Neural dynamics of induced vocal tract vibrations during vocal emotion recognition

Garance Selosse MSc (a, b)*

Damien Benis, PhD (a,b)*

Jennifer Greilsamer, MSc (c)

Ben Meuleman, PhD (b)

Didier Grandjean, PhD† (a, b)

Leonardo Ceravolo, PhD† (a, b)

(a) *Neuroscience of Emotion and Affective Dynamics Lab, Department of Psychology, University of Geneva, Geneva, Switzerland,*

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

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

Corresponding author:
Garance Selosse
Garance.Selosse@unige.ch

* Shared first authors

† Shared last authors

Abstract

Emotional prosody is defined as suprasegmental and segmental changes in voice and related acoustic parameters that can inform the listener about the emotional state of the speaker. Despite a large corpus of literature in psychological and brain mechanisms in emotional prosody perception, the perspective of embodied cognition in these mechanisms have been largely neglected. Here we investigated the influence of induced bodily vibrations in the categorization of ambiguous emotional vocalizations in an event-related potential study (*N*=24). The factorial design included Vocal emotion [anger and fear] and external Vibration [anger, fear, and none] as fixed factors. 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 through transducers placed close to the vocal cords. We hypothesized that induced bodily vibrations would constitute an interoceptive and proprioceptive feedbacks that would influence the perception of emotions, especially for more ambiguous voices as ambiguity would favour the processing of other available sensory information, here toward the tactile sensory modality. Behavioural results revealed that induced vibrations skewed the participants' emotional ratings by biasing responses congruent with the vibration. Event-related potentials results indicated that N100 and P200 components subtending the early processing of emotional prosody were significantly modulated by induced vibrations in the congruent setting, which could be considered as a facilitation effect for emotion recognition at early stage of processing. A significant modulation of the late positive component was also observed in the incongruent setting, suggesting an error processing mechanism. EEG source reconstruction highlighted significant contrasts between vibration types in prefrontal, motor, somatosensory, and insular cortices. Altogether, our results suggest that voice-associated vibrations would play a significant role in vocal emotion perception and recognition through embodied mechanisms at both behavioral and neural levels.

Key words: *electroencephalography, emotional prosody, bodily vibrations, interoception, embodiment*

Introduction

Oral communication is essential for a social animal such as a human being, both in terms of semantic meaning of words and non-verbal parameters. The ‘emotional prosody’ of the voice refers to the changes in these acoustic parameters, and it also refers to the melody of the voice, which is related, among others, to the temporal fluctuations of the pitch, envelope amplitude, as well as the spectral fluctuations related to the timbre (Grandjean et al., 2006; Laukka et al., 2005; Risberg & Lubker, 1978; Scherer & Oshinsky, 1977). According to Scherer's adaptation of the Brunswik lens model, perception of emotional prosody—also named ‘vocal emotion’—variations allow the listener to make subjective attribution(s) of the speaker's emotional state(s) in social interactions, often influenced by context (Brunswik, 1956; Scherer, 2003, Grandjean et al., 2006).

One of the key mechanisms allowing this accurate attribution of a speaker's emotional state based on voice signals relies on the embodied ‘simulation’ of what we hear (Hawk et al., 2012). Emotional embodiment is defined as the motor, sensory and perceptual reexperiencing or simulating of the emotion perceived in a peer by the perceiver in the framework of the Simulation of Smiles model (SIMS model; Niedenthal, 2007, Niedenthal et al., 2010, p.418). This process has primarily been described in the visual modality, as authors have demonstrated that the perception of facial expressions in others generates facial mimicry in the observer, which is stronger for emotional compared to non-emotional facial expressions (Dimberg, 1982; Korb et al., 2010; Künecke et al., 2017; Moody et al., 2007; Moody & McIntosh, 2011; Chartrand & Bargh, 1999). Recently, many studies have supported the idea that facial mimicry represents the sensorimotor simulation of an observed emotion rather than a mere muscular reproduction of an observed facial expression, and is context- and affiliation-dependent (Borgomaneri et al., 2020; Hess & Fischer, 2014; Drimalla et al., 2019; Hess, 2021; Philip et al., 2018). Importantly, some studies suggest that this process may play a role in 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 this claim is contradicted by other studies (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). Another study

showed that language comprehension was facilitated when the primary motor areas controlling the lips or the tongue were stimulated transcranially, and this was true for sounds respectively produced by either of these oral organs (D'Ausilio et al., 2009). One of the mechanisms which may underly the effect of emotional embodiment in the emotional prosody domain relies on body resonances. In fact, such resonances—vibrations, especially in the throat region close to the vocal folds—may represent a series of interoceptive and proprioceptive feedbacks during vocal emotion production. Indeed, during voice production, vibrations from the vocal cords are transmitted through the speaker's skin and body tissues (Munger & Thompson, 2008; Švec et al, 2005). The frequency of these body resonances mainly corresponds to the fundamental frequency of the auditory tone that is produced (Sundberg, 1992). The sensation of body resonances would constitute for the speaker a stable and dynamic feedback on several characteristics of his or her own voice (Sundberg, 1992), and these can hardly be disturbed by the environment. 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 further research is needed to gain a deeper and clearer understanding of the different types of mechanisms that may be engaged in the embodiment of emotional prosody, both at the behavioral and brain levels. To investigate precise dynamic neural correlates of the embodiment of vocal emotion perception—a novel topic in the literature, we turned to surface electroencephalography (EEG). We will now start by introducing research on vocal emotion perception and show how this work constitutes a solid ground to build up and further investigate the dynamic mechanisms of the embodiment of emotion in the vocal domain.

Event-related potential (ERP) correlates of emotional prosody decoding have been extensively studied in the last two decades. Models of emotional prosody decoding suggest a processing in three-stage (Schirmer & Kotz, 2006). First, sensory processing in the auditory cortex and superior temporal sulcus (STS, 0-150ms) takes place with low-level acoustic analysis. The second stage (150-250ms) involves the integration of emotionally significant acoustic cues in the bilateral superior temporal gyrus (STG) and anterior STS and the third stage (>250ms) consists of emotional processing with high-level cognitive processes in frontal areas. 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 appear

to be involved in auditory-sensory processing, attentional processing and early recognition of meaning. Firstly, the N100 component is a negative deflection peaking around 100ms after stimulus onset and located most of the time in the frontocentral area. It might reflect an early component of low-level prosody analysis as well as emotion discrimination (Gädeke et al., 2013; Paquette et al., 2020; Tarai & Srinivasan, 2019). In other studies, basic emotions are also distinguished in the P200 component located mostly in the frontal, frontocentral and parietal areas (Duville et al., 2022; Maltezou-Papastylianou et al., 2022; Paulmann & Uskuul, 2017; Paulmann et al., 2013; Gädeke et al., 2013). The P200 component is a positive deflection around 200 milliseconds that is believed to reflect an increased attention to emotional stimuli so that they can be processed more effectively when needed (Paulmann & Kotz, 2008; Thomas et al., 2007). Later ERP 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). Indeed, negative or positive deflections around 300 and 400 milliseconds (P300/N400) located in frontal and frontocentral areas were observed during emotional prosody processing (Carminati et al., 2018; Proverbio et al., 2020; Tarai & Srinivasan, 2019) and were associated with the detection of deviant stimuli in an auditory oddball paradigm (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 that is 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). A study investigating the processing of emotional facial expression reported that the LPC may reflect attentional processes rather than emotional processing (Ashley et al., 2004), which is consistent with the attentional capture effect associated with violation or incongruence designs (Vachon et al., 2012). Other studies showed that patterns of mismatching prosodies or musical stimuli elicited an LPC that could indicate an integration process of the different sources of information (Brattico et al., 2006; Paulmann et al., 2012).

