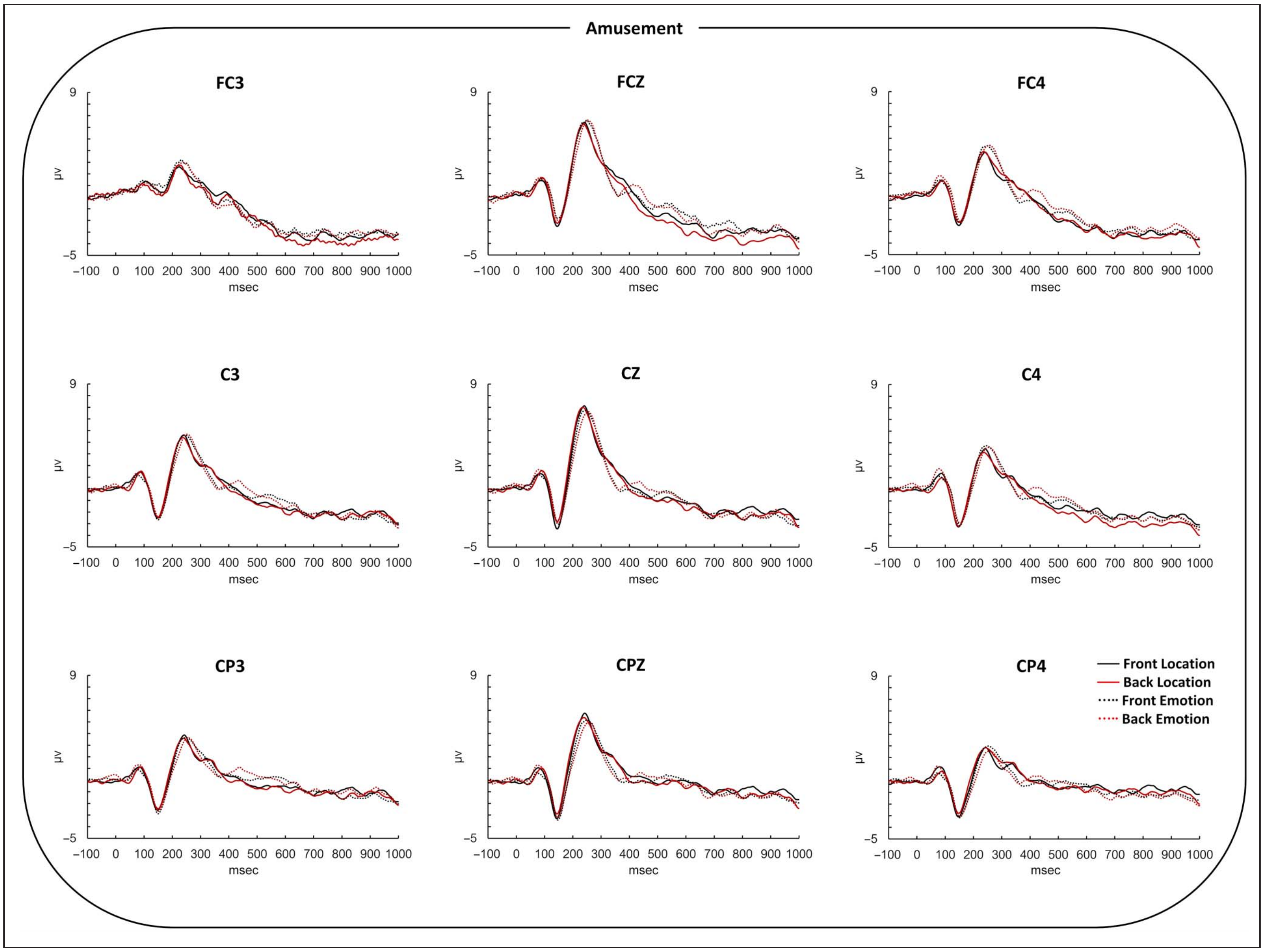

Figure A5. Grand average waveforms contrasting amused vocalizations presented at the front and back locations in the spatial location discrimination and emotion recognition tasks. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

