Neural dynamics of induced throat vibrations during vocal emotion recognition

Damien Benis, PhD (a,b)*

Garance Selosse MSc (a, b)*

Jennifer Greilsamer, MSc (c)

Ben Meuleman, PhD (b)

Didier Grandjean, PhD† (a, b)

Leonardo Ceravolo, PhD† (a, b)

* Shared first authors

† Shared last authors

(a) *Department of Psychology, University of Geneva, Geneva, Switzerland,*

(b) *Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland,*

(c) *Department of Medicine, University of Fribourg, Fribourg, Switzerland*

Corresponding author:

Leonardo Ceravolo, PhD
University of Geneva
Boulevard Pont-d'Arve 40
CH-1205 Switzerland
email: leonardo.ceravolo@unige.ch

Abstract

Despite a large corpus of literature in psychological and brain mechanisms on emotional prosody perception, the perspective of embodied cognition in these mechanisms have been largely neglected. Here we investigated the influence of induced bodily vibrations on the categorization of ambiguous emotional vocalizations using event-related potentials (ERPs). Emotional voices were morphed between a fearful expression with the speaker's identity-matching angry expression, creating blends of emotions in each voice. Emotional congruent and incongruent vibrations were delivered on the skin close to the vocal cords. Congruent with our hypotheses, behavioural results revealed that induced vibrations skewed the participants' emotional ratings by biasing responses towards the vibration's emotion. ERPs indicated that N100 and P200 components subtending the early processing of emotional prosody were significantly modulated by induced vibrations in the congruent setting, considered as a facilitation effect for emotion recognition at early stages of processing. A modulation of the late positive component was also observed in the incongruent setting, suggesting an error processing mechanism. Source reconstruction highlighted effects of vibration types in prefrontal, motor, somatosensory, and insular cortices. Our results suggest that voice-associated vibrations may play a significant role in vocal emotion processing and recognition through an embodied mechanism.

Keywords: *voice prosody, emotion, bodily vibrations, feedback, embodiment, electroencephalography*

Word count (excluding abstract, references, tables, legends): 4996

Introduction

Verbal and non-verbal communication is essential for humans. Emotional prosody represents the temporal fluctuations of the pitch, amplitude and spectrum (Grandjean et al., 2006; Laukka et al., 2005; Risberg & Lubker, 1978; Scherer & Oshinsky, 1977). Processing emotional prosody variations allows listeners to subjectively attribute the emotional state of speakers, influenced by context (Brunswik, 1956; Grandjean et al., 2006; Scherer, 2003). One key mechanism is the embodied 'simulation' of what we hear (Hawk et al., 2012), and emotional 'embodiment' is defined as the motor, sensory and perceptual reexperiencing or simulating of the emotion perceived in a peer in the framework of the Simulation of Smiles (SIMS) model (Niedenthal, 2007; Niedenthal et al., 2010). This process has primarily been described in the visual modality, with the perception of facial expressions in others generating facial mimicry in the observer—stronger for emotional vs. non-emotional faces (Chartrand & Bargh, 1999; Dimberg, 1982; Korb et al., 2010; Künecke et al., 2017; Moody & McIntosh, 2011; Moody et al., 2007). Recent studies have supported the idea that facial mimicry represents context-dependent sensorimotor simulation of an observed emotion rather than a mere muscular reproduction of an observed facial expression (Borgomaneri et al., 2020; Drimalla et al., 2019; Hess & Blairy, 2001; Hess & Fischer, 2014; Philip et al., 2018). Importantly, some studies suggest that this process may impact the accurate decoding of facial emotional information especially in women (Hyniewska & Sato, 2015; Korb et al., 2010; Schneider et al., 2013; Stel & van Knippenberg, 2008), although not exclusively (Blairy et al., 1999; Hess & Blairy, 2001; Holland et al., 2021).

While studies on the contribution of embodiment in the field of emotional prosody decoding are scarcer than in the visual domain, some point to a contribution of this mechanism in the auditory modality. Indeed, several studies showed that passive listening of speech signals elicits activations in motor regions of the brain similar to those involved during speech production (Fadiga et al., 2002; Watkins et al., 2003; Wilson et al., 2004). Facilitated language comprehension when the primary motor areas controlling the lips/tongue were stimulated transcranially was also observed (D'Ausilio et al., 2009). One mechanism underlying emotion embodiment in voice signals may be body resonances or vibrations. In fact, vibrations—especially close to the vocal folds—may represent a reliable interoceptive and/or exteroceptive feedback during vocal emotion expression. Indeed, during voice expression, vibrations from the vocal cords are transmitted through the speaker's skin and body tissues (Munger & Thomson, 2008; Švec et al., 2005), with a frequency corresponding

to the fundamental frequency of the produced auditory tone (Sundberg, 1992). These bodily vibrations would constitute, for the speaker, a dynamic feedback on several characteristics of his/her own voice (Sundberg, 1992). Lylykangas and colleagues (2009) highlighted the potential to successfully use vibrotactile feedback as relevant information for speed regulation messages. Similarly, Tuuri and colleagues (2010) compared audio and tactile feedback from the same speech stimuli as speed regulation cues and found that the information was delivered in a similar manner across modalities. However, there is still a paucity of research in this field and we intend to fill this gap. To investigate precise dynamic neural correlates of the embodiment of vocal emotion perception concurrent to induced throat vibrations, we used surface electroencephalography (EEG) using event-related potentials (ERPs) as the underlying dependent measure.

ERP correlates of emotional prosody decoding have been extensively studied, with evidence for a three-stage processing (Schirmer & Kotz, 2006). In line with this model, a wealth of ERP studies has shown the implication of several early and late ERP components in the processing of emotional vocalizations. Early components underlie auditory-sensory processing, attention and early recognition of meaning around 100ms after stimulus onset (N100, frontocentral area). It might reflect early low-level prosody analysis and emotion discrimination (Gädeke et al., 2013; Paquette et al., 2020; Tarai & Srinivasan, 2019). Basic emotions are also distinguished in the frontal/parietal P200 component (Duville et al., 2022; Gädeke et al., 2013; Maltezou-Papastylianou et al., 2022; Paulmann et al., 2013; Paulmann & Uskul, 2017). The P200 is a positive deflection around 200 milliseconds reflecting increased attention to emotional stimuli for facilitated processing (Paulmann & Kotz, 2008; Thomas et al., 2007). Later components are commonly associated with higher-level mechanisms of emotional prosody processing such as appraisal, contextual relevance and mental representations (Duville et al., 2022; Pell & Kotz, 2021). This includes ~300-400 milliseconds deflections (P300/N400) in frontal/frontocentral areas during emotional prosody processing (Carminati et al., 2018; Proverbio et al., 2020; Tarai & Srinivasan, 2019), underlying deviant auditory stimuli detection (Lindín et al., 2004; Tsolaki et al., 2015). Finally, the Late Positive Component (LPC) is a positive deflection observed at a later stage in frontal/frontocentral and parietal areas, associated with higher-order cognitive processes such as arousal information (Paulmann et al., 2013), expectancy violation in prosody (Astésano et al., 2004; Brattico et al., 2006; Paulmann et al., 2012) and emotional ‘meaning’ (Wei et al., 2022; Zora et al., 2020). The LPC may reflect attentional processes rather than emotional face processing *per se*

(Ashley et al., 2004), consistent with the attentional capture effect associated with violation or incongruence (Vachon et al., 2012). Mismatching prosody stimuli also elicits an LPC, potentially indicating an integration process of the different sources of information (Brattico et al., 2006; Paulmann et al., 2012).

Concerning functional MRI studies, processing emotional prosody activates subcortical and cortical networks (Ceravolo et al., 2021; Grandjean, 2021; Peron et al., 2016; Witteman et al., 2012), including the bilateral superior temporal, inferior frontal (Ceravolo et al., 2026) and orbitofrontal cortex, cerebellum, etc. Additionally, the motor cortex is also involved during speech perception (Fadiga et al., 2002; Watkins et al., 2003) while interoceptive signals trigger insular activity, a region responsible for the integration of internal and external physiological signals and emotional processes (Adolfi et al., 2017; Casals-Gutierrez & Abbey, 2020). The somatosensory cortex or the supplementary motor area (SMA) also play a critical role in interoceptive and proprioceptive processing (Engelen et al., 2023; Gibson, 2019; Khalsa et al., 2018).