Functional MRI studies have demonstrated that emotional prosody perception activates both subcortical and cortical networks (Grandjean, 2012; Witteman et al., 2012). These networks include the bilateral temporal voice areas, inferior frontal gyrus, orbitofrontal area, cerebellum, and amygdala, among others. Additionally, studies have shown the involvement of the motor cortex during speech perception (Fadiga et al., 2002; Watkins et al., 2003). Furthermore, studies examining interoceptive signals have reported activations in the insular

cortex, a region responsible for the integration both internal and external physiological signals and emotional processes (Adolfi et al., 2017; Casals-Gutierrez & Abbey, 2020). Other structures, such as the somatosensory cortex, the supplementary motor area (SMA), the prefrontal cortex and the cingulate cortex, have been found to play a 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 proprioceptive feedbacks represented by induced bodily resonances—through induced vibrations, and their impact on emotional prosody perception. Revealing such as link between vocal emotion perception and induced body resonances would constitute convincing evidence for a crucial role of vibro-tactile feedback as a physiological mechanism of vocal emotion embodiment. An association between interoception and emotional processing was shown in numerous studies but never when using the vibrations concurrent to expressed voice signals (Couto et al., 2015; Critchley, & Garfinkel, 2017; Burleson & Quigley et al., 2021). For example, Mendoza-Medialdea & Ruiz-Padial (2021) revealed the role of interoceptive accuracy on exogenous attention to disgusting and fearful distractors of a competing task. Several ERP studies investigating interoception in emotion processing showed that there is a link between the interoceptive sensitivity as measured with a heartbeat detection task and the cortical processing of emotional stimuli. Indeed, significant differences in the P300 component and in a slow wave window (550-900 ms) have been reported all over the scalp between good and poor heartbeat perceivers while rating emotional pictures (Herbert et al., 2007; Pollatos et al., 2005). A somatosensory-specific P300 component was also revealed in parietal regions in studies investigating somatosensory perception (Al et al., 2020; Auksztulewicz, & Blankenburg, 2013; Has et al., 2021). Moreover, Salamone et al. (2021), showed that negative facial emotion recognition was improved after focusing on one's own heartbeat and that heart-evoked potentials around 250-300 milliseconds were modulated by this interoceptive priming task. To our knowledge, no study to date has specifically examined the role of vibrations as interoceptive and proprioceptive feedback signals in vocal emotion perception using EEG recordings. While our previous behavioral study explored a related question, its design differed by focusing on perceptual rather than emotional ambiguity, and its findings did not provide conclusive evidence. This highlights the need for further investigations into the potential impact of bodily resonances on auditory emotion perception.

Nevertheless, multimodal integration/multisensorial studies using audiovisual emotional stimuli have been conducted in the past and they showed that multimodal information

seems to facilitate emotional processing compared to unimodal information (Föcker & Röder, 2019; Kotz, Jessen & Paulmann, 2009) and that multimodal signals interact at several stages of the audiovisual integration process (Klasen et al., 2012). Moreover, it appears that ERP components usually involved in emotional processing are modulated by the interaction of auditory and visual information (Klasen et al., 2012). Indeed, Föcker & Röder (2019) showed several interesting results: i) performance in an emotion recognition task was higher in congruent bimodal trials compared to unimodal ones; ii) the earliest multisensory interactions they reported were represented by a greater positivity for congruent face-voice stimuli in the P200 time range (180-230ms); iii) the N400 component was more negative for incongruent than for congruent emotional expressions in face-voice pairs. Furthermore, Pourtois and colleagues (2000) showed in an audiovisual task of emotion recognition that the audiovisual integration was probably taking place at a very early stage (110ms post-stimulus), with the amplitude of the auditory N1 being enhanced for bimodal stimuli. In addition, behavioral (Brosch, Grandjean et al., 2008) and EEG (Brosch, Sander et al., 2008) studies further support the idea that emotionally relevant stimuli can modulate attention and early perceptual processing across modalities. Although the current study involves different modalities as compared to existing literature (vibro-tactile 'interoception' and audition), we rely on these promising results to investigate the impact of interoceptive and proprioceptive information on emotional auditory processing.

To the best of our knowledge, no study has specifically investigated the contribution of body vibrations to the auditory perception of emotions using EEG. While our previous behavioral study addressed a related question, its design differed, and its findings were inconclusive. The present study aims to further explore this mechanism. The main hypothesis is that vocal cord vibrations, which are a 'mechanical' consequence of the production of vocalized sounds, would represent interoceptive and proprioceptive feedbacks that could impact the auditory perception of emotions from 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. Since the SIMS model specifies that the effect of embodied simulation is observed when perceiving non-prototypical emotional stimuli, our study uses ambiguous emotional vocal stimuli in order to favor the triggering of such effect. Our study targets 'emotional ambiguity', with stimuli simultaneously conveying two emotions—i.e., anger and fear—at varying percentages using an auditory morphing procedure (Latinus & Belin, 2011). The vibrations induced in two-thirds of the trials

either expressed pure anger or pure fear, and were either fully congruent, partially incongruent, or fully incongruent with the auditorily-presented emotional vocalizations (see Methods section for details). At the behavioral level, two key hypotheses were tested for this study. First, we postulated that vibrations would affect the discriminative perceptual processes of emotional vocalizations. Specifically, it was predicted that participants would respond "anger" more frequently when the vibration also expressed anger and likewise for fear trials—when compared to the 'no-vibration' condition. Second, these effects were expected to be observed especially for ambiguous stimuli (i.e., voices containing 40-60% of each emotion) and not for unambiguous stimuli (i.e., voices containing 90% of each emotion). Indeed, the more emotionally ambiguous stimuli would transmit unclear input and would therefore be the conditions in which information is sought elsewhere, based on the SIMS model (Niedenthal, 2007, Niedenthal et al., p.418)—e.g., via embodied cognition here represented and/or triggered by induced vibrations. At the neural level, we first expected amplitude changes for ERP components associated with emotional prosody processing—such as the N100, P200 and LPC in fronto-temporal electrodes (Paquette et al., 2020; Paulmann & Kotz; Paulmann, & Uskul, 2017, 2013; Pourtois et al., 2004; Wei et al., 2022; Zora et al., 2020). Second, we postulated the presence of vibration-specific ERPs between 200 and 300ms modulated by interoceptive and 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 or pre-SMA (Adolfi et al., 2017; Casals-Gutierrez & Abbey, 2020; Engelen et al., 2023; Gibson, 2019; Khalsa et al., 2018), for anger and fear vibrations compared to the conditions without vibrations. Third, as the design included in some trials an incongruence between emotional vocalizations and vibrations–anger/fear vocalizations mismatched with fear/anger, we expected as well the presence of early (100ms and 200-300ms) and late (400-700ms) ERP components with source modulations in the above-mentioned brain areas, which can be related to to error prediction and/or integration of interoceptive and proprioceptive feedbacks in the context of emotional prosody information processing.

Results

In this study, the participants had to categorize emotional voices that were morphed with a continuum between anger and fear, at varying percentages. In two-thirds of the trials, bodily vibrations were induced on their throat through a vibrating device, and these vibrations expressed either pure anger or pure fear. We analyzed the influence of these vibrations on behavioral responses and ERPs and we further extracted source modulations in the brain.

Behavioral responses

The main effect of Vibration was significant ($\chi^2(2) = 18.10$, $p<.001$, Figure 1**A**). The main effect of Emotional Voice was also significant ($\chi^2(1) = 55.93$, $p<.001$, Supplementary Figure 1) as well as their interaction ($\chi^2(2) = 9.29$, $p< .01$, Figure 1**B**). The effect size of this model was medium ($R^2_{marginal} = .10$, $R^2_{conditional} = .58$[1]).

Further investigation of the impact of Vibration on the participants' responses was performed using planned contrasts related to our hypotheses and the results were the following (Fear vibrations: *m*=.57, *sd*=.50, Anger vibrations: *m*=.53, *sd*=.50, No vibrations: *m*=.55, *sd*=.50): i) [Vibration Anger > Fear] ($\chi^2(1) = 7.25$, $p<.05$), ii) [Vibration Fear > None] ($\chi^2(1) = 4.23$, p=12, and iii) [Vibration Anger > None] ($\chi^2(1) = 2.07$, p=.45; Figure 1**A**.). Concerning the impact of the interaction between Vibration and Emotional Voice on the participants' responses, planned contrasts were also performed: [Vibration Anger > Fear] x [Ambiguous (A_{60}, A_{50}, A_{40}) voice stimuli] ($\chi^2(1) = 7.25$, $p<.05$), [Vibration Anger > None] x [Ambiguous (A_{60}, A_{50}, A_{40}) voice stimuli] ($\chi^2(1) = 7.78$, $p<.05$), [Vibration Fear > None] x [Ambiguous (A_{60}, A_{50}, A_{40}) voice stimuli] ($\chi^2(1) = 1.31$, p=.76; Figure 1**B**). More specific contrasts on the most ambiguous condition of Emotional Voice were computed: [Vibration Anger > Fear] x [Ambiguous (A_{50}) voice stimuli] ($\chi2(1) = 3.96$, p=.14), [Vibration Anger > None] x [Ambiguous (A_{50}) voice stimuli] ($\chi2(1) = 6.64$, p<.05), [Vibration Fear > None] x [Ambiguous (A_{50}) voice stimuli] ($\chi2(1) = 1.307$, p=1; Figure 1B). All p-values were adjusted using the Bonferroni correction method (Bland & Altman, 1995).