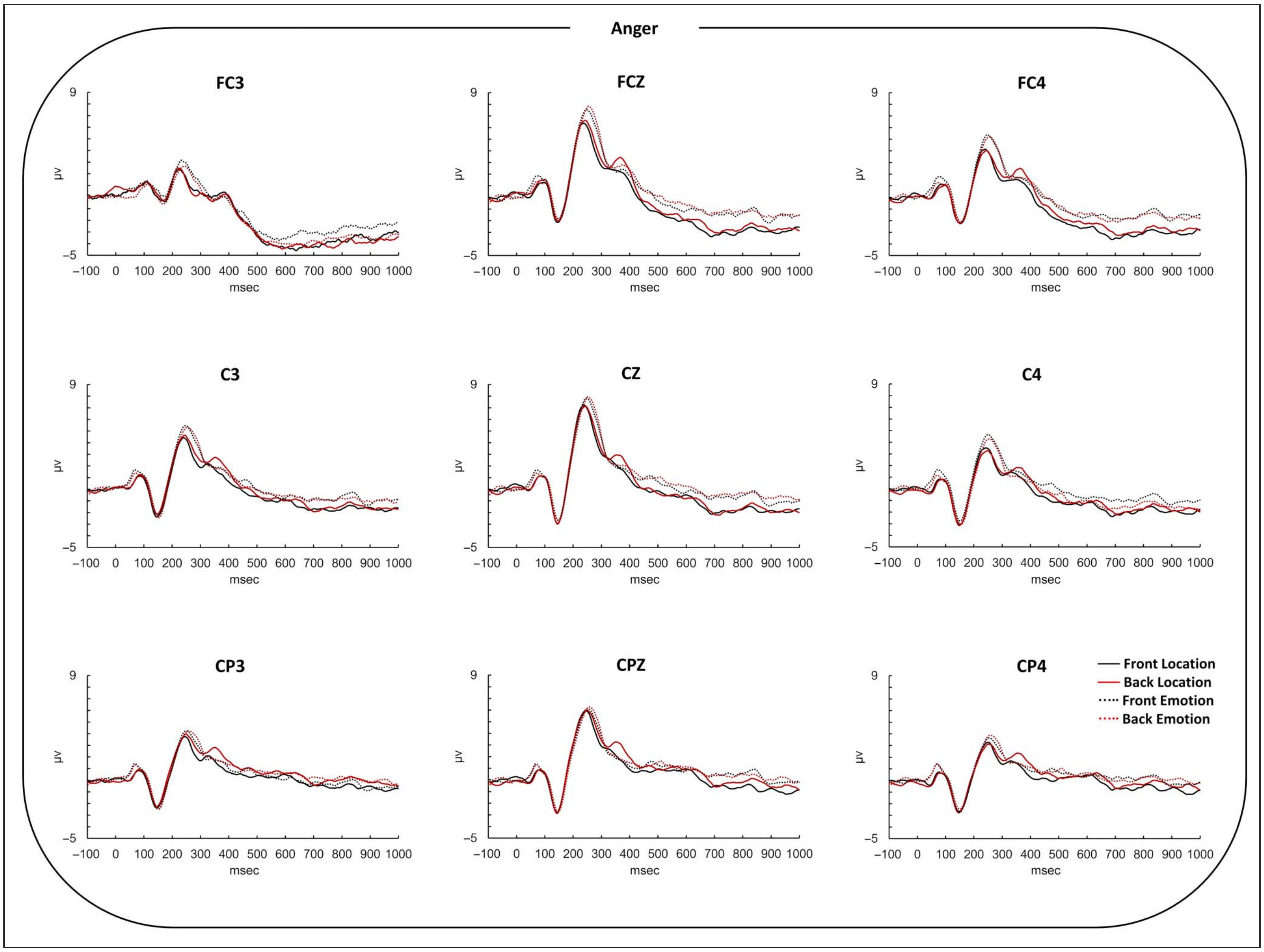

Figure A6. Grand average waveforms contrasting angry vocalizations presented at the front and back locations in the spatial location discrimination and emotion recognition tasks. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