Our study aims to explore the potential interoceptive and/or exteroceptive feedbacks represented by bodily resonances—induced external throat vibrations, and their impact on emotional prosody recognition. Revealing such a link between vocal emotion perception and induced vibrations would constitute convincing evidence for a crucial role of vibrotactile feedback as a physiological mechanism of vocal emotion embodiment. We recently demonstrated the neural correlates of measured throat vibrations during emotional voice expression (Selosse et al., 2025) and our results are coherent with above mentioned literature—although pertaining to throat vibrations specifically. Associations between interoception/proprioception and emotional processing were previously studied but never using vibrations concurrent to expressed voice signals (Burleson & Quigley, 2021; Couto et al., 2015; Critchley & Garfinkel, 2017). Other related work includes: interoceptive accuracy (Mendoza-Medialdea & Ruiz-Padial, 2021), ERPs of interoception in emotion processing (P300 component; Herbert et al., 2007; Pollatos et al., 2005), a somatosensory-specific P300 component (parietal regions) involved in somatosensory perception (Al et al., 2020; Auksztulewicz & Blankenburg, 2013; Has et al., 2021), interoceptive priming on faces (Salamone et al., 2021). In addition, other studies further support the idea that emotionally relevant stimuli can modulate attention and early perceptual processing across modalities (Brosch, Grandjean, et al., 2008; Brosch, Sander, et al., 2008)—in the case of the present study, auditory and vibrotactile modalities.

No study has specifically investigated the contribution of throat vibrations to the auditory perception of emotions using EEG. The present study aims to further explore this mechanism. The main mechanistic rationale is that vocal cords vibrations, seen as a 'mechanical' consequence of the production of vocalized signals, would represent a feedback that could impact the auditory perception and recognition of vocal emotions. Namely, according to an embodied perception perspective: as a valid feedback, vibratory signals would influence voice production for the speaker and should therefore also have an impact on perception if induced in the listener. At the behavioral level, we first postulated that: i) vibrations would affect the discriminative perceptual processes of emotional vocalizations (more frequent "anger" response when the vibrations also express anger and likewise for fear) or when compared to the 'no-vibration' condition; ii) these effects would occur especially for ambiguous stimuli (i.e., voices containing 40-60% of each emotion), as these would trigger sensory information seeking for clarification (Niedenthal, 2007; Niedenthal et al., 2010), here induced throat vibrations. At the neural level, we postulated: i) vibration-specific ERPs between 200 and 300ms modulated by interoceptive/proprioceptive processes (Herbert et al., 2007; Mendoza-Medialdea & Ruiz-Padial, 2021; Pollatos et al., 2005) with source modulations in the insular, somatosensory, prefrontal, and cingulate cortices as well as in the SMA/pre-SMA (Adolfi et al., 2017; Casals-Gutierrez & Abbey, 2020; Engelen et al., 2023; Gibson, 2019; Khalsa et al., 2018), for anger or fear vibrations compared to no-vibration; ii) incongruence between emotional voice and vibrations–angry voice with fear vibration and *vice versa*, would trigger early (100ms and 200-300ms) and late (400-700ms) ERPs with source modulations in the above-mentioned brain areas, related to error prediction and/or integration of vibratory feedbacks in the context of emotional prosody information processing.

Method

Participants

Sample size was decided a priori based on repeated measures MANOVA in 'G*Power 3.1' software, with: effect size=.25, power=.8, alpha=.05, number of groups=2 (fixed, no other option here), number of measures=9 and correlation among repeated measures=0.5 (Faul et al., 2007). This approximation—the closest possible to mixed-effects modeling—resulted in a required sample size of N=24. We therefore recruited twenty-four healthy participants (18 females and 6 males; Age=18-41; M_{Age}=23.67; SD_{Age}=5.11). Participants had

normal hearing, no neurologic or psychiatric history and gave informed written consent. The study was approved by state-wise ethics committee (nr. 2020-02262) and conducted according to the declaration of Helsinki.

Material

Voices (‘Aah’, Montreal affective voices database; Belin et al., 2008) were presented through headphones (HD25, Sennheiser, DE) and morphed on the emotion dimension (anger, fear) using Matlab (The Mathworks Inc., Natick, MA, USA) and the PsychToolbox (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Morphing fearful expression with the speaker’s identity-matching angry expression resulted in five conditions of 10 voices each [fear%-anger%: 10-90 (A_{90}), 40-60 (A_{60}), 50-50 (A_{50}), 60-40 (A_{40}), and 90-10 (A_{10})], using the STRAIGHT toolbox (Kawahara & Matsui, 2003). Voice stimuli and vibrations were acoustically assessed (amplitude, pitch, loudness, duration, spectral centroid and slope, F1 frequency; Supplementary Table 1 & 2, respectively). Morphing conditions and actors were randomly distributed across trials. The choice of using anger and fear was due to the fact that these emotions are among the most biologically-relevant for survival.

Throat vibrations were induced using a small vibrator (BC-10, Ortofon, DK; dimensions: 13.5*29.5*18.0mm, weight: 16.5gr, sensitivity: 118dB, harmonic distortion: 1.5%, impedance: 15Ω, sensibility: 100-1000Hz, sampling rate: 1000Hz). These expressed pure anger or pure fear in 2/3rd of the trials (unmorphed ‘Aah’; anger, fear) while no vibrations were induced in the remaining trials (control, 1/3rd). The vibrator was placed close to the vocal cords (Figure 1), taped on the left side to the laryngeal prominence with kinesiology tape. This location fits our general hypothesis that throat vibrations may be used as a vibrotactile feedback. It also matches relevant locations in this context (Munger & Thomson, 2008; Nolan et al., 2009; Sundberg, 1992). The same speaker was used for both vibrator and voice stimuli for the sake of consistency and to avoid intonation issues. There were 20 stimuli for the vibrations (two emotions, ten actors).

Task Procedure

The experiment comprised six runs (two fear-induced, two anger-induced vibrations, two no-vibration) of N=113 randomly presented trials each (Figure 1). Each trial started with a fixation cross (500ms) concurrent to a voice stimulus (1000ms), then explicit emotion categorization for 3000ms (buttons: anger, fear; key labels counterbalanced between). Participants were instructed to answer as fast and as accurately as possible. Vibrations were

induced simultaneously to the voices and were emotionally congruent or incongruent with the voice stimulus. The final factorial design is a 5 EmotionalVoice (A_{90}, A_{60}, A_{50}, A_{40}, A_{10}) by 3 Vibration (No Vibrations, Anger, Fear) within-subjects design.

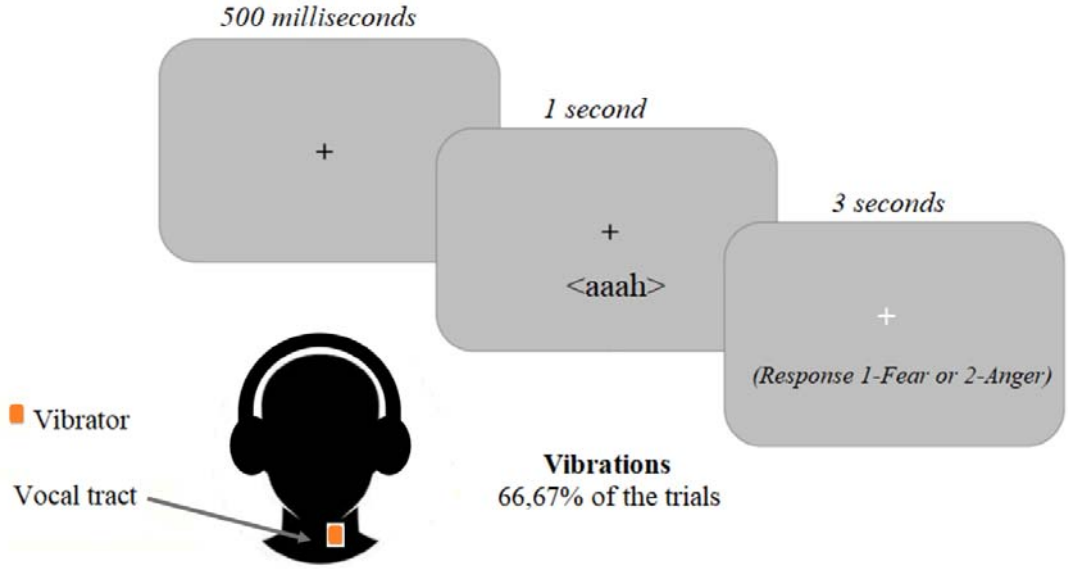

Figure 1. **Experimental design.** Participants were instructed to focus on a black central fixation cross displayed on a grey screen for five hundred milliseconds. Emotional voice stimuli were presented through headphones for one second. Vibrations were induced at the same time in two thirds of the trials and expressed pure anger or pure fear. After each voice stimulus, the cross turned white to indicate that the participant had to categorize the emotion expressed auditorily using the response keys, as fast as accurately as possible and within the 3-seconds response window.