---

[1] R^2marginal represents the proportion of variance explained by the fixed effects alone, while R^2conditional accounts for both fixed and random effects, providing a measure of the total variance explained by the model (Nakagawa & Schielzeth, 2013)

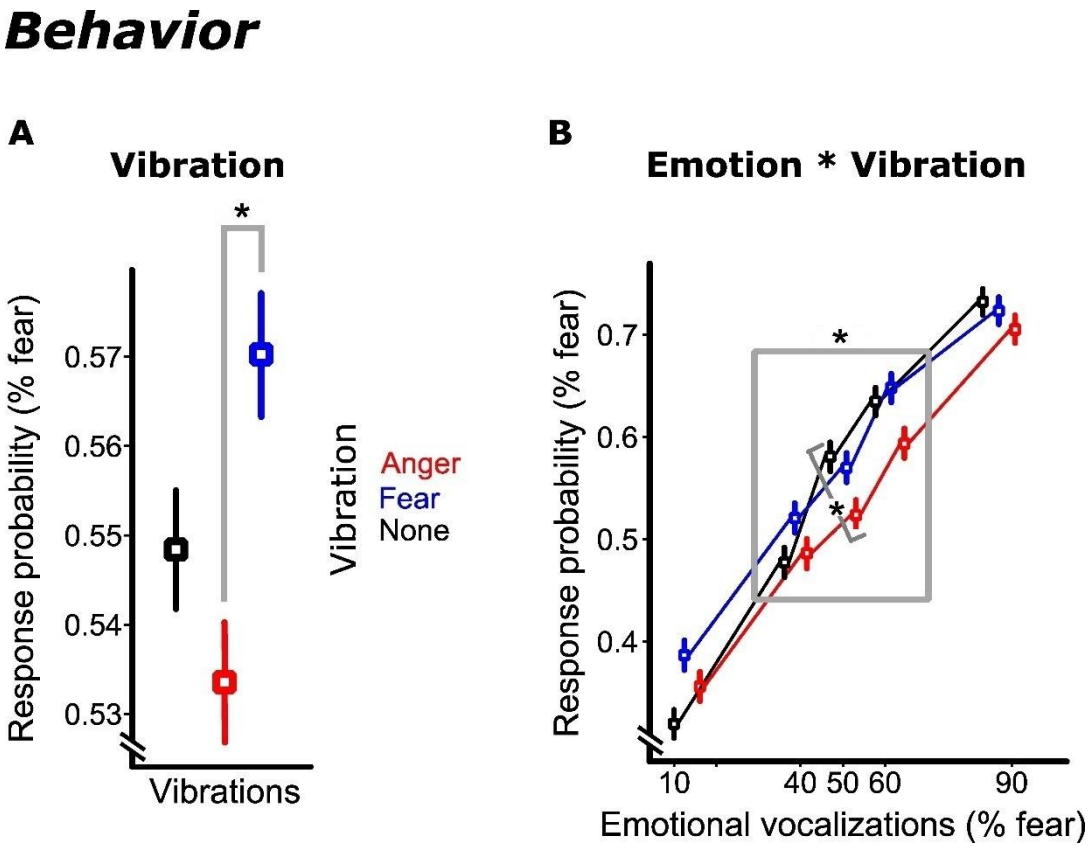

Figure 1. **A. Main effect of Vibration on the probability of fear responses.** On the X axis, each point represents a Vibration condition (i.e., None = no vibrations, Anger, Fear). **B.** Interaction between Emotional Voice and Vibration on the probability of fear responses. On the X axis, each point 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<.001.

EEG results

Event-Related Potentials

We started by analyzing ERPs to investigate the influence of induced vibrations during vocal emotion processing on brain dynamics. Significant differences between our clusters in terms of response to our conditions is materialized by a significant three-way Cluster X Emotional Voice X Vibration interaction along trial length (Supplementary Table 6, $F(24)$=[3.04-23.36 min-max]; p<.001 FDR corrected). Following clustering data reduction (see Methods), three clusters contained at least one 50ms timebin presenting an Emotional voice X Vibration two-way interaction. In the paragraph below, we describe the results from the functionally best-defined cluster, i.e., the frontocentral cluster (cluster C4, Figure 2**G**, Figure 3**K**). Moreover, the frontocentral zone is the most represented in the literature on emotional prosody (Duville et al., 2022; Paquette et al., 2020), as well as being relevant for somatosensory

perception (Desmedt & Bourguet, 1985; Macerollo et al., 2018). Results from the two other ERP clusters (occipital and temporal) are described in the supplementary material ('Supplementary EEG results', *I.*). The results presented below pertain directly to our hypotheses while exhaustive results are reported in the supplementary material for ERP data ('Supplementary EEG results', *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 2**G** & Figure 3**K**) presented an Emotional Voice X Vibration two-way interaction at the 50-150ms, 300-400ms, 450-550ms, 600-650ms, 750-850ms, and 900-950ms time intervals (Supplementary Table 6, Figure 2 and Figure 3). Three clearly defined peaks exceeded the baseline threshold (see Methods): a positive deflection from 100 to 150ms, a positive deflection from 200 to 350ms, and a negative deflection starting 500ms post-stimulus onset.

Low emotional voice ambiguity

At an early stage—the 100ms peak, no expected results were observed while other non-planned contrasts revealed significant. These are described and reported in the supplementary material (Figure S2, Supplementary Table 5). At 250ms, anger vibrations—compared to fear vibrations and no-vibration—led to a congruent *increase* of the P200 peak amplitude for mostly-angry (A_{90}) voices (Figure 2**A**, Supplementary Table 5). At a later stage—at 750ms, anger vibration was associated in congruent A_{90} voices to a significantly weaker negative deflection compared to the no-vibration condition (Figure 2**A**).

High emotional voice ambiguity

In the high ambiguity setting, fear vibration was associated with an incongruent decrease of P200 peak amplitude for A_{60}/A_{50} voices, which persists in the descending phase of the P200 peak (Figure 3**A**). Additionally, in the descending phase of the P200 peak, a significant power decrease was also observed for A_{60} voices under pure anger vibration induction (Figure 3**A**). Moreover, at a later stage—around 600ms, we observed a congruent significant increase in power amplitude for A_{40} voice for fear vibrations compared to the no-vibration condition (Figure 3**F**). Finally, at 750ms, induced fear vibration was associated with a significantly stronger negative deflection for congruent A_{40} voices (Figure 3**F**). Induced anger vibration was associated with a significant relative power decrease compared to no-vibration in A_{60} voices only (Figure 3**A**). Anger vibration also led to a significantly stronger negative deflection

compared to no-vibration in A_{60} and A_{40} voices (Figure 3**A**,**F**). Finally, induced fear vibration was associated to a significantly stronger negative deflection compared to no-vibration when perceiving A_{40} voices (Figure 3**F**).

Source reconstruction

Source reconstruction analysis revealed correlates at the brain level of the vibration ERP-dissociations observed at the sensor level (Figure 2**B**-**D**,**F**, Figure 3**B**-**E**,**G**-**J**).

Low emotional voice ambiguity

We observe in the 300-350ms time interval a significant fear vs anger vibrations source-ERP contrast in the superior posterior lobe (Figure 2**B**,**C**). At a later stage, we observed a significant incongruent fear vs no-vibrations source-ERP contrast in the medial orbitofrontal cortex (OFC), insula and primary somatosensory cortex (550-600ms, Figure S4) followed by the supplementary motor area (750-800ms) and primary motor cortex (900-950ms) (see Figure 2**D**). For fearful voices, significant congruent fear vs no-vibrations contrasts were observed in the supplementary motor area (700-750ms) and superior parietal lobe (700-750ms; 850-900ms). See Figure 2**F**.