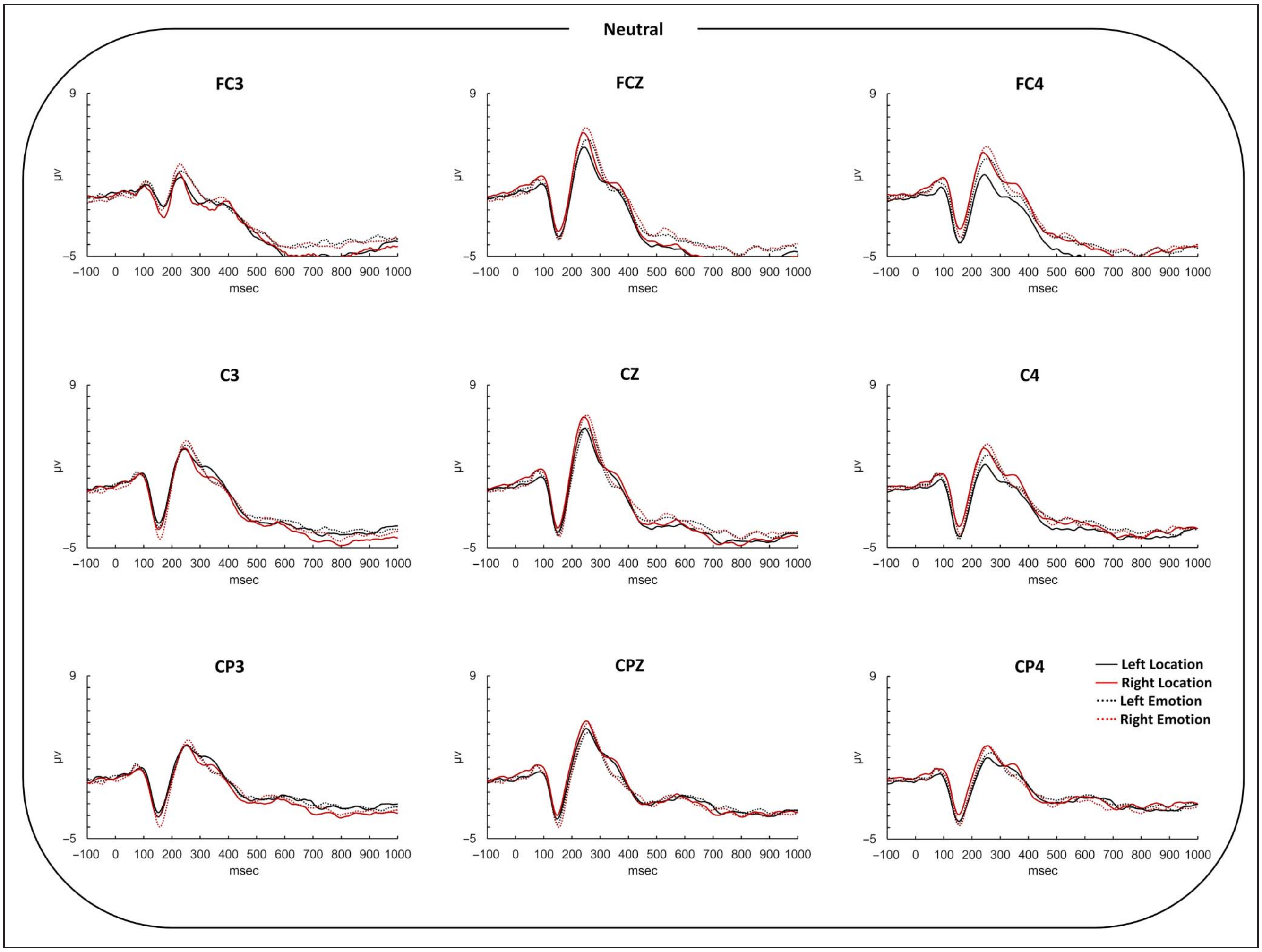

Figure A7. Grand average waveforms contrasting neutral vocalizations presented at the left and right locations in the spatial location discrimination and emotion recognition tasks. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