Statistical Analysis

Trial-level behavioral analysis

Responses were categorical: anger=0, fear=1. A Generalized Linear Mixed Model (GLMM) was performed in R through Rstudio (Venables & Smith, 2003) to analyze the responses as a function of Vibration and EmotionalVoice. An LMM was chosen to analyze reaction times with the same factors (no significant effects and not part of our hypotheses; Supplementary Table 3). Both models were computed using the 'lme4' package (Bates et al., 2015). For both models, fixed effects included the interaction between Vibration and Emotional voice. Random effects included the random slope of Vibration and EmotionalVoice as a function of participant identity (uncorrelated) and the random intercept

of stimulus identity, participant and speaker gender, runs. The formula was the following (glmer for response, lmer for reaction times):

Model.response / rt <- glmer / lmer

*(response/rt~Vibration*EmotionalVoice+ParticipantGender+SpeakerGender+(1+Vibration+EmotionalVoice||ParticipantID)+(1|Run)+(1|StimulusID))*

The effect size of each model was computed and labelled according to the thresholds defined by Cohen (1988). Planned contrasts according to our hypotheses use Bonferroni correction for multiple testing (Bland & Altman, 1995) and estimated marginal means package for odds ratios (Lenth et al., 2019).

EEG preprocessing

Surface EEG was acquired with 64 amplified electrodes (Biosemi ActiveThree system, BioSemi B.V., Netherlands), preprocessing was performed using Brainvision analyser 2.2 (Brain Products, Munich, Germany) and Fieldtrip (Oostenveld et al., 2011). Spline interpolation was applied to artifacted channels (5.34±4.74% of channels) and an average reference was applied. Blink and saccade artifacts were rejected using independent component analysis. Raw data were split into two-seconds epochs (-1s to +1s) centered on voice onset. Peak-to-peak amplitude was calculated for all epochs across all channels, and a trial rejection threshold was calculated as the median +1.96 standard deviation of the obtained peak-to-peak amplitude distribution (~114μv). For each channel, epochs exceeding the threshold were rejected (average rejection: 9.44±11.95% trials). For each participant/channel, ERPs were obtained by averaging epochs for each emotional condition and each vibration type, and their conjunction (anger90-fear10 with anger vibration for example). Grand-average ERPs were then obtained by averaging the resulting ERPs across participants for each channel (final average number of trials per participant/condition: 40.95±6.08; Supplementary Table 7). Topographical clustering methods are reported in the supplementary material (‘Supplementary EEG methods’, *I.*).

ERP statistical analysis

ERP for each channel, condition, trial was averaged into twenty 50ms bins from stimulus onset to 1s post-stimulus-onset. A database was created with average power at each of the 20 timebins, 3 Vibration and 5 EmotionalVoice conditions, for each EEG channel for the 4 clusters and electrode placement (left, center, right) as factors. As power followed a linear distribution, a Bayesian general LMM approach was used to validate the result of

topographical clustering. For each timebin, data was entered into a GLMM to test the Cluster*EmotionalVoice(A_{90}, A_{60}, A_{50}, A_{40}, A_{10})*Vibration(Anger, Fear, No vibrations) interaction. Random factors included participant identity and EEG channels.

Cluster significance calculations and source reconstruction methods are reported in the supplementary material ('Supplementary EEG methods', *II. & III.*).

Results

In this study, the participants had to categorize morphed voice emotions concomitant to induced throat vibrations in 2/3rd of the trials. We analyzed the influence of these vibrations on voice recognition responses, ERPs and source modulations in the brain.

Behavioral responses

The main effect of Vibration ($\chi^2(2)=8.25$, $p<.05$, Figure 2**A**), EmotionalVoice ($\chi^2(4)=17.81$, $p<.01$, Supplementary Figure 1) and their interaction ($\chi^2(8)=23.92$, $p< .01$, Figure 2**B**) were significant. Model effect size was small-to-medium (Fixed effects $R^2=.14$; Fixed and random effects $R^2=.36$).

The impact of Vibration on responses (Fear vibrations: *m*=.57, *sd*=.50, Anger vibrations: *m*=.53, *sd*=.50, No vibrations: *m*=.55, *sd*=.50) was tested using hypothesis-driven planned contrasts: [Vibration Anger>Fear], ($\chi^2(1)=6.86$, $p<.01$, *OR*=0.83); [Vibration Fear>None], ($\chi^2(1)=4.87$, $p<.05$, *OR*=1.12; [Vibration Anger>None], ($\chi^2(1)=1.68$, $p=.19$, *OR*=0.93; Figure 2**A**). These results already partly confirm our first hypothesis (more 'anger' response for an anger vibration, and similarly for fear).

Vibration*Emotional Voice interaction on responses revealed significant planned contrasts: [Vibration Anger>Fear]*[Ambiguous (A_{60}, A_{50}, A_{40}) voice stimuli], ($\chi^2(1)=7.99$, $p<.01$, *OR*=0.51, [Vibration Anger>None]*[Ambiguous (A_{60}, A_{50}, A_{40}) voice stimuli], ($\chi^2(1)=4.21$, $p<.05$, *OR*=0.67, [Vibration Fear>None]*[Ambiguous (A_{60}, A_{50}, A_{40}) voice stimuli], ($\chi^2(1)=4.20$, $p<.14$, *OR*=1.31; Figure 2**B**). More specific contrasts on the most ambiguous condition of Emotional Voice were computed: [Vibration Anger>Fear]*[Ambiguous (A_{50}) voice stimuli], ($\chi2(1)=3.94$, $p<.05$, *OR*=0.81), [Vibration Anger>None]*[Ambiguous (A_{50}) voice stimuli], ($\chi2(1)=7.08$, $p<.01$, *OR*=0.77), [Vibration Fear>None]*[Ambiguous (A_{50}) voice stimuli], ($\chi2(1)=0.28$, $p=.59$, *OR*=0.95; Figure 2**B**). These results also partly confirm our second hypothesis, especially the influence of angry

vibrations on ambiguous voice stimuli. Results for non-ambiguous voices are reported in Supplementary Table 8, with only an effect of incongruence for fear vibrations on A_{90} voices.

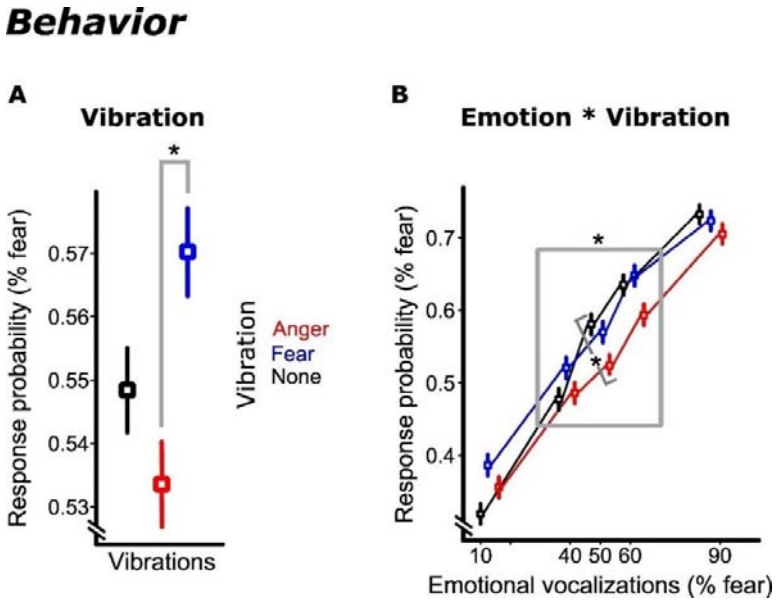

Figure 2. **A. Main effect of Vibration on the probability of fear responses.** On the X axis, each point represents a Vibration condition (i.e., None = no vibrations, Anger, Fear). **B.** Interaction between Emotional Voice and Vibration on the probability of fear responses. On the X axis, each point represents an Emotional Voice condition (i.e., A_{90}, A_{60}, A_{50}, A_{40}, A_{10}) and each coloured curve represents a Vibration condition (i.e., None = no vibrations; Anger, Fear). Error bars indicate the standard error of the mean. *p<.05.