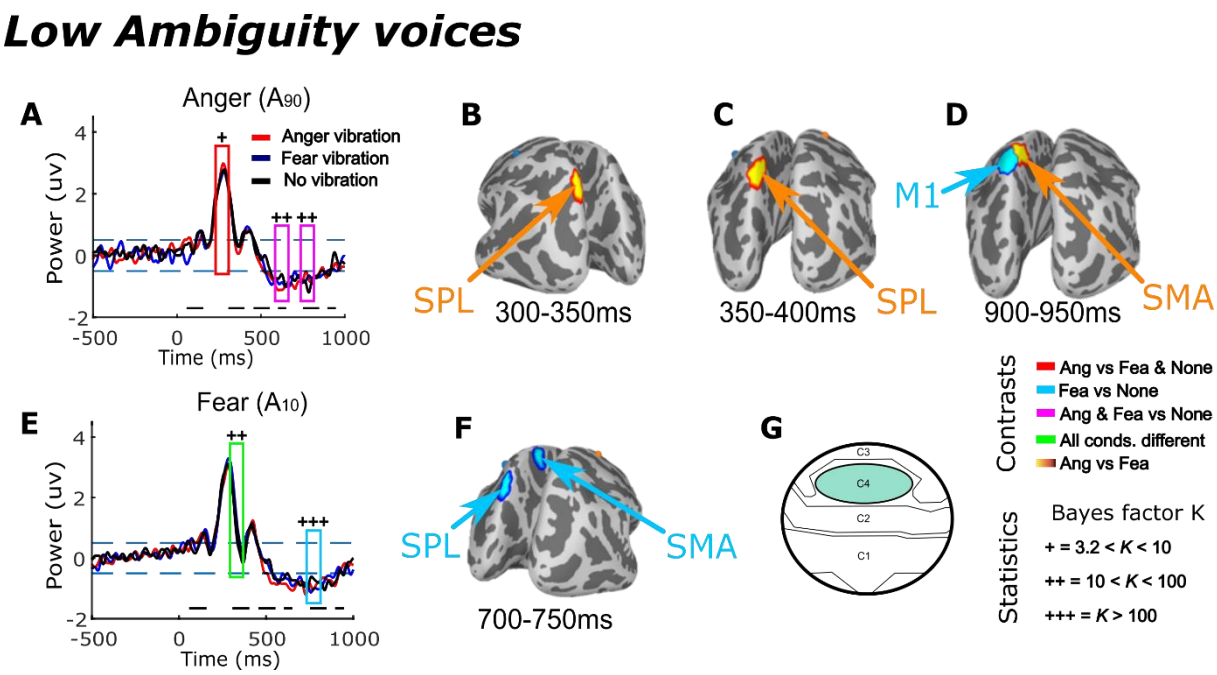

Figure 2. **Contrasts between vibrations for Low Ambiguity vocalizations.** Top panel: Contrasts between vibrations for low ambiguity angry voices (A_{90}) for **A:** Event-related potentials and **B-C-D**: Brain activations from source reconstruction. Bottom panel: Contrasts between vibrations for low 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 significant. 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

At the 350-400ms peak, significant source-ERP fear vs no vibration contrasts were observed in the anterior cingulate cortex and anterior insula for fearful voices (Figure 3**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 3**G**). At the 450-500ms peak, a significant congruent anger vibration vs no-vibration contrast was observed for angry voices in the primary sensory, primary motor, and premotor areas, orbitofrontal cortex, dorsolateral prefrontal cortex (DLPFC) as well as in the intraparietal sulcus (IPS; Figure 3**B**). In the late period, a significant congruent anger vibration vs no-vibration contrast was observed for angry voices in the precuneus (700-750ms, Figure 3**D**), as well as superior parietal lobule and IPS (900-950ms, Figure 3**E**). A significant congruent fear vibration vs no-vibration contrast was also observed for fearful voices in the primary motor cortex, IFG *pars opercularis* and OFC from 650 to 700ms while a significant fear and anger (global concatenation of emotion) vibration vs no-vibration contrast reached significance in the primary motor area at the same timing (Figure 3**I**). Finally, at 750-800ms, an incongruent anger vibration vs no-vibration contrast was observed for fearful voices in primary motor cortex and SMA (Figure 3**J**).

High Ambiguity voices

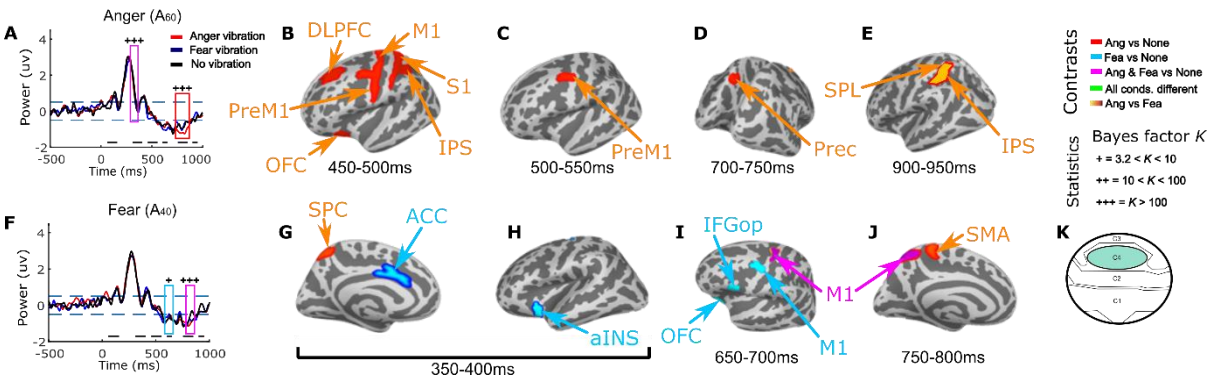

Figure 3. **Contrasts between vibrations for High Ambiguity vocalizations.** Top panel: Contrasts between vibrations for high ambiguity angry voices (A_{10}) for **A**: Event-related potentials and **B-C-D-E**: Brain activations from source reconstruction. Bottom panel: Contrasts between vibrations for 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 pertaining to the impact of embodied cognition on vocal emotion processing is so far lacking in the literature. Our study sought to demonstrate and better understand the influence of induced bodily vibrations on the perception of emotionally ambiguous voices, at the behavioural and neural levels. The research question revolved around the possible role of the vocal tract vibrations —here induced and systematically manipulated through a device in the throat region, necessary for vocalization and triggered by the use of vibrators, and whether these could represent some type of feedback—potentially interoceptive and proprioceptive feedbacks—that would influence auditory emotion perception, especially in the case of ambiguous voice stimuli. Both behavioural and EEG results revealed effects that support our assumptions: induced vibrations triggered a modulation of behavioral responses to emotional vocalizations as well as of early and late ERP components—highlighting relevant and expected brain modulators through source reconstruction.

As mentioned above, our study showed 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 significant, it was clearly observed that our participants categorized emotional vocalizations as ‘anger’ significantly more often when a concomitant anger-like vibration was induced than when it was a fear-like vibration. This is consistent with the study of D’Ausilio et al. (2009) showing the facilitation effect of language comprehension by the transcranial stimulation of the primary motor areas controlling the lips or the tongue. Indeed, in our study, the vibrations would stimulate the mechanoreceptors in the throat in an emotion-specific pattern and this mechanism would then facilitate the recognition of the emotion expressed auditorily and/or bias the response curve—since there are no correct responses for ambiguous voices since they are a blend of fear and anger. This is also coherent

with studies investigating multimodal integration, which showed a congruence-specific facilitation effect in multimodal processing of emotion (Lin & Ding, 2019; Niedenthal, Brauer, Halberstadt, & Innes-Ker, 2001).

Following a similar reasoning, participants categorized ambiguous emotional vocalizations (A_{60}, A_{50}, and A_{40}) as 'anger' more often when anger-like vibrations were induced than when fear-like vibrations were induced—or when there were no vibrations induced at all. This specific contrast between anger-like vibrations and no vibration is consistent with the SIMS model (Niedenthal et al., 2010), postulating that interoceptive feedback would be integrated for non-prototypical emotional stimuli, when classic available emotional information are not clear enough to identify these specific stimuli— here our ambiguous morphed voices.