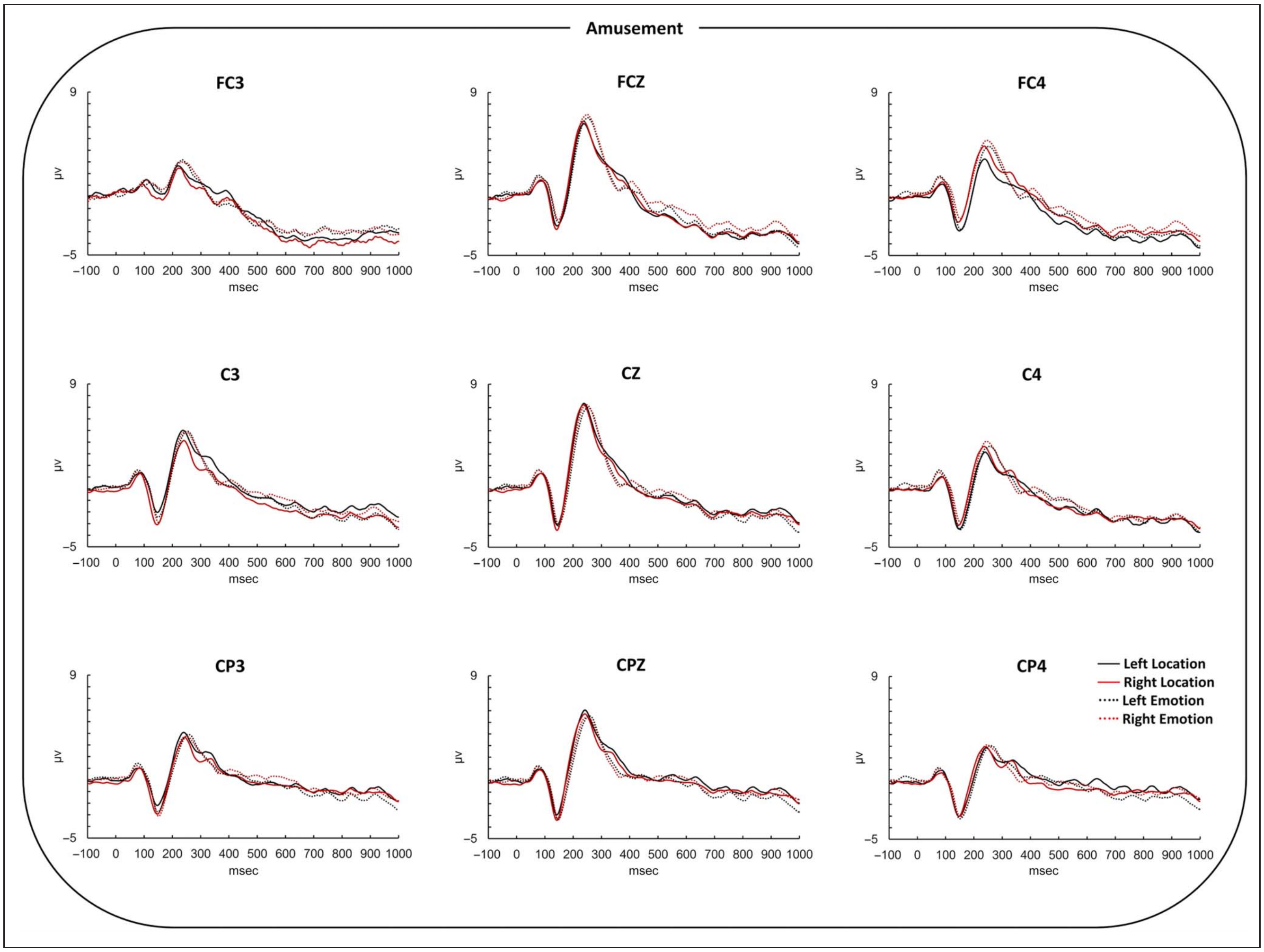

Figure A8. Grand average waveforms contrasting amused vocalizations presented at the left and right locations in the spatial location discrimination and emotion recognition tasks. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