EEG results

Event-Related Potentials

Influence of induced vibrations during vocal emotion processing on brain dynamics revealed differences between our clusters considering conditions response with a three-way Cluster*EmotionalVoice*Vibration interaction along trial length (Supplementary Table 6, $F(24)$=[3.04-23.36 min-max]; p<.001 FDR corrected). Three clusters contained at least one 50ms timebin presenting an EmotionalVoice*Vibration interaction. In the paragraph below, we describe the results from the functionally best-defined cluster, i.e., the frontocentral cluster (cluster C4, Figure 3**G** & 4**K**). Results from the two other ERP clusters (occipital and temporal) are described in the supplementary material, together with posthoc ERP results not pertaining to our hypotheses ('Supplementary EEG results', *I.* & *II*, and Supplementary Figures 2-4) and for source reconstruction (Supplementary Figures 4-5).

Effect of Emotional voice X Vibration: frontocentral cluster C4

The frontocentral cluster C4 (Figure 3**G** & 4**K**) presented an EmotionalVoice* Vibration interaction at 50-150ms, 300-400ms, 450-550ms, 600-650ms, 750-850ms, and 900-950ms time intervals (Supplementary Table 6; Figure 3 & 4). Three clearly defined peaks exceeded baseline threshold: a positive deflection (100-150ms), another positive deflection (200-350ms) and a negative one (from 500ms post-stimulus onset). For the sake of clarity, a complete list of all statistical contrasts is available in the supplementary material ('Supplementary EEG results', *II.)*.

Low emotional voice ambiguity ERPs

At 125ms, A_{90} voices with anger vibrations elicited a weaker positive deflection compared to both A_{50} and A_{10} voices. At the same time point, we observed that in A_{90} voices, anger vibration presented a weaker deflection compared to both fear and no-vibration conditions. No effects were observed for the P200 peak amplitude. Around 625ms, A_{10} voices showed a weaker negative deflection in the fear vibration condition compared to both A_{50} and A_{90} voices. Full results are detailed in the supplementary material (Figure S2, Supplementary Table 5).

High emotional voice ambiguity ERPs

For ambiguous voices, no expected results were observed around 100ms. In the descending phase of the P200, at 325ms, A_{40} voices differed from A_{50} voices (fear vibration). At 475ms, A_{60} voices differed from A_{50} voices (anger vibration). Around 625ms, A_{60} voices differed from A_{40} voices (anger vibration) and A_{40} voices showed a weaker negative deflection than both A_{50} and A_{60} voices (fear vibration). As for vibration contrasts, fear vibration elicited a greater negative deflection than both other conditions for A_{60} and A_{50} voices. At 825ms, all vibration conditions differed for A_{40} and A_{50} voices, while anger vibration led to a greater negative deflection than both other conditions for A_{60} voices. Finally, at 925ms, fear vibration showed a weaker negative deflection than both other conditions for A_{50} and A_{60} voices. Full results in Figure S2 and Supplementary Table 5. These results already support our EEG hypotheses, namely the presence of early (100-300ms) and later-stage (400-700ms) ERPs specific to the vibration conditions, both when these were mostly congruent or incongruent with the emotion of the voices presented auditorily.

Source reconstruction

Source reconstruction analysis revealed correlates at the brain level of the vibration ERP-dissociations observed at the sensor level (Figure 3**B**-**D**,**F** and 4**B**-**E**,**G**-**J**).

Low emotional voice ambiguity brain sources

We observed in the 300-350ms time interval a fear vs anger vibrations source-ERP contrast in the superior posterior lobe (Figure 3**B**,**C**) but also for incongruent fear vs no-vibrations in the medial orbitofrontal cortex (OFC), insula and primary somatosensory cortex (550-600ms, Figure S4) followed by the SMA (750-800ms) and primary motor cortex (900-950ms; Figure 3**D**). For fearful voices, congruent fear vs no-vibration contrasts were observed in the SMA (700-750ms) and superior parietal lobe (700-750ms; 850-900ms; Figure 3**F**).

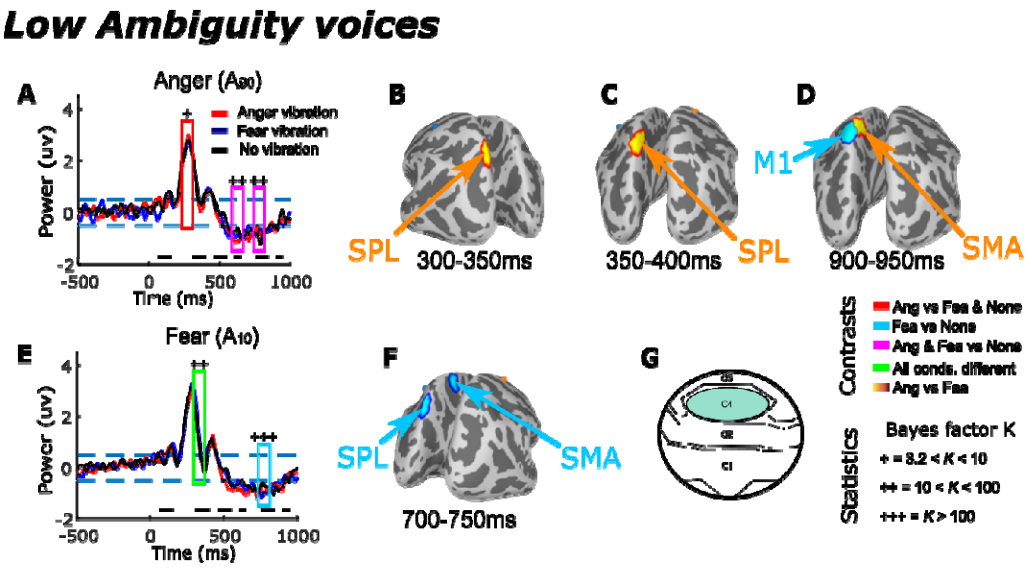

Figure 3. **Contrasts between vibrations for Low Ambiguity vocalizations.** Top panel: Contrasts between vibrations for low ambiguity angry voices (A_{90}) for **A:** Event-related potentials and **B**-**C**-**D**: Brain activations from source reconstruction. Bottom panel: Contrasts between vibrations for low ambiguity fearful voices (F_{90}) for **E**: Event-related potentials and **F**: Brain activations from source reconstruction. **G**: Graphical representation of the topographical location of the analyzed cluster. For ERPs graphs, line colors for each condition represent the vibration applied (anger in red, fear in blue, no-vibration control in black). Upper horizontal lines represent contrast with substantial proof of difference (BF>3, see Methods) between conditions. Black lower horizontal line represents the period where the two-way Condition X Vibration interaction was important. Blue dashed horizontal line represents the 3 X baseline standard deviation interval. All statistics reported here have Bayes factors > 3. M1: primary motor cortex, SMA: supplementary motor area, SPL: superior parietal lobule.

High emotional voice ambiguity brain sources

At the 350-400ms peak, source-ERP fear vs no vibration contrasts were observed in the anterior cingulate cortex and insula for fearful voices (Figure 4**G**-**H**). A significant anger vs fear vibration contrast was also observed at the same timing in the superior parietal cortex for fearful voices (Figure 4**G**). At the 450-500ms peak, a congruent (anger vibration vs. no-vibration) contrast was observed for angry voices in the primary sensory, primary motor, and premotor areas, OFC, dorsolateral prefrontal cortex (DLPFC) as well as in the intraparietal sulcus (IPS; Figure 4**B**). In the late period, a congruent anger vibration vs no-vibration contrast was observed for angry voices in the precuneus (700-750ms, Figure 4**D**), as well as superior parietal lobule and IPS (900-950ms, Figure 4**E**). A congruent fear vibration vs no-vibration contrast was also observed for fearful voices in the primary motor cortex, IFG and OFC (650-700ms) while a fear and anger (global) vibration vs no-vibration contrast was observed in the primary motor area at the same timing (Figure 4**I**). Finally, an incongruent anger vibration vs no-vibration contrast was observed for fearful voices in primary motor cortex and SMA (750-800ms; Figure 4**J**).

Source reconstruction results therefore also fit well with our EEG hypotheses both for vibration-specific locations and for error-prediction sites. Taken together, ERP and source reconstruction results mostly confirm our hypotheses and paint a clearer picture of the dynamic neural underpinnings of induced throat vibrations on the recognition of both emotionally ‘pure’ and ambiguous voices.