At the neural level, we chose to focus on the functionally best-defined cluster as well as the most relevant one in terms of location (Duville et al., 2022; Gädeke et al., 2013; Maltezou-Papastylianou et al., 2022; Paquette et al., 2020), i.e., the frontocentral cluster (C4). The frontocentral cluster was the locus of an interaction between voice ambiguity and induced vibrations—for 50-150ms, 300-400ms, 450-550ms, 600-650ms, 750-850ms, and 900-950ms time intervals. Three clearly defined peaks exceeded the baseline threshold (see Methods): a positive deflection from 100 to 150ms, a positive deflection from 200 to 350ms and a negative deflection starting 500ms post-stimulus onset. Our results showed that the induction of both congruent fear and anger-like vibrations led to an increase in P100 amplitude in both low- and high-ambiguity situations, then including the most ambiguous voices. This positive deflection around 100ms recalls the study of Mendoza-Medialdea & Ruiz-Padial (2021) in which interoceptive accuracy was measured in a heartbeat detection paradigm, for which the participants had to indicate whether an audible tone was played at the same time as their heartbeat or presented with a delay (500ms). In this study, interoception accuracy impacted the capture of exogenous attention by emotional stimuli as early as 100ms post-stimulus onset—namely, the P100 component. Also—and although it was in the audiovisual modalities, the results of Pourtois and colleagues (2000) in their emotion recognition study indicated that multimodal integration could occur at a very early stage of processing, around 110ms post-stimulus onset. Furthermore, studies investigating interoception and somatosensory stimulation observed effects on the whole scalp and particularly in the parietal area—in line with our results, albeit at a slightly later stage with a modulation of the P300 component (Herbert et al., 2007; Pollatos et al., 2005; Al et al., 2020; Auksztulewicz, & Blankenburg, 2013; Has et al., 2021). Moreover, the fact that emotional vibrations have this effect notably for ambiguous voices

supports the assumption of Niendenthal et al (2010) described in their SIMS model. Indeed, ambiguous voices here represent non-protoypical emotional stimuli and therefore other sources of information should be sought to better understand and interpret the expressed emotion. Therefore, and according to our results, vibrations in the throat region near the vocal cords would therefore constitute a reliable source of feedback to be considered and further integrated when emotional processing is initiated and ongoing.

In our EEG results, we also observed the presence of a P200 component in this frontocentral area. This is consistent with the vast majority of studies investigating emotional voice processing (Duville et al., 2022; Maltezou-Papastylianou et al., 2022; Paulmann & Uskuul, 2017; Paulmann et al., 2013; Gädeke et al., 2013). Our results showed that anger-like vibrations lead to a congruent *increase* of the P200 peak for low-ambiguity voices, while fear-like vibrations lead to a *decrease* of the P200 peak for ambiguous voices, and for incongruent high-ambiguity angry voices. These results support the above-mentioned findings on multimodal integration and interoception and occur at a slightly later stage. Indeed, we have here congruent vibrations that seem to impact the perception of emotional voice stimuli, which is in line with studies showing that multimodal information seems to facilitate emotional processing when compared to unimodal information (Föcker & Röder, 2019; Kotz et al., 2009). The results also align with previous research on interoception and somatosensory stimulation instigating a P300 component, throughout the scalp and specifically in the parietal region, as observed in the past (Herbert et al., 2007; Pollatos et al., 2005; Al et al., 2020; Auksztulewicz, & Blankenburg, 2013; Has et al., 2021). Moreover, the effect of fear-like vibrations for ambiguous voice categorization only is also supported by the SIMS model (Niedenthal, 2007, Niedenthal et al., 2010, p.418). Concerning the P200 decrease for incongruent fear-like vibrations in high ambiguity mentioned above, this result may be due to the fact that, at high ambiguity, voice stimuli are almost equally composed of congruent and incongruent information. In this case, it is possible that the 40% of fear contained in the high-ambiguity angry voices may have taken over from the angry part and that would lead to this apparent and subtle incongruent effect.

Our data finally show late, negative ERP modulations by induced vibrations—most notably between 500 and 800ms post-stimulus onset. These modulations involve both congruent and incongruent deflections, especially for low-ambiguity voices and for both emotions. For instance, at 500ms, anger vibrations produce an incongruent negative deflection for low-ambiguity fearful voices. These results are consistent with literature on vocal emotion

processing, as the negative deflections around 500 and 600ms recalls 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, an increase in power is observed for both emotional vibrations compared to no-vibration for ambiguous voices. In the literature, at this point in time in the processing of emotional voices, there is usually instead a positive deflection, called the Late Positive Component (LPC), which is 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). However, in our study, although the power increases during this period, it still remains negative. Nevertheless, this echoes once again the study of Niedenthal and colleagues (2010), which postulates that embodied simulation is only engaged when the emotional stimulus is non-prototypical. 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).

The ERP data were also used for source reconstruction, revealing source modulations in regions involved in vibratory processing, emotional prosody perception, multimodal integration, proprioception and interoception. Regarding vibratory processing, we observed source modulations in the dorsolateral prefrontal cortex, involved in vibrotactile working memory (Burton et al., 2010). Other regions such as the superior parietal lobule, the inferior parietal sulcus, the precuneus, the insula, the supplementary motor area, as well as the premotor, and the primary somatosensory cortices are known for their involvement in vibratory processing. In fact, they are crucially observed in vibration rate and/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 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, importantly associated with interoceptive awareness (Herbet et al, 2019; Longarzo et al., 2021; Wilson-Mendenhall et al., 2019). The modulation of the primary somatosensory cortex might be more related to proprioceptive information

processing. Further source modulations were observed in the orbitofrontal cortex, which has also been reported in interoception studies (Singer & Klimecki, 2014; Stern et al., 2017). Regarding emotional prosody perception, we observed source modulations in the dorsolateral prefrontal cortex known for its implication in prosodic emotion working memory (Mitchell, 2007) and the superior parietal lobule found in explicit judgments of emotional prosody (Bach et al., 2008), as well as other regions also known to be involved in emotional prosody processing such as the intraparietal sulcus and the supplementary motor area (Alba-Ferrara et al., 2011; Belyk & Brown, 2014; Kotz et al., 2003). Interestingly, as already mentioned, we also observed source modulations in the orbitofrontal cortex, associated with higher-level processing of emotional prosody (Wildgruber et al., 2006), especially in ambiguous situations (Grandjean, 2021). Regarding multimodal integration, we observed source modulations in the insula, also crucially involved in integrating external emotional stimuli and emotional experiences with interoceptive signals (Adolfi et al., 2017; Nguyen et al., 2016; Zaki et al., 2012). As mentioned above, the superior parietal lobule, intraparietal sulcus, premotor and primary somatosensory cortices were also modulated, and they are regions known for their involvement in multimodal integration for auditory, visual, sensorimotor and/or tactile stimuli (Baker et al., 2018; Culham & Valyear, 2006; De Borst & De Gelder, 2016; Gentile et al., 2011; Kreifelts et al., 2007; Parker et al., 2020). Taken together, these results using linear constraint minimum variance (LCMV) beamformer source reconstruction suggest the implication of a wide-spread network of brain areas in the processing of induced vibrations and their integration with emotional vocalizations, especially in the case of highly ambiguous stimuli, for which such information is most relevant for improved decision making. More specifically, since the above regions are strongly involved in interoception, our results suggest that the vibrations induced on our participants' throat are processed as interoceptive feedback, which would be integrated into the processing of emotional vocalizations and thus impact auditory perception and decision.

Limitations and perspectives

Our study has some limitations that we need to state. First, our sample consists mainly of female participants, despite studies indicating gender differences in emotional prosody recognition (Imaizumi et al., 2004; Lambrecht et al., 2014) and facial mimicry (Dimberg, & Lundquist, 1990; Niedenthal et al., 2012, Korb et al., 2015). Second, it would be necessary to replicate the effects we observed in our study with different emotions, particularly positive ones, to enable a more reliable generalization of the results. However, morphing on other emotions leads to less natural blends and sometimes strange-sounding voices, which might

confuse the participants. Third, since the vibrations induced in our participants are supposed to be similar to the ones initiated by the vocal cords during vocal production, it could be more accurate to use voice stimuli created by the participants themselves. Indeed, acoustic parameters of the vibrations would be much closer to their own bodily vibrations created during vocal production and therefore potentially amplify or solidify the observed effect. We would like also to stress that the induced vibrations were based on recorded sound and the vibrations were initiated on the skin. Such skin vibrations were thought to mimic the vocal chords vibrations but, of course, these latter might induce subtle tissue vibrations around them, encoded by different mechano-receptors, which might not be induced by the vibration delivered at the surface of the skin, even if the vibrators were localized at the closest possible location related to vocal chords. Further research is needed to better clarify the relationships between vibrations induced by the vocal chords itself and the induced vibrations at the skin level (i.e., the ones that can be systematically manipulated at the experimental level). Fourth, the correction of *p*-values for multiple comparisons resulted in the elimination of significant trends in some of the results, particularly those concerning the contrasts between vibration conditions. It may therefore be related to the sample size, and with a larger number of participants or with a better balance in the number of men and women participants, these trends could emerge as significant. Finally, future studies would benefit from the inclusion of the Multidimensional Assessment of Interoceptive Awareness (MAIA; Mehling et al., 2012) questionnaire in the procedure. While the questionnaire does not address vibrations specifically—but rather other bodily sensations, it would still be relevant for such studies. Due to variability in this dimension amongst individuals, using the interoceptive awareness questionnaire as a participant-level covariate could improve the control over the data as far as interoception is concerned and it could also allow for an interesting clustering of the participants on that relevant dimension.

