*Cerebral Cortex*, 2022;32: 1978–1992

https://doi.org/10.1093/cercor/bhab329
Advance Access Publication Date: 15 October 2021
Original Article

ORIGINAL ARTICLE

# EEG Spatiotemporal Patterns Underlying Self-other Voice Discrimination

Giannina Rita Iannotti[1,2,†], Pavo Orepic[3,†], Denis Brunet[1,4], Thomas Koenig[5], Sixto Alcoba-Banqueri[3], Dorian F. A. Garin[2], Karl Schaller[2], Olaf Blanke[3] and Christoph M. Michel[1,4]

[1]Functional Brain Mapping Lab, Department of Fundamental Neurosciences, University of Geneva, 1202, Switzerland, [2]Department of Neurosurgery, University Hospitals of Geneva and Faculty of Medicine, University of Geneva, 1205, Switzerland, [3]Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), 1202, Switzerland, [4]CIBM Center for Biomedical Imaging, Lausanne and Geneva, 1015, Switzerland and [5]Translational Research Center, University Hospital of Psychiatry and Psychotherapy, University of Bern, Bern 3000, Switzerland

Address correspondence to Prof. Christoph M. Michel, Campus Biotech, Neuroscience Department, 9 chemin des Mines, 1211 Genève 20, Switzerland. Email: christoph.michel@unige.ch
[†]Giannina Rita Iannotti and Pavo Orepic have contributed equally to