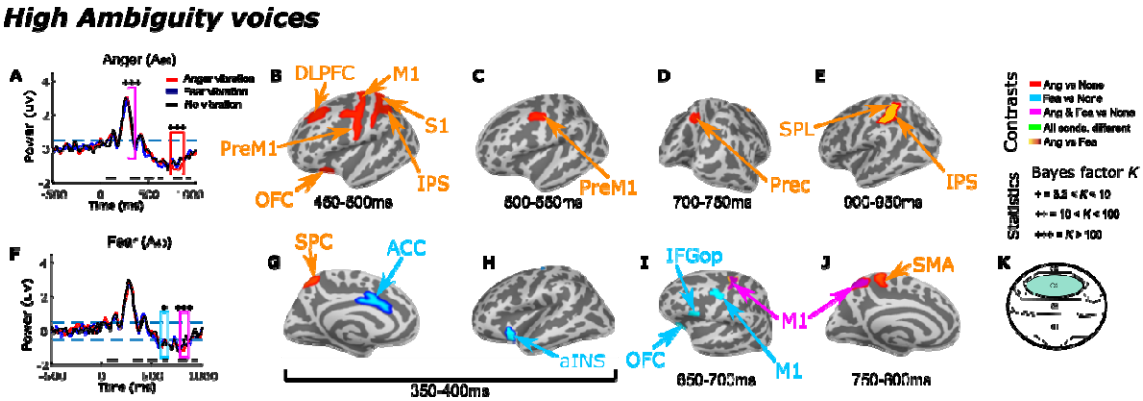

Figure 4. **Contrasts between vibrations for High Ambiguity vocalizations.** Top panel: Contrasts between vibrations for high ambiguity angry voices (A_{10}) for **A**: Event-related potentials and **B**-**C**-**D**-**E**: Brain activations from source reconstruction. Bottom panel: Contrasts between vibrations for high ambiguity fearful voices (F_{10}) for **F:** Event-related potentials and **G**-**H**: Brain activations from source reconstruction. **K:** Graphical representation

of the topographical location of the analyzed cluster. For ERPs graphs, line colors for each condition represent the vibration applied (anger in red, fear in blue, no-vibration control in black). Upper horizontal lines represent contrast with substantial proof of difference (BF>3, see Methods) between conditions. Black lower horizontal line represents the period where the two-way Condition X Vibration interaction was significant. Blue dashed horizontal line represents the 3 X baseline standard deviation interval. All statistics reported here have Bayes factors > 3. ACC: anterior cingular cortex, DLPFC: dorsolateral prefrontal cortex, IFG: inferior frontal gyrus, IPS: intraparietal sulcus, M1: primary motor cortex, OFC: orbitofrontal cortex, op: pars opercularis, Prec: precuneus, PreM1: premotor cortex, S1: primary somatosensory cortex, SMA: supplementary motor area, SPL: superior parietal lobule.

Discussion

Research on the impact of embodied cognition on vocal emotion processing is very scarce. Our study investigated whether vocal tract vibrations could represent useable vibrotactile feedback, and whether this feedback could influence auditory emotion perception, especially for emotionally ambiguous voices. Behavioural and EEG results revealed effects that support our assumptions: induced throat vibrations modulated behavioral responses to emotional voices as well as early and late ERP components, within relevant and expected brain modulators.

As mentioned, our study highlighted an impact of induced vibrations on the behavioural responses to morphed emotional voices, which is in line with our main hypothesis. Although not all the expected contrasts were observed, it was clear that our participants categorized emotional vocalizations as ‘anger’ significantly more often when a concomitant anger vibration was induced than when it was a fear vibration. This result is consistent with language comprehension facilitation effect of by the transcranial stimulation of the primary motor areas controlling the lips/tongue (D'Ausilio et al., 2009). We therefore interpret the results of our study as an emotion-specific stimulation of throat mechanoreceptors that would facilitate vocal emotion recognition. This is also coherent with studies on congruence-specific facilitation effect in multimodal processing of emotion (Lin & Ding, 2019; Niedenthal et al., 2001). Following a similar reasoning, participants categorized ambiguous emotional vocalizations (A_{60}, $A_{50,}$ and A_{40}) as ‘anger’ more often when anger (vs. fear) vibrations were induced—or when there were no vibrations were induced. This latter contrast is consistent with the SIMS model (Niedenthal et al., 2010): interoceptive feedback

would be integrated especially for non-prototypical emotional stimuli—here, ambiguous voices, when available emotional information is not clear enough.

At the neural level, the frontocentral cluster C4 was the locus of an interaction between voice ambiguity and induced throat vibrations—both for early and late time-windows. Our results showed that the induction of both congruent fear and anger vibrations led to an increase in P100 amplitude—for low, high and highest (A_{50}) emotional ambiguity. In the literature, the P100 was linked to attention-driven interoceptive accuracy effects—auditory tone appearing in synchrony with heartbeat or with a delay (Mendoza-Medialdea & Ruiz-Padial, 2021). In cross-modal studies, multimodal integration of emotion recognition occurred very early, around 110ms post-stimulus onset (Pourtois et al., 2000). Furthermore, studies investigating interoception and somatosensory stimulation observed P300 modulations—slightly later that in our data, particularly in the parietal area and therefore in line with our results (Al et al., 2020; Auksztulewicz & Blankenburg, 2013; Has et al., 2021; Herbert et al., 2007; Pollatos et al., 2005).

We also observed the presence of a P200 component in this frontocentral area, consistent with emotional voice processing (Duville et al., 2022; Gädeke et al., 2013; Maltezou-Papastylianou et al., 2022; Paulmann et al., 2013; Paulmann & Uskul, 2017). Our results showed anger vibrations lead to P200 amplitude *increase* (low-ambiguity), while fear vibrations lead to P200 *decrease* (ambiguous voices), and for incongruent high-ambiguity angry voices. These results are in line with above-mentioned findings on multimodal integration and interoception and occur at a slightly later stage. These data therefore support the impact of vibrotactile throat feedback on behavior through P100 and P200 modulations. They also align with interoception and somatosensory stimulation instigating a P300 component—especially in the parietal region (Al et al., 2020; Auksztulewicz & Blankenburg, 2013; Has et al., 2021; Herbert et al., 2007; Pollatos et al., 2005). Moreover, the effect of fear vibrations for ambiguous voice categorization only is also supported by the SIMS model (Niedenthal, 2007; Niedenthal et al., 2010).

Concerning later ERP components, we observed modulations by induced vibrations most notably between 500 and 800ms post-stimulus onset. These modulations involve both congruent and incongruent cases, especially for low-ambiguity voices—and both emotions. For instance, at 500ms, anger vibrations produce an incongruent negative deflection for low-ambiguity fearful voices. These results recall the N400 component observed in frontal, parietal and frontocentral areas during emotional voice processing (Föcker & Röder, 2019;

Maltezou-Papastylianou et al., 2022; Tarai & Srinivasan, 2019). At 800ms, power increase—although still negative—for both emotional vibrations compared to no-vibration for ambiguous voices compares to the Late Positive Component (LPC). The LPC is usually associated with higher-order cognitive processes: arousal information (Paulmann et al., 2013), prosody expectancy violation (Astésano et al., 2004; Brattico et al., 2006; Paulmann et al., 2012), emotional 'meaning' (Wei et al., 2022; Zora et al., 2020). Nevertheless, this echoes once again with the SIMS model—embodied simulation is engaged when the emotional stimulus is non-prototypical (Niedenthal et al., 2010). Here, the vibrations were integrated specifically for ambiguous voices, as prototypical emotional voice processing would not be enough to identify the emotion. At the interoception and proprioception levels, these results globally show that vibrations are integrated during emotional processing at multiple stages, dynamically and throughout the time course of an event, similarly to the multimodal integration shown in audiovisual studies (Föcker & Röder, 2019; Klasen et al., 2012).

ERP source reconstruction revealed source modulations in several key regions. Some are involved in vibratory processing—the DLPFC, vibrotactile working memory (Burton et al., 2010) or in vibratory processing, crucially in vibration rate or pattern discrimination (Christensen et al., 2007; Coghill et al., 1994; Golaszewski et al., 2006; Hernández et al., 2002; Lenoir et al., 2017; Li Hegner et al., 2010; Macar et al., 2002; Romo et al., 2004; Woolgar & Zopf, 2017; Wu et al., 2018). Concerning more internal feedback or interoception, we observed source modulations in the insula, known to be crucially involved in interoceptive processing (Adolfi et al., 2017; Casals-Gutierrez & Abbey, 2020) and precuneus (interoceptive awareness; Herbert et al., 2007; Longarzo et al., 2021; Wilson-Mendenhall et al., 2019). The modulation of the primary somatosensory cortex might be related to proprioceptive information processing while modulations in the OFC were also reported in interoception studies (Singer & Klimecki, 2014; Stern et al., 2017).