Conclusion

In conclusion, our study provides fresh insights into the topics of embodied cognition and interoception mechanisms in the context of vocal emotion perception. Indeed, both our behavioral and EEG findings revealed effects that supported our hypotheses. Specifically, induced vibrations exerted an impact on emotional vocalizations perceptual decisions, and also influenced several early and late ERP components. Moreover, the source reconstruction analysis revealed source modulations in regions involved in vibratory processing, i.e. encoded by the different kinds of mechano-receptors, emotional prosody perception, multimodal integration, and interoception. Our findings suggest that vibrations do indeed constitute an

interoceptive signal that is integrated during several stages throughout the process of vocal emotion perception. Further investigation is needed to explore the mechanism(s) involved more deeply, as we believe we are just getting started on this journey toward a more integrated understanding of the influence of somatosensory inputs on auditory perception.

Method

Participants

Twenty-four healthy participants were recruited through posters and via social media dedicated to students from the University of Geneva (18 females and 6 males; Age range = 18-41; M_{Age}=23.67; SD_{Age}=5.11). Sample size (N=24) was decided upon using a power analysis adjusted to the design of our study with an effect size of .25, a power of .8 and an alpha of .05 (G*Power 3; Faul et al., 2007). All participants were at least 18 years old, had normal hearing and no neurologic or psychiatric history. Participants gave informed and written consent for their participation in the experiment. The study was approved by the local ethics committee in accordance with ethical and data security guidelines of the University of Geneva and conducted according to the declaration of Helsinki.

Material

Participants were presented with voice stimuli morphed including both anger and fear voices, presented through headphones (HD25, Sennheiser, DE) with Matlab (The Mathworks Inc., Natick, MA, USA), using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). The stimuli were presented while a fixation cross was displayed on the screen. The morphing was performed by mixing (STRAIGHT toolbox; Kawahara & Matsui, 2003) a fearful expression with the speaker's identity-matching angry expression, resulting in the five conditions (fear%-anger%): 10-90 (A_{90}), 40-60 (A_{60}), 50-50 (A_{50}), 60-40 (A_{40}), and 90-10 (A_{10}). We used fifty stimuli (five morphing conditions times ten actors) and their acoustic parameters were analyzed (Supplementary Table 1). Voice stimuli were not significantly different on any parameter (amplitude, pitch, loudness, duration, spectral centroid and slope, and F1 frequency). Morphing conditions and actors were randomly distributed across trials. The choice of emotions to be morphed was restricted to anger and fear because (i) mixing positive and negative emotions does not give consistent, human-like stimuli, (ii) including more morphing types would drastically increase the duration of the study that already lasted 2 hours,

(iii) these two emotions are among the strongest, biologically-relevant ones for survival. Voice and vibration stimuli included ‘Aah’s’ from the Montreal affective voices database (Belin et al., 2008).

Bodily resonances were induced in the form of vibrations expressing either pure anger or pure fear in two-thirds of the trials while no vibrations were induced in the remaining third of the trials as a control measure. Three conditions of bodily resonances were therefore implemented (anger, fear, no vibrations). Bodily resonances were created by transmitting emotional voice stimuli to a device that mechanically converted these sound waves into vibrations. The vibrator’s (BC-10, Ortofon, DK) dimensions are 13.5*29.5*18.0 millimeters, it weighs 16.5 grams, has a sensitivity of 118 decibels, a total harmonic distortion of 1.5%, an impedance of 15 ohm and sensibility ranging from 100 to 1000 Hz with a sampling rate of 1000Hz. It was positioned close to the vocal cords (Figure 4), taped on the left side to the laryngeal prominence, also known as Adam’s apple, with kinesiology tape—being both soft and strong at the same time. To create the vibration, the vibrator transforms the waveform of an input sound into mechanical energy, here a vibratory signal. Many studies have explored the location of these vibrations, and it appears that regions such as the nasal bone, zygomatic bones, temples, above the upper lip and in the upper neck are the most appropriate to consider (Munger & Thomson, 2008; Nolan et al., 2009; Sundberg, 1992). These vibrations came from the same database as the voice stimuli, but were not morphed (100% anger or 100% fear) and were matched for stimulus duration. The same speaker was used for both vibrator and voice stimuli for the sake of consistency and to avoid biases and strange intonation effects that could arise when mixing actors. We therefore used twenty stimuli for the vibration (two emotions x ten actors) and their acoustic parameters were analyzed (Supplementary Table 2). The acoustic parameters of these vibrations were those of the emotional voice stimuli used, and after a Bonferroni correction (Bland & Altman, 1995), no significant differences emerged.

Task Procedure

The experiment comprised six experimental blocks (two fear-induced and two anger-induced vibrations blocks, alternately), each including one hundred and thirteen randomly presented stimuli (Figure 4). A trial consisted of the presentation of a centered fixation cross screen for five hundred milliseconds, then the presentation of a voice stimulus for one second while the display of the same centered fixation cross screen, and then the categorization of this stimulus by the participant in a two-alternative forced choice between anger and fear using two keyboard keys. Key labels were counterbalanced across participants. Participants were

instructed to answer as fast and as accurately as possible and the response screen lasted for three seconds. Bodily resonances were induced in two-thirds of the trials ($N_{VibrationTrials}$= 452, $N_{TotalTrials}$= 678) as mentioned above, with vibrations expressing pure anger or pure fear, matching for stimuli duration and speaker identity with the auditory voice stimuli. The occurrence of the vibrations—in the vibration trials—was either emotionally congruent or incongruent with the emotional vocalization they accompanied. Therefore, the final factorial design includes two factors: Emotional Voice (A_{90}, A_{60}, A_{50}, A_{40}, A_{10}) and Vibration (No Vibrations, Anger, Fear) for a 5 x 3 design.

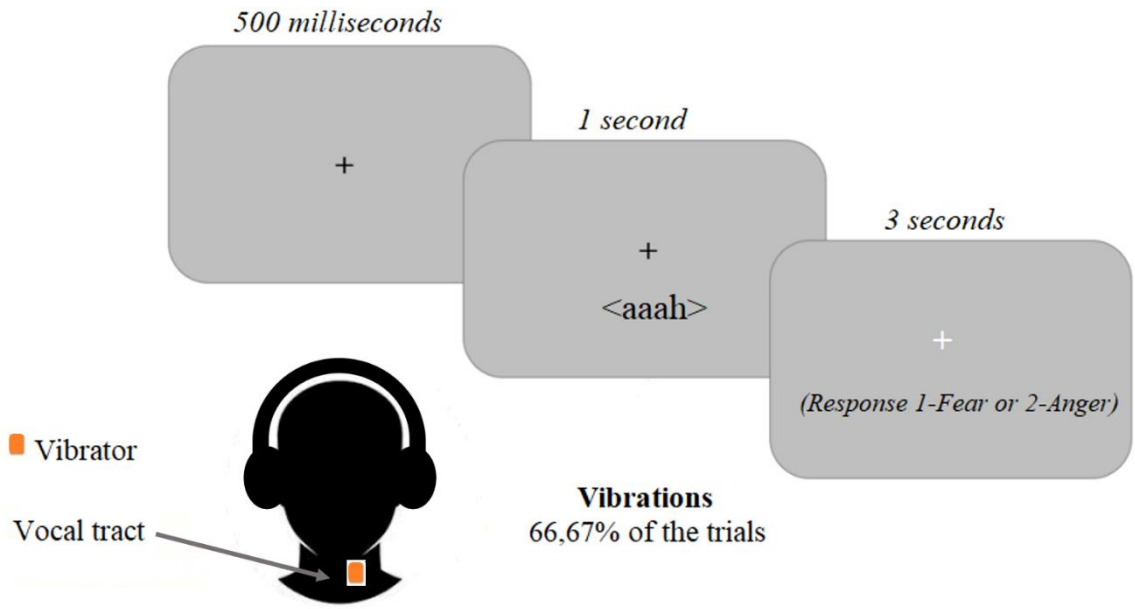

Figure 4. **Experimental design.** Participants were instructed to focus on a black central fixation cross displayed on a grey screen for five hundred milliseconds. Emotional voice stimuli were presented through headphones for one second. Vibrations were induced at the same time in two thirds of the 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