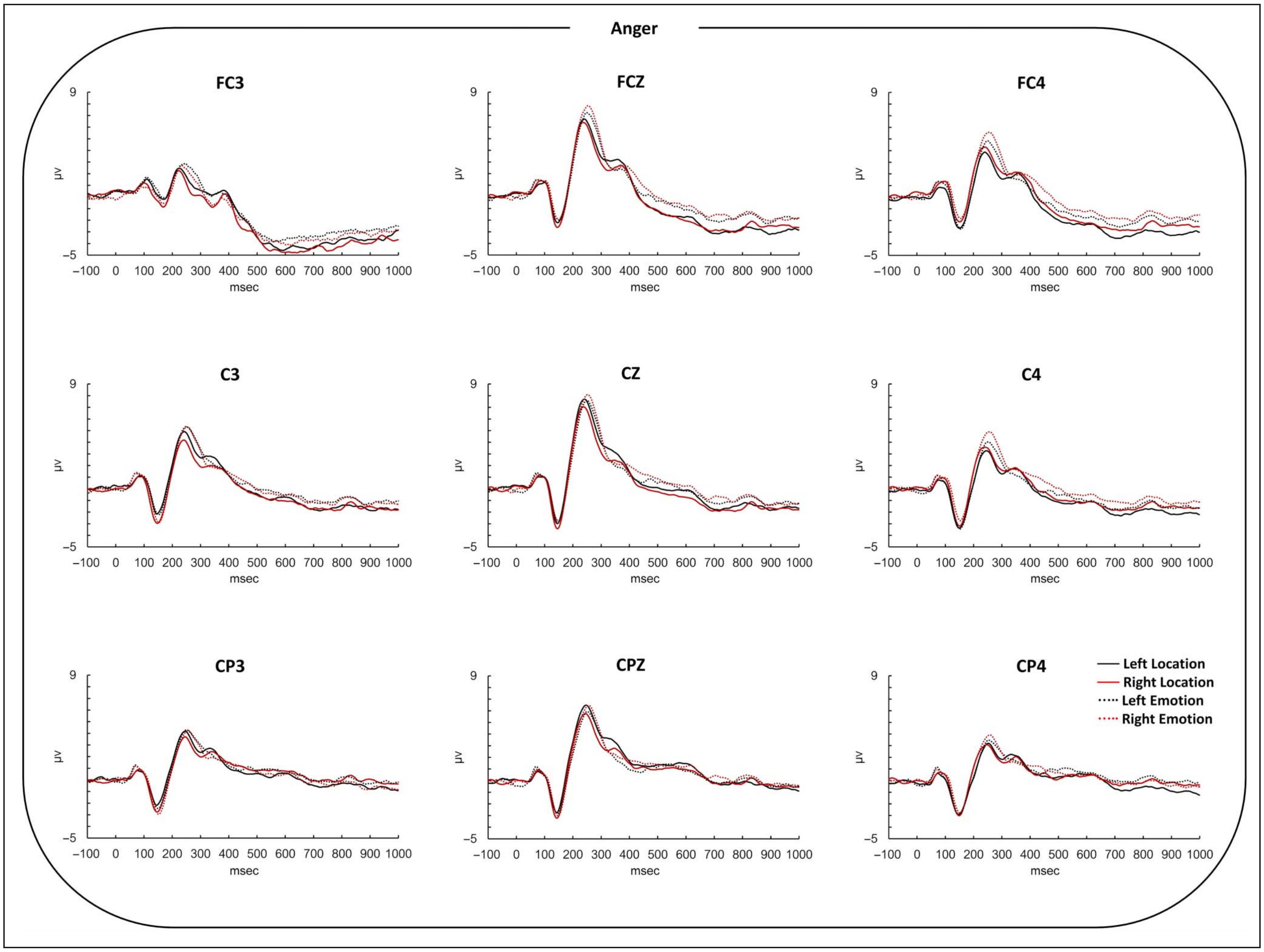

Figure A9. Grand average waveforms contrasting angry vocalizations presented at the left and right locations in the spatial location discrimination and emotion recognition tasks. A high cutoff filter of 20 Hz was applied to the grand average waveforms for visualization purposes.

Acknowledgments

The authors thank all the participants who collaborated in the study and particularly Marta Pereira, Maria Amorim, and João Sarzedas for collaboration in data collection.

Corresponding author: Ana P. Pinheiro, Faculty of Psychology, University of Lisbon, Alameda da Universidade, 1649-013 Lisboa, Portugal, via e-mail: appinheiro@psicologia.ulisboa.pt.

Data Availability Statement

The conditions of our ethics approval do not permit sharing of the data supporting the conclusions in this study with any individual outside the author team under any circumstances. The scripts of the experimental tasks are available at https://osf.io/7rwzm/.

Author Contributions

Both authors designed the study, wrote the protocol, and managed the literature searches. Sara Temudo collected and analyzed the data, and undertook the statistical analyses, under Ana P. Pinheiro's supervision. Both authors wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Funding Information

This work was supported by the Portuguese Science Foundation (Fundação para a Ciência e a Tecnologia; https://dx.doi.org/10.13039/501100001871; grant number: PTDC/MHC-PCN/0101/2014, awarded to A. P. P.) and BIAL Foundation (https://dx.doi.org/10.13039/501100005032) grant number: BIAL 146/2020, awarded to A. P. P. This work received Portuguese national funding from FCT – Fundação para a Ciência e a Tecnologia, I.P, through the Research Center for Psychological Science of the Faculty of Psychology, University of Lisbon (UIDB/04527/2020; UIDP/04527/2020) and DOI identifier https://doi.org/10.54499/UIDB/04527/2020 e https://doi.org/10.54499/UIDP/04527/2020.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