Regarding emotional prosody perception, we observed source modulations in the DLPFC (prosodic working memory; Mitchell, 2007), superior parietal lobule (emotional prosody judgement; Bach et al., 2008), as well as in other regions also involved in emotional prosody processing (IPS, SMA; Alba-Ferrara et al., 2011; Belyk & Brown, 2014; Kotz et al., 2003). OFC modulations fit well with higher-level processing of emotional prosody (Wildgruber et al., 2006)—especially when ambiguous (Grandjean, 2021), or for accurately categorizing human voice (Ceravolo et al., 2023). Source modulations in the insula may contribute to external emotional stimuli integration (vibrotactile feedback) and emotional

experiences with interoceptive signals (Adolfi et al., 2017; Nguyen et al., 2016; Zaki et al., 2012). Finally, above-mentioned modulations—superior parietal lobule, IPS, premotor, primary somatosensory cortex—are in line with multimodal integration for auditory, visual, sensorimotor and tactile stimuli (Baker et al., 2018; Culham & Valyear, 2006; de Borst & de Gelder, 2017; Gentile et al., 2011; Kreifelts et al., 2007; Parker et al., 2020). Source modulation locations for the Vibration and Emotion factors are fully in line with our previous functional MRI results on neural correlates of emitted, measured vocal tract vibrations during emotional expression (Selosse et al., 2025). Combined with the present data, they help draw a clearer picture of emotion embodiment in voice signals.

Taken together, these results suggest the implication of a dynamic, wide-spread network of brain areas in the processing of throat induced vibrations and their integration with emotional voices. This is especially true for highly ambiguous voice signals, for which such information is most relevant for improved decision making. This interpretation relies on previous literature on interoception and multisensory integration, and suggests that induced throat vibrations are a type of feedback critical enough to impact auditory perception and decision.

Limitations and perspectives

Our sample consists mainly of female participants, despite gender differences in emotional prosody recognition (Imaizumi et al., 2004; Lambrecht et al., 2014) and facial mimicry (Korb et al., 2015; Niedenthal et al., 2012). Second, it would be necessary to replicate the observed effects we with different emotions to generalize our results, although morphing positive and negative emotions may be problematic. Third, induced vibrations should be those of the participants for improved ecological validity—and better consideration of interindividual differences. Fourth, vibrations were induced on the skin, therefore more subtle internal tissue vibrations encoded by different mechano-receptors may not be triggered. Further research is needed to clarify the relationships between actual vibrations created by the vocal cords and those induced on the skin level. Future studies would benefit from the inclusion of measures of bodily self-consciousness or interoceptive awareness. In fact, interindividual variability in this dimension probably impacts subsequent data as the feedback mechanism highlighted in the present work might highly depend on such sensibility.

Conflict of interest

The authors declare no conflict of interest whatsoever.

Data availability statement

Data and codes will be made available upon request to the corresponding author.

Funding

This work as well as Open Access publication charges were funded by the Swiss National Science Foundation (grant number 10531C_189397/DG-LC).

References

Adolfi, F., Couto, B., Richter, F., Decety, J., Lopez, J., Sigman, M., & Ibáñez, A. (2017). Convergence of interoception, emotion, and social cognition: a twofold fMRI meta-analysis and lesion approach. *Cortex*, *88*, 124--142.

Al, E., Iliopoulos, F., Forschack, N., Nierhaus, T., Grund, M., Motyka, P., & Villringer, A. (2020). Heart–brain interactions shape somatosensory perception and evoked potentials. *Proceedings of the National Academy of Sciences*, *117*(19), 10575--10584.

Alba-Ferrara, L., Hausmann, M., Mitchell, R. L., & Weis, S. (2011). The Neural Correlates of Emotional Prosody Comprehension: Disentangling Simple from Complex Emotion. *PLoS ONE*, *6*(12), 28701. https://doi.org/10.1371/journal.pone.0028701

Ashley, V., Vuilleumier, P., & Swick, D. (2004). Time course and specificity of event-related potentials to emotional expressions. *Neuroreport*, *15*(1), 211--216.

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, *18*(2), 172--184.

Auksztulewicz, R., & Blankenburg, F. (2013). Subjective rating of weak tactile stimuli is parametrically encoded in event-related potentials. *Journal of Neuroscience*, *33*(29), 11878--11887.

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage*, *42*(2), 919--927. https://doi.org/10.1016/j.neuroimage.2008.05.034

Baker, C. M., Burks, J. D., Briggs, R. G., Conner, A. K., Glenn, C. A., Taylor, K. N., Sali, G., McCoy, T. M., Battiste, J. D., O'Donoghue, D. L., & Sughrue, M. E. (2018). A Connectomic Atlas of the Human Cerebrum—Chapter 7: The Lateral Parietal Lobe. *Operative Neurosurgery*, *15*(suppl_1), 295--349. https://doi.org/10.1093/ons/opy261

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *convergence*, *12*(1), 2.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior research methods*, *40*(2), 531--539.

Belyk, M., & Brown, S. (2014). Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, *9*(9), 1395--1403. https://doi.org/10.1093/scan/nst124

Blairy, S., Herrera, P., & Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *Journal of Nonverbal Behavior*, *23*, 5--41.

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ*, *310*(6973), 170.

Borgomaneri, S., Bolloni, C., Sessa, P., & Avenanti, A. (2020). Blocking facial mimicry affects recognition of facial and body expressions. *PLoS ONE*, *15*(2), e0229364.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433--436. https://doi.org/10.1163/156856897x00357

Brattico, E., Tervaniemi, M., Näätänen, R., & Peretz, I. (2006). Musical scale properties are automatically processed in the human auditory cortex. *Brain Research*, *1117*(1), 162--174.

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2008). Behold the voice of wrath: Cross-modal modulation of visual attention by anger prosody. *Cognition*, *106*(3), 1497--1503.

Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008). Beyond fear: Rapid spatial orienting toward positive emotional stimuli. *Psychological science*, *19*(4), 362--370.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Burleson, M. H., & Quigley, K. S. (2021). Social interoception and social allostasis through touch: legacy of the somatovisceral afference model of emotion. *Social neuroscience*, *16*(1), 92--102.

Burton, H., Sinclair, R. J., & Dixit, S. (2010). Working memory for vibrotactile frequencies: Comparison of cortical activity in blind and sighted individuals. *Human Brain Mapping*, *31*(11), 1686--1701. https://doi.org/10.1002/hbm.20966

Carminati, M., Fiori-Duharcourt, N., & Isel, F. (2018). Neurophysiological differentiation between preattentive and attentive processing of emotional expressions on French vowels. *Biological psychology*, *132*, 55--63.

Casals-Gutierrez, S., & Abbey, H. (2020). Interoception, mindfulness and touch: A meta-review of functional MRI studies. *International Journal of Osteopathic Medicine*, *35*, 22--33.

Ceravolo, L., Debracque, C., Pool, E., Gruber, T., & Grandjean, D. (2023). Frontal mechanisms underlying primate calls recognition by humans. *Cerebral Cortex Communications*, *4*(4), tgad019.

Ceravolo, L., Frühholz, S., Pierce, J., Grandjean, D., & Péron, J. (2021). Basal ganglia and cerebellum contributions to vocal emotion processing as revealed by high-resolution fMRI. *Scientific reports*, *11*(1), 10645.

Ceravolo, L., Moisa, M., Grandjean, D., Ruff, C., & Frühholz, S. (2026). Behavioral and neural aftermath of right inferior frontal cortex disruption on ambiguous vocal emotion decisional processes. *Scientific reports*. https://doi.org/10.1038/s41598-026-39668-0

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893.

Christensen, M. S., Lundbye-Jensen, J., Geertsen, S. S., Petersen, T. H., Paulson, O. B., & Nielsen, J. B. (2007). Premotor cortex modulates somatosensory cortex during voluntary movements without proprioceptive feedback. *Nature Neuroscience*, *10*(4), 417--419. https://doi.org/10.1038/nn1873

Coghill, R., Talbot, J., Evans, A., Meyer, E., Gjedde, A., Bushnell, M., & Duncan, G. (1994). Distributed processing of pain and vibration by the human brain. *The Journal of Neuroscience*, *14*(7), 4095--4108. https://doi.org/10.1523/jneurosci.14-07-04095.1994

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale.

Couto, B., Adolfi, F., Velasquez, M., Mesow, M., Feinstein, J., Canales-Johnson, A., & Ibanez, A. (2015). Heart evoked potential triggers brain responses to natural affective scenes: a preliminary study. *Autonomic Neuroscience*, *193*, 132--137.

Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current opinion in psychology*, *17*, 7--14.