Behavioral analysis

Participants' responses were categorical, with only two possible responses coded as follows: anger = “0” and fear = “1”. A Generalized Linear Mixed Model (GLMM) was performed in R (R Core Team, 2020, Version 1.4.1103) to analyze the responses as a function of both the Vibration (No vibrations, Anger, Fear) and Emotional Voice (A_{90}, A_{60}, A_{50}, A_{40}, A_{10}) factors and actor gender (male, female) was controlled for as well. An LMM was chosen to analyze the reaction times with the same factors (results in Supplementary Table 1). These

models are relevant to our study for their integration of random effects (Bolker et al., 2009; Venables, & Dichmont, 2004). For both models, fixed effects included the interaction between Vibration and Emotional voice, the main effect of participant and speaker gender. In addition, random effects included the random slope of Vibration and Emotional voice as a function of participant identity (uncorrelated) and the random intercept of the stimulus identity. The formula was the following (glmer for response, lmer for reaction times):

Model.response<-
*glmer(Response~Vibration*EmotionalVoice+ParticipantGender+SpeakerGender+(1+Vibration+EmotionalVoice||ParticipantID)+(1|StimulusID), data=data, family = binomial)*

The effect size of each model was computed and labelled according to the thresholds defined by Cohen (1988; *r* effects: small ≥ .10, medium ≥ .30, large ≥ .50). Planned contrasts have been performed according to our hypotheses, using Bonferroni correction for multiple testing (Bland & Altman, 1995).

EEG preprocessing

A classical EEG was acquired with a surface EEG system with 64 amplified electrodes (Biosemi ActiveThree system, BioSemi B.V., Netherlands) and preprocessing was performed using Brainvision analyser 2.2 (Brain Products, Munich, Germany) and Fieldtrip (Oostenveld et al., 2011). A spline interpolation was applied to artifacted channels (5.34±4.74% of channels) and an average reference was applied to the data. Blink and saccade artifacts were rejected using independent component analysis (ICA). Raw data were split into two seconds epochs from -1s to +1s centered on voice onset. Peak-to-peak amplitude was calculated for all epochs across all channels, and a trial rejection's threshold was calculated as the median + 1.96 standard deviation of the peak-to-peak amplitude distribution obtained (~114 μv). For each channel, epochs exceeding the rejection threshold were rejected from any future analysis (9.44±11.95% of trials rejected in average). For each participant and each channel, ERPs were obtained by averaging epochs for each emotional condition and each vibration type, as well as their conjunction (anger90-fear10—A_{90}—with anger vibration for example). Grand-average event-related potentials were then obtained by averaging the resulting ERPs across patients for each channel (final average number of trials per participant per condition 40.95+-6.08, see SI Table 7).

Event-related topographical clustering analysis

For each channel, the grand-average ERP for each condition was averaged into forty 25ms bins from voice onset to 1s post-stimulus onset. To increase the specificity of the topographical clustering algorithm to relevant temporal regions of interest, a baseline threshold was defined as ± 3 standard deviations of the grand average ERP for all conditions in the 1s baseline period. A database was created with, as lines the 40 timebins * 15 conditions (3 Vibration * 5 Emotional Voice), and as columns the 64 EEG channels for a final matrix size of 600 lines * 64 columns = 38'400 cells. If the average power in the bin exceeded the baseline threshold, the average power was reported in the table and if the average power in the bin was within the baseline confidence interval, a 0 was reported in the database instead.

Dimension reduction and topographical clustering was performed using Uniform Manifold Approximation and Projection (UMAP) for Dimension Reduction and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) for clustering both available in python (McInnes et al., 2017). Previous publications highlighted these methods as the most efficient for noisy high-dimensional data (Allaoui et al., 2020; McInnes et al., 2018; Weijler et al, 2022; Yang et al., 2021). UMAP algorithm was applied to the data with 40 neighbor parameters. The number of components was chosen using a measure of specificity—the average proportion of components significantly correlated (person's test, $p<.05$) with each parameter, and inclusivity—the total proportion of parameters significantly correlated to at least one component. Using this method, the optimal number of components (highest inclusivity *and* specificity) for this dataset was determined to be 6.

The UMAP components were entered into a HDBSCAN algorithm with Cityblock metric. The clustering yielded to 4 homogeneous topographical clusters with 0 non-attributed electrodes (see Figure 2**G** & Figure 3**K**).

ERP statistical analysis

For this analysis, ERPs for each channel, in each trial for each condition was averaged into twenty 50ms bins from stimulus onset to 1s post-stimulus onset. A database was created with the average power at each of the 20 timebins as for the 3 Vibration and 5 Emotional Voice conditions, for each 64 EEG channel with the 4 clusters and electrode placement (left, center, right) as factors. As power followed a linear distribution, a bayesian general linear mixed model approach was used to validate the result of our topographical clustering. For each timebin, data was entered into a GLMM to test the 3-way interaction between Cluster X Emotional Voice (5

levels, A_{90}, A_{60}, A_{50}, A_{40}, A_{10}) X Vibration (3 levels, Anger, Fear, No vibrations). Random factors included: the gender of the actor‘s voice—to account for gender differences and participant identity—as well as EEG channels.

Then, to test significant effects for each cluster and potential lateralization of these effects—that escaped topographical clustering analysis, data for each cluster in each timebin was entered into a Bayesian General Linear Mixed model testing the main effects, 2-way interaction between Emotional Voice (5 levels, A_{90}, A_{60}, A_{50}, A_{40}, A_{10}) X Vibration (3 levels, Anger, Fear, No vibrations), via the BayesFactor package in R. This approach, computing an index of the relative Bayesian likelihood of an Hypothesis 1 (H1) model compared with the null hypothesis (H0) model, supplements the classic GLMM approach, as the Bayes factor computed by this method allows H0 to be assessed. A Bayes factor > 3 constitutes proof of H1, while a Bayes factor < 1/3 is proof of H0. Strong proof for each is linked to a Bayes factor within the [10 30] or [1/10 1/30] boundaries, and extreme proof to a Bayes factor > 100 or < 1/100). Finally, the comparative nature of Bayesian GLMM makes it unsusceptible to multiple comparison. The gender of the actor’s voice was entered as random factor to account for gender differences and participant identity and as well as EEG channel. If a significant two-way Emotional Voice X Vibration interaction was found for a cluster in a timebin, contrast analysis was performed to test for each significant vibration differences between Emotional Voice conditions, and for each Emotional Voice condition, significant differences between Vibration conditions. Contrast analysis was performed using two-conditions Bayesian GLMM using the same random factor structure as the main analysis in R. Bayes factors being unsusceptible to multiple comparisons, no correction was used.

Significant results were considered when ERPs for at least one condition exceeded the average baseline ±3 standard deviations as threshold and for at least the whole 50ms bin.

Source reconstruction and analysis

Source reconstruction was performed using the MNE python from preprocessed, filtered, artifacts cleaned and epoched data (see EEG preprocessing section above). After forward solution computation using the fsaverage template (https://surfer.nmr.mgh.harvard.edu/fswiki/FsAverage), a LCMV spatial filter was computed using unit-noise-gain as weight normalization and applied to the epoched data. The average timecourse was computed for each epoch and each condition in each label of the HCP-MMP1 atlas (mean-flip parameter; Glasser et al., 2016). Then, as performed in the sensor domain (see

above), source-level ERP for each HCP-MMP1 atlas label, in each trial for each condition was averaged into twenty non-overlapping 50ms bins from stimulus onset to 1s post-stimulus onset. A database was created with the average power at each of the 20 timebins as for the 3 Vibration and 5 Emotional Voice conditions. Significant differences between conditions were investigated using two-conditions Bayesian GLMM with the Participant ID as random factor.

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), e28701. https://doi.org/10.1371/journal.pone.0028701

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

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

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), S295-S349. https://doi.org/10.1093/ons/opy261

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology, 70*(3), 614.

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT). *Emotion, 9*(5), 691.

Bänziger, T., & Scherer, K. R. (2010). Introducing the geneva multimodal emotion portrayal (gemep) corpus. *Blueprint for affective computing: A sourcebook*, 2010, 271-94.

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). *Social embodiment.Psychology of learning and motivation, 43*, 43-92.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological science, 21*(11), 1605-1610.

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.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution, 24*(3), 127-135.