Note

1. The vocalizations selected were: Anger - Anger_C_5, Anger_C_3, Anger_C_2, Anger_C_1, Anger_C_4, Anger_C_6, Anger_T_11, Anger_T_12, Anger_M_7, Anger_M_8; Amusement - Amusement_C_3, Amusement_C_4, Amusement_MS_10, Amusement_C_2, Amusement_C_1, Amusement_T_14, Amusement_T_16, Amusement_M_5, Amusement_M_6, Amusement_T_15; Neutral - 6_neutral, 42_neutral, 45_neutral, 46_neutral, 53_neutral, 55_neutral, 58_neutral, 59_neutral, 60_neutral, 61_neutral.

REFERENCES

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M., et al. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 14608–14613. https://doi.org/10.1073/pnas.0510480103, PubMed: 16983092

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 12301–12306. https://doi.org/10.1073/pnas.211209098, PubMed: 11572938

Amorim, M., & Pinheiro, A. P. (2019). Is the sunny side up and the dark side down? Effects of stimulus type and valence on a spatial detection task. *Cognition and Emotion*, *33*, 346–360. https://doi.org/10.1080/02699931.2018.1452718, PubMed: 29564964

Anourova, I., Nikouline, V. V., Ilmoniemi, R. J., Hotta, J., Aronen, H. J., & Carlson, S. (2001). Evidence for dissociation of spatial and nonspatial auditory information processing. *Neuroimage*, *14*, 1268–1277. https://doi.org/10.1006/nimg.2001.0903, PubMed: 11707083

Asutay, E., & Västfjäll, D. (2015). Attentional and emotional prioritization of the sounds occurring outside the visual field. *Emotion*, *15*, 281–286. https://doi.org/10.1006/nimg.2001.0903, PubMed: 25603132

Bach, D. R., Neuhoff, J. G., Perrig, W., & Seifritz, E. (2009). Looming sounds as warning signals: The function of motion cues. *International Journal of Psychophysiology*, *74*, 28–33. https://doi.org/10.1016/j.ijpsycho.2009.06.004, PubMed: 19615414

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*, 129–135. https://doi.org/10.1016/j.tics.2004.01.008, PubMed: 15301753

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*, 531–539. https://doi.org/10.3758/brm.40.2.531, PubMed: 18522064

Bertels, J., Kolinsky, R., Coucke, D., & Morais, J. (2013). When a bang makes you run away: Spatial avoidance of threatening environmental sounds. *Neuroscience Letters*, *535*, 78–83. https://doi.org/10.1016/j.neulet.2012.12.058, PubMed: 23328437

Bourquin, N. M.-P., Murray, M. M., & Clarke, S. (2013). Location-independent and location-linked representations of sound objects. *Neuroimage*, *73*, 40–49. https://doi.org/10.1016/j.neuroimage.2013.01.026, PubMed: 23357069

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*, 49–59. https://doi.org/10.1016/0005-7916(94)90063-9, PubMed: 7962581

Burra, N., Kerzel, D., Munoz Tord, D., Grandjean, D., & Ceravolo, L. (2019). Early spatial attention deployment toward and away from aggressive voices. *Social Cognitive and Affective Neuroscience*, *14*, 73–80. https://doi.org/10.1093/scan/nsy100, PubMed: 30418635

Campbell, J. I. D., & Thompson, V. A. (2012). MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods*, *44*, 1255–1265. https://doi.org/10.3758/s13428-012-0186-0, PubMed: 22437511

Canavarro, M. C. (1999). Inventário de Sintomas Psicopatológicos [Brief Symptoms Inventory]: BSI. In R. S. Mário, S. A. Leandro, & M. G. Miguel (Eds.), *Testes e provas psicológicas em Portugal: Vol. II* (pp. 95–109). Braga: SHO/APPORT.