Culham, J. C., & Valyear, K. F. (2006). Human parietal cortex in action. *Current Opinion in Neurobiology*, *16*(2), 205--212. https://doi.org/10.1016/j.conb.2006.03.005

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*(5), 381--385.

de Borst, A. W., & de Gelder, B. (2017). fMRI-based multivariate pattern analyses reveal imagery modality and imagery content specific representations in primary somatosensory, motor and auditory cortices. *Cerebral Cortex*, *27*(8), 3994--4009. https://doi.org/10.1093/cercor/bhw211

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology*, *19*(6), 643--647.

Drimalla, H., Landwehr, N., Hess, U., & Dziobek, I. (2019). From face to face: the contribution of facial mimicry to cognitive and emotional empathy. *Cognition and Emotion*.

Duville, M. M., Alonso-Valerdi, L. M., & Ibarra-Zarate, D. I. (2022). Neuronal and behavioral affective perceptions of human and naturalness-reduced emotional prosodies. *Frontiers in Computational Neuroscience*, *16*, 1022787.

Engelen, T., Solcà, M., & Tallon-Baudry, C. (2023). Interoceptive rhythms in the brain. *Nature Neuroscience*, *26*(10), 1670--1684.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*(2), 399--402.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175--191.

Föcker, J., & Röder, B. (2019). Event-related potentials reveal evidence for late integration of emotional prosody and facial expression in dynamic stimuli: An ERP study. *Multisensory Research*, *32*(6), 473--497.

Gädeke, J. C., Föcker, J., & Röder, B. (2013). Is the processing of affective prosody influenced by spatial attention? An ERP study. *BMC neuroscience*, *14*(1), 1--15.

Gentile, G., Petkova, V. I., & Ehrsson, H. H. (2011). Integration of Visual and Tactile Signals From the Hand in the Human Brain: An fMRI Study. *Journal of Neurophysiology*, *105*(2), 910--922. https://doi.org/10.1152/jn.00840.2010

Gibson, J. (2019). Mindfulness, interoception, and the body: A contemporary perspective. *Frontiers in Psychology*, *10*.

Golaszewski, S. M., Siedentopf, C. M., Koppelstaetter, F., Fend, M., Ischebeck, A., Gonzalez-Felipe, V., Haala, I., Struhal, W., Mottaghy, F. M., Gallasch, E., Felber, S. R., & Gerstenbrand, F. (2006). Human brain structures related to plantar vibrotactile stimulation: A functional magnetic resonance imaging study. *NeuroImage*, *29*(3), 923--929. https://doi.org/10.1016/j.neuroimage.2005.08.052

Grandjean, D. (2021). Brain networks of emotional prosody processing. *Emotion Review*, *13*(1), 34--43.

Grandjean, D., Bänziger, T., & Scherer, K. R. (2006). Intonation as an interface between language and affect. *Progress in brain research*, *156*, 235--247.

Has, N. A., Jamaludin, M. N., Rejab, S. M., Ahmad, Z., Zainuddin, N. F., Salam, A. H. A., & As'ari, M. A. (2021). P300 Somatosensory Validation of Vibrotactile Haptic Feedback for Upper Limb Prosthesis. 2021 IEEE National Biomedical Engineering Conference (NBEC),

Hawk, S. T., Fischer, A. H., & van Kleef, G. A. (2012). Face the noise: embodied responses to nonverbal vocalizations of discrete emotions. *Journal of personality and social psychology*, *102*(4), 796.

Herbert, B. M., Pollatos, O., & Schandry, R. (2007). Interoceptive sensitivity and emotion processing: an EEG study. *International Journal of Psychophysiology*, *65*(3), 214--227.

Hernández, A., Zainos, A., & Romo, R. (2002). Temporal Evolution of a Decision-Making Process in Medial Premotor Cortex. *Neuron*, *33*(6), 959--972. https://doi.org/10.1016/s0896-6273(02)00613-x

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology*, *40*(2), 129--141.

Hess, U., & Fischer, A. (2014). Emotional mimicry: Why and when we mimic emotions. *Social and personality psychology compass*, *8*(2), 45--57.

Holland, A. C., O'Connell, G., & Dziobek, I. (2021). Facial mimicry, empathy, and emotion recognition: a meta-analysis of correlations. *Cognition and Emotion*, *35*(1), 150--168.

Hyniewska, S., & Sato, W. (2015). Facial feedback affects valence judgments of dynamic and static emotional expressions. *Frontiers in Psychology*, *6*, 291.

Imaizumi, S., Homma, M., Ozawa, Y., Maruishi, M., & Muranaka, H. (2004). Gender differences in emotional prosody processing—An fMRI study. *Psychologia*, *47*(2), 113--124.

Kawahara, H., & Matsui, H. (2003). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing,

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., & Zucker, N. (2018). Interoception and mental health: a roadmap. *Biological psychiatry: cognitive neuroscience and neuroimaging*, *3*(6), 501--513.

Klasen, M., Chen, Y. H., & Mathiak, K. (2012). Multisensory emotions: perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*(4), 381--392.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3. In.

Korb, S., Grandjean, D., & Scherer, K. R. (2010). Timing and voluntary suppression of facial mimicry to smiling faces in a Go/NoGo task—An EMG study. *Biological psychology*, *85*(2), 347--349.

Korb, S., Malsert, J., Rochas, V., Rihs, T. A., Rieger, S. W., Schwab, S., & Grandjean, D. (2015). Gender differences in the neural network of facial mimicry of smiles–An rTMS study. *Cortex*, *70*, 101--114.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, *86*(3), 366--376. https://doi.org/10.1016/s0093-934x(02)00532-1

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *NeuroImage*, *37*(4), 1445--1456. https://doi.org/10.1016/j.neuroimage.2007.06.020

Künecke, J., Wilhelm, O., & Sommer, W. (2017). Emotion recognition in nonverbal face-to-face communication. *Journal of Nonverbal Behavior*, *41*, 221--238.

Lambrecht, L., Kreifelts, B., & Wildgruber, D. (2014). Gender differences in emotion recognition: Impact of sensory modality and emotional category. *Cognition & Emotion*, *28*(3), 452--469.

Laukka, P., Juslin, P., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion*, *19*(5), 633--653.

Lenoir, C., Huang, G., Vandermeeren, Y., Hatem, S. M., & Mouraux, A. (2017). Human primary somatosensory cortex is differentially involved in vibrotaction and nociception. *Journal of Neurophysiology*, *118*(1), 317--330. https://doi.org/10.1152/jn.00615.2016

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package ‘emmeans’. *R package version*, *1*(3.2).

Li Hegner, Y., Lee, Y., Grodd, W., & Braun, C. (2010). Comparing Tactile Pattern and Vibrotactile Frequency Discrimination: A Human fMRI Study. *Journal of Neurophysiology*, *103*(6), 3115--3122. https://doi.org/10.1152/jn.00940.2009

Lin, Y., & Ding, H. (2019). Multisensory integration of emotions in a face-prosody-semantics Stroop task. In *NeuroManagement and Intelligent Computing Method on Multimodal Interaction* (pp. 1--5).

Lindín, M., Zurrón, M., & Díaz, F. (2004). Changes in P300 amplitude during an active standard auditory oddball task. *Biological psychology*, *66*(2), 153--167.

Longarzo, M., Mele, G., Alfano, V., Salvatore, M., & Cavaliere, C. (2021). Gender Brain Structural Differences and Interoception. *Frontiers in Neuroscience*, *14*, 586860. https://doi.org/10.3389/fnins.2020.586860

Lylykangas, J., Surakka, V., Rantala, J., Raisamo, J., Raisamo, R., & Tuulari, E. (2009). Vibrotactile information for intuitive speed regulation. In *People and Computers XXIII Celebrating People and Technology* (pp. 112--119).

Macar, F., Lejeune, H., Bonnet, M., Ferrara, A., Pouthas, V., Vidal, F., & Maquet, P. (2002). Activation of the supplementary motor area and of attentional networks during temporal processing. *Experimental Brain Research*, *142*(4), 475--485.

Maltezou-Papastylianou, C., Russo, R., Wallace, D., Harmsworth, C., & Paulmann, S. (2022). Different stages of emotional prosody processing in healthy ageing–evidence from behavioural responses, ERPs, tDCS, and tRNS. *PLoS ONE*, *17*(7), e0270934.

Mendoza-Medialdea, M. T., & Ruiz-Padial, E. (2021). Understanding the capture of exogenous attention by disgusting and fearful stimuli: the role of interoceptive accuracy. *International Journal of Psychophysiology*, *161*, 53--63.