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. *Univ 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

Caldwell, D. G., Lawther, S., & Wardle, A. (1996). Tactile perception and its application to the design of multi-modal cutaneous feedback systems. *In Proceedings of IEEE International Conference on Robotics and Automation, 4*, 3215-3221.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends in cognitive sciences, 11*(12), 535-543.

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.

Cespedes-Guevara, J., & Dibben, N. (2022). The role of embodied simulation and visual imagery in emotional contagion with music. *Music & Science, 5*, 20592043221093836.

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. Hillsdle.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain research*, *1242*, 126-135.

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.

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, *7*(2), 189-195. https://doi.org/10.1038/nn1176

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

Desmedt, J. E., & Bourguet, M. (1985). Color imaging of parietal and frontal somatosensory potential fields evoked by stimulation of median or posterior tibial nerve in man. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *62*(1), 1-17.

Dijkstra, K., Kaschak, M. P., & Zwaan, R. A. (2007). Body posture facilitates retrieval of autobiographical memories. *Cognition*, *102*(1), 139-149.

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

Dimberg, U., & Lundquist, L. O. (1990). Gender differences in facial reactions to facial expressions. *Biological psychology*, *30*(2), 151-159.

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.

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4(*3), 319-325.

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.

Gauffin, J., & Sundberg, J. (1974). Masking effect of one's own voice. *Speech Transmission Lab Q Prog Status Rep, 1,* 35-41.

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,* 2012.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature, 1–11. http://doi.org/10.1038/nature18933

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, November). P300 Somatosensory Validation of Vibrotactile Haptic Feedback for Upper Limb Prosthesis. In 2021 IEEE National Biomedical Engineering Conference (NBEC) (pp. 19-24). IEEE.

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.

Herbet, G., Lemaitre, A.-L., Moritz-Gasser, S., Cochereau, J., & Duffau, H. (2019). The antero-dorsal precuneal cortex supports specific aspects of bodily awareness. *Brain, 142*(8), 2207-2214. https://doi.org/10.1093/brain/awz179

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, April). Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation. In 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). (Vol. 1, pp. I-I). IEEE.

Kaschak, M. P., & Maner, J. K. (2009). Embodiment, evolution, and social cognition: An integrative framework. *European Journal of Social Psychology, 39*(7), 1236-1244.

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.

Kitamura, T., & Ohtani, K. (2015). Non-contact measurement of facial surface vibration patterns during singing by scanning laser Doppler vibrometer. *Frontiers in Psychology, 6*, 1682.

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?

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., Jessen, S., & Paulmann, S. (2009). Investigating the Multimodal Nature of Human *Communication. Journal of Psychophysiology, 23*(2), 63-76.

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.

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology, 21*(4), R143-R145.

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. V. (2021). emmeans: Estimated Marginal Means, aka Least-Squares Means. (Version X.X.X) [R package]. URL: https://cran.r-project.org/package=emmeans

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, R. (2013). Intuitiveness of vibrotactile speed regulation cues. *ACM Transactions on Applied Perception (TAP), 10*(4), 1-15.

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

Macerollo, A., Brown, M. J., Kilner, J. M., & Chen, R. (2018). Neurophysiological changes measured using somatosensory evoked potentials. *Trends in Neurosciences, 41*(5), 294-310.

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.

Maxwell, G. M., Cook, M. W., & Burr, R. (1985). The encoding and decoding of liking from behavioral cues in both auditory and visual channels. *Journal of Nonverbal Behavior, 9*(4), 239-263.

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. J. Open Source Softw., 2(11), 205.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

Mehling, W. E., Price, C., Daubenmier, J. J., Acree, M., Bartmess, E., & Stewart, A. (2012). The multidimensional assessment of interoceptive awareness (MAIA). *PloS one, 7*(11), e48230.

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.

Meteyard, L., Cuadrado, S. R., Bahrami, B., & Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex, 48*(7), 788-804.

Mills, K. A., Scholes, L., & Brown, A. (2022). Virtual Reality and Embodiment in Multimodal Meaning Making. *Written Communication, 07410883221083517.*

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.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current biology, 20*(8), 750-756.

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.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution, 4*(2), 133-142.

Neumann, R., & Strack, F. (2000). Approach and avoidance: The influence of proprioceptive and exteroceptive cues on encoding of affective information. *Journal of personality and social psychology, 79*(1), 39.

Niedenthal, P. M. 2007. “Embodying Emotion.” *Science, 316*(5827): 1002–5.

Niedenthal, P. M. (2008). Emotion concepts. *Handbook of emotions, 3*, 587-600.

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.

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

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*, 4453-4456.

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, 1-9.

Paquette, S., Rigoulot, S., Grunewald, K., & Lehmann, A. (2020). Temporal decoding of vocal and musical emotions: Same code, different timecourse?. *Brain Research, 1741,* 146887.

Parker G. Allan, Allan, P. G., Robert G. Briggs, Robert G. Briggs, Briggs, R. G., Andrew K. Conner, Conner, A. K., Christen M. O’Neal, O’Neal, C. M., Phillip A. Bonney, Bonney, P. A., B. David Maxwell, Maxwell, B. D., Cordell M Baker, Baker, C. M., Joshua D. Burks, Joshua D. Burks, Burks, J. D., Goksel Sali, … Sughrue, M. E. (2020). Parcellation-based tractographic modeling of the ventral attention network. *Journal of the Neurological Sciences, 408,* 116548-116548.

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. (2011). On the time course of vocal emotion recognition. *PLoS One, 6*(11).

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. https://doi.org/ 10.1163/156856897X00366

Petrushin, V. A. (2000). Emotion recognition in speech signal: experimental study, development, and application. *In Sixth international conference on spoken language processing.*

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.

Pollatos, O., Traut-Mattausch, E., Schroeder, H., & Schandry, R. (2007). Interoceptive awareness mediates the relationship between anxiety and the intensity of unpleasant feelings. *Journal of anxiety disorders, 21*(7), 931-943.

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.

Price, S., & Jewitt, C. (2013). A multimodal approach to examining'embodiment'in tangible learning environments. *In Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction* (pp. 43-50).

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

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

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci., 27*, 169-192.

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.

Salminen, K. (2015). Emotional Responses to Friction-based, Vibrotactile, and Thermal Stimuli.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *Neuroimage, 28*(4), 848-858.

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of cognitive neuroscience, 22*(3), 474-481.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The quarterly journal of experimental psychology, 63*(11), 2251-2272.

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.

Singer, T., & Klimecki, O. M. (2014). Empathy and compassion. *Current biology, 24*(18), R875-R878.

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.

Stevens, L. E., & Fiske, S. T. (1995). Motivation and cognition in social life: A social survival perspective. *Social cognition, 13*(3), 189-214.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *Journal of personality and social psychology, 54(*5), 768.

Sundberg, J. (1977). The acoustics of the singing voice. *Scientific American, 236*(3), 82-91.

Sundberg, J. (1983). Chest wall vibrations in singers. *Journal of Speech, Language, and Hearing Research, 26*(3), 329-340.

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. *Progress in Brain Research, 247*, 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, 1603*, 32-49.

Turella, L., Pierno, A. C., Tubaldi, F., & Castiello, U. (2009). Mirror neurons in humans: Consisting or confounding evidence?. *Brain and language, 108*(1), 10-21.

Tuuri, K., Eerola, T., & Pirhonen, A. (2010, September). Leaping across modalities: Speed regulation messages in audio and tactile domains. *In International Workshop on Haptic and Audio Interaction Design* (pp. 10-19). Springer, Berlin, Heidelberg.

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., & Dichmont, C. M. (2004). GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. *Fisheries research, 70*(2-3), 319-337.

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 2022*, 629-633.

Weijler, L., Kowarsch, F., Wödlinger, M., Reiter, M., Maurer-Granofszky, M., Schumich, A., & Dworzak, M. N. (2022). UMAP based anomaly detection for minimal residual disease quantification within acute myeloid leukemia. *Cancers, 14*(4), 898.

Wiens, S., Mezzacappa, E. S., & Katkin, E. S. (2000). Heartbeat detection and the experience of emotions. *Cognition & Emotion, 14*(3), 417-427.

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, p. 249-268). Elsevier. https://doi.org/10.1016/S0079-6123(06)56013-3

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.

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

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.

Yang, Y., Sun, H., Zhang, Y., Zhang, T., Gong, J., Wei, Y., ... & Yu, D. (2021). Dimensionality reduction by UMAP reinforces sample heterogeneity analysis in bulk transcriptomic data. *Cell reports, 36*(4).

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., & Montell Magnusson, A. K. (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.