Ceravolo, L., Frühholz, S., & Grandjean, D. (2016a). Modulation of auditory spatial attention by angry prosody: An fMRI auditory dot-probe study. *Frontiers in Neuroscience*, *10*, 216. https://doi.org/10.3389/fnins.2016.00216, PubMed: 27242420

Ceravolo, L., Frühholz, S., & Grandjean, D. (2016b). Proximal vocal threat recruits the right voice-sensitive auditory cortex. *Social Cognitive and Affective Neuroscience*, *11*, 793–802. https://doi.org/10.1093/scan/nsw004, PubMed: 26746180

Clarke, S., Bellmann, A., Meuli, R. A., Assal, G., & Steck, A. J. (2000). Auditory agnosia and auditory spatial deficits following left hemispheric lesions: Evidence for distinct processing pathways. *Neuropsychologia*, *38*, 797–807. https://doi.org/10.1016/s0028-3932(99)00141-4, PubMed: 10689055

Conde, T., Correia, A. I., Roberto, M. S., Scott, S. K., Lima, C. F., & Pinheiro, A. P. (2022). The time course of emotional authenticity detection in nonverbal vocalizations. *Cortex*, *151*, 116–132. https://doi.org/10.1016/j.cortex.2022.02.016, PubMed: 35405538

Derey, K., Rauschecker, J. P., Formisano, E., Valente, G., & de Gelder, B. (2017). Localization of complex sounds is modulated by behavioral relevance and sound category. *Journal of the Acoustical Society of America*, *142*, 1757. https://doi.org/10.1121/1.5003779, PubMed: 29092572

Du, F., Wang, X., Abrams, R. A., & Zhang, K. (2017). Emotional processing is enhanced in peri-hand space. *Cognition*, *165*, 39–44. https://doi.org/10.1016/j.cognition.2017.04.009, PubMed: 28486136

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews*, *68*, 96–110. https://doi.org/10.1016/j.neubiorev.2016.05.002, PubMed: 27189782

Gädeke, J. C., Föcker, J., & Röder, B. (2013). Is the processing of affective prosody influenced by spatial attention? An ERP study. *BMC Neuroscience*, *14*, 14. https://doi.org/10.1186/1471-2202-14-14, PubMed: 23360491

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484. https://doi.org/10.1016/0013-4694(83)90135-9, PubMed: 6187540

Grisendi, T., Clarke, S., & Da Costa, S. (2023). Emotional sounds in space: Asymmetrical representation within early-stage auditory areas. *Frontiers in Neuroscience*, *17*, 1164334. https://doi.org/10.3389/fnins.2023.1164334, PubMed: 37274197

Hart, H. C., Palmer, A. R., & Hall, D. A. (2004). Different areas of human non-primary auditory cortex are activated by sounds with spatial and nonspatial properties. *Human Brain Mapping*, *21*, 178–190. https://doi.org/10.1002/hbm.10156, PubMed: 14755837

Ho, H. T., Schröger, E., & Kotz, S. A. (2015). Selective attention modulates early human evoked potentials during emotional face-voice processing. *Journal of Cognitive Neuroscience*, *27*, 798–818. https://doi.org/10.1162/jocn_a_00734, PubMed: 25269113

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*, 242–249. https://doi.org/10.1093/scan/nsl027, PubMed: 17607327

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*, 770–814. https://doi.org/10.1037/0033-2909.129.5.770, PubMed: 12956543

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, PubMed: 17391654

Koiwa, N., Masaoka, Y., Kusumi, T., & Homma, I. (2010). Sound localization difficulty affects early and late processing of auditory spatial information: Investigation using the dipole tracing method. *Clinical Neurophysiology*, *121*, 1526–1539. https://doi.org/10.1016/j.clinph.2010.03.016, PubMed: 20378400

Kong, L., Michalka, S. W., Rosen, M. L., Sheremata, S. L., Swisher, J. D., Shinn-Cunningham, B. G., et al. (2014). Auditory spatial attention representations in the human cerebral cortex. *Cerebral Cortex*, *24*, 773–784. https://doi.org/10.1093/cercor/bhs359, PubMed: 23180753

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, *5*, 108–125. https://doi.org/10.1111/j.1749-818X.2010.00267.x

Kryklywy, J. H., Macpherson, E. A., Greening, S. G., & Mitchell, D. G. V. (2013). Emotion modulates activity in the "what" but not "where" auditory processing pathway. *Neuroimage*, *82*, 295–305. https://doi.org/10.1016/j.neuroimage.2013.05.051, PubMed: 23711533

Lewald, J., & Getzmann, S. (2011). When and where of auditory spatial processing in cortex: A novel approach using electrotomography. *PLoS One*, *6*, e25146. https://doi.org/10.1371/journal.pone.0025146, PubMed: 21949873

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods*, *45*, 1234–1245. https://doi.org/10.3758/s13428-013-0324-3, PubMed: 23444120