Mitchell, R. L. C. (2007). fMRI delineation of working memory for emotional prosody in the brain: Commonalities with the lexico-semantic emotion network. *NeuroImage*, *36*(3), 1015--1025. https://doi.org/10.1016/j.neuroimage.2007.03.016

Moody, E. J., & McIntosh, D. N. (2011). Mimicry of dynamic emotional and motor-only stimuli. *Social Psychological and Personality Science*, *2*(6), 679--686.

Moody, E. J., McIntosh, D. N., Mann, L. J., & Weisser, K. R. (2007). More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion*, *7*(2), 447.

Munger, J. B., & Thomson, S. L. (2008). Frequency response of the skin on the head and neck during production of selected speech sounds. *The Journal of the Acoustical Society of America*, *124*(6), 4001--4012.

Nguyen, V. T., Breakspear, M., Hu, X., & Guo, C. C. (2016). The integration of the internal and external milieu in the insula during dynamic emotional experiences. *NeuroImage*, *124*, 455--463. https://doi.org/10.1016/j.neuroimage.2015.08.078

Niedenthal, P. M. (2007). Embodying Emotion. *Science*, *316*(5827), 1002--1005.

Niedenthal, P. M., Augustinova, M., Rychlowska, M., Droit-Volet, S., Zinner, L., Knafo, A., & Brauer, M. (2012). Negative relations between pacifier use and emotional competence. *Basic and Applied Social Psychology*, *34*(5), 387--394.

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, Å. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition & Emotion*, *15*(6), 853--864.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) model: Embodied simulation and the meaning of facial expression. *Behavioral and brain sciences*, *33*(6), 417.

Nolan, M., Madden, B., & Burke, E. (2009). Accelerometer based measurement for the mapping of neck surface vibrations during vocalized speech. Annual International Conference of the IEEE Engineering in Medicine and Biology Society,

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*(1), 156869.

Paquette, S., Rigoulot, S., Grunewald, K., & Lehmann, A. (2020). Temporal decoding of vocal and musical emotions: Same code, different timecourse? *Brain Research*, 146887.

Parker, A. G., Briggs, R. G., Conner, A. K., O'Neal, C. M., Bonney, P. A., Maxwell, B. D., Baker, C. M., Burks, J. D., Sali, G., & Sughrue, M. E. (2020). Parcellation-based tractographic modeling of the ventral attention network. *Journal of the Neurological Sciences*, *408*, 116548.

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, 345.

Paulmann, S., Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, *50*(7), 1609--1620.

Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo-and lexical-sentence context. *Brain and Language*, *105*(1), 59--69.

Paulmann, S., & Uskul, A. K. (2017). Early and late brain signatures of emotional prosody among individuals with high versus low power. *Psychophysiology*, *54*(4), 555--565.

Pell, M. D., & Kotz, S. A. (2021). Comment: The next frontier: Prosody research gets interpersonal. *Emotion Review*, *13*(1), 51--56.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437--442.

Peron, J., Frühholz, S., Ceravolo, L., & Grandjean, D. (2016). Structural and functional connectivity of the subthalamic nucleus during vocal emotion decoding. *Social Cognitive and Affective Neuroscience*, *11*(2), 349-356.

Philip, L., Martin, J. C., & Clavel, C. (2018). Rapid facial reactions in response to facial expressions of emotion displayed by real versus virtual faces. *i-Perception*, *9*(4), 2041669518786527.

Pollatos, O., Kirsch, W., & Schandry, R. (2005). On the relationship between interoceptive awareness, emotional experience, and brain processes. *Cognitive Brain Research*, *25*(3), 948--962.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, *11*(6), 1329--1333.

Proverbio, A. M., Benedetto, F., & Guazzone, M. (2020). Shared neural mechanisms for processing emotions in music and vocalizations. *European Journal of Neuroscience*, *51*(9), 1987--2007.

Risberg, A., & Lubker, J. (1978). Prosody and speechreading. *Speech Transmission Laboratory Quarterly Progress Report and Status Report*, *4*, 1--16.

Romo, R., Hernández, A., & Zainos, A. (2004). Neuronal Correlates of a Perceptual Decision in Ventral Premotor Cortex. *Neuron*, *41*(1), 165--173. https://doi.org/10.1016/s0896-6273(03)00817-1

Salamone, P. C., Legaz, A., Sedeño, L., Moguilner, S., Fraile-Vazquez, M., Campo, C. G., & Ibañez, A. (2021). Interoception primes emotional processing: multimodal evidence from neurodegeneration. *Journal of Neuroscience*, *41*(19), 4276--4292.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech communication*, *40*(1-2), 227--256.

Scherer, K. R., & Oshinsky, J. S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion*, *1*(4), 331--346.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in cognitive sciences*, *10*(1), 24--30.

Schneider, K. G., Hempel, R. J., & Lynch, T. R. (2013). That "poker face" just might lose you the game! The impact of expressive suppression and mimicry on sensitivity to facial expressions of emotion. *Emotion*, *13*(5), 852.

Selosse, G., Grandjean, D., & Ceravolo, L. (2025). Neural correlates of embodied and vibratory mechanisms associated with emotional prosody production. *Social Cognitive and Affective Neuroscience*, *20*(1), nsaf084.

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current Biology*, *24*(18), 875--878.

Stel, M., & van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological science*, *19*(10), 984.

Stern, E. R., Grimaldi, S. J., Muratore, A., Murrough, J., Leibu, E., Fleysher, L., & Burdick, K. E. (2017). Neural correlates of interoception: Effects of interoceptive focus and relationship to dimensional measures of body awareness. *Human Brain Mapping*, *38*(12), 6068--6082.

Sundberg, J. (1992). Phonatory vibrations in singers: A critical review. *Music Perception: An Interdisciplinary Journal*, *9*(3), 361--381.

Švec, J. G., Titze, I. R., & Popolo, P. S. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *The Journal of the Acoustical Society of America*, *117*(3), 1386--1394.

Tarai, S., & Srinivasan, N. (2019). Emotional prosody Stroop effect in Hindi: an event related potential study. In *Progress in brain research* (Vol. 247, pp. 193--217).

Thomas, S. J., Johnstone, S. J., & Gonsalvez, C. J. (2007). Event-related potentials during an emotional Stroop task. *International Journal of Psychophysiology*, *63*(3), 221--231.

Tsolaki, A., Kosmidou, V., Hadjileontiadis, L., Kompatsiaris, I. Y., & Tsolaki, M. (2015). Brain source localization of MMN, P300 and N400: Aging and gender differences. *Brain Research*, 32--49.

Tuuri, K., Eerola, T., & Pirhonen, A. (2010). Leaping across modalities: Speed regulation messages in audio and tactile domains. International Workshop on Haptic and Audio Interaction Design,

Vachon, F., Hughes, R. W., & Jones, D. M. (2012). Broken expectations: violation of expectancies, not novelty, captures auditory attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 164.

Venables, W. N., & Smith, D. M. (2003). The R development core team. *An Introduction to R, Version*, *1*(0).

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989--994.

Wei, H., He, Y., Kauschke, C., Scharinger, M., & Domahs, U. (2022). An EEG-study on L2 categorization of emotional prosody in German. Speech Prosody,

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. In *Progress in brain research* (Vol. 156, pp. 249--268). https://doi.org/10.1016/s0079-6123(06)56013-3

Wilson-Mendenhall, C. D., Henriques, A., Barsalou, L. W., & Barrett, L. F. (2019). Primary Interoceptive Cortex Activity during Simulated Experiences of the Body. *Journal of Cognitive Neuroscience*, *31*(2), 221--235. https://doi.org/10.1162/jocn_a_01346

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701--702.

Witteman, J., van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, *50*(12), 2752--2763.

Woolgar, A., & Zopf, R. (2017). Multisensory coding in the multiple-demand regions: Vibrotactile task information is coded in frontoparietal cortex. *Journal of Neurophysiology*, *118*(2), 703--716. https://doi.org/10.1152/jn.00559.2016

Wu, Y., Uluç, I., Schmidt, T. T., Tertel, K., Kirilina, E., & Blankenburg, F. (2018). Overlapping frontoparietal networks for tactile and visual parametric working memory representations. *NeuroImage*, *166*, 325--334.

Zaki, J., Davis, J. I., & Ochsner, K. N. (2012). Overlapping activity in anterior insula during interoception and emotional experience. *NeuroImage*, *62*(1), 493--499.

Zora, H., Rudner, M., & Magnusson, A. K. M. (2020). Concurrent affective and linguistic prosody with the same emotional valence elicits a late positive ERP response. *European Journal of Neuroscience*, *51*(11), 2236--2249.