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*, 108–112. https://doi.org/10.1097/WNR.0b013e32834ea757, PubMed: 22134115

Luck, S. J. (2005). An introduction to event-related potentials and their neural origins. In *An introduction to the event-related potential technique* (pp. 1–50). MIT Press.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J.-P., et al. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *Neuroimage*, *14*, 802–816. https://doi.org/10.1006/nimg.2001.0888, PubMed: 11554799

Mostardinha, A. R., Bártolo, A., Bonifácio, J., & Pereira, A. (2019). Validation of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) among university students. *Acta Medica Portuguesa*, *32*, 279–288. https://doi.org/10.20344/amp.10650, PubMed: 31067422

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4, PubMed: 5146491

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, PubMed: 23801973

Paulmann, S., & Kotz, S. A. (2008). Early emotional prosody perception based on different speaker voices. *NeuroReport*, *19*, 209–213. https://doi.org/10.1097/WNR.0b013e3282f454db, PubMed: 18185110

Paulmann, S., Ott, D. V. M., & Kotz, S. A. (2011). Emotional speech perception unfolding in time: The role of the basal ganglia. *PLoS One*, *6*, e17694. https://doi.org/10.1371/journal.pone.0017694, PubMed: 21437277

Pell, M. D., & Kotz, S. A. (2021). Comment: The next frontier: Prosody research gets interpersonal. *Emotion Review*, *13*, 51–56. https://doi.org/10.1177/1754073920954288

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology*, *111*, 14–25. https://doi.org/10.1016/j.biopsycho.2015.08.008, PubMed: 26307467

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, PubMed: 28942367

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*, 127–139. https://doi.org/10.1093/scan/nsv103, PubMed: 26468268

Pinheiro, A. P., Lima, D., Albuquerque, P. B., Anikin, A., & Lima, C. F. (2019). Spatial location and emotion modulate voice perception. *Cognition and Emotion*, *33*, 1577–1586. https://doi.org/10.1080/02699931.2019.1586647, PubMed: 30870109

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, PubMed: 26894680

Pinheiro, A. P., Sarzedas, J., Roberto, M. S., & Kotz, S. A. (2023). Attention and emotion shape self-voice prioritization in speech processing. *Cortex*, *158*, 83–95. https://doi.org/10.1016/j.cortex.2022.10.006, PubMed: 36473276

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30. https://doi.org/10.1016/j.tics.2005.11.009, PubMed: 16321562

Spreckelmeyer, K. N., Kutas, M., Urbach, T., Altenmüller, E., & Münte, T. F. (2009). Neural processing of vocal emotion and identity. *Brain and Cognition*, *69*, 121–126. https://doi.org/10.1016/j.bandc.2008.06.003, PubMed: 18644670

Tajadura-Jiménez, A., Larsson, P., Väljamäe, A., Västfjäll, D., & Kleiner, M. (2010). When room size matters: Acoustic influences on emotional responses to sounds. *Emotion*, *10*, 416–422. https://doi.org/10.1037/a0018423, PubMed: 20515229

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin*, *134*, 383–403. https://doi.org/10.1037/0033-2909.134.3.383, PubMed: 18444702

Valdés-Conroy, B., Sebastián, M., Hinojosa, J. A., Román, F. J., & Santaniello, G. (2014). A close look into the near/far space division: A real-distance ERP study. *Neuropsychologia*, *59*, 27–34. https://doi.org/10.1016/j.neuropsychologia.2014.04.009, PubMed: 24747210

Van Strien, J. W., De Sonneville, L. M. J., & Franken, I. H. A. (2010). The late positive potential and explicit versus implicit processing of facial valence. *NeuroReport*, *21*, 656–661. https://doi.org/10.1097/WNR.0b013e32833ab89e, PubMed: 20453693

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*, 239–267. https://doi.org/10.1007/s10919-017-0253-4

Xie, J., Wang, R., & Chang, S. (2014). The mechanism of valence-space metaphors: ERP evidence for affective word processing. *PLoS One*, *9*, e99479. https://doi.org/10.1371/journal.pone.0099479, PubMed: 24923271

Zündorf, I. C., Lewald, J., & Karnath, H.-O. (2016). Testing the dual-pathway model for auditory processing in human cortex. *Neuroimage*, *124*, 672–681. https://doi.org/10.1016/j.neuroimage.2015.09.026, PubMed: 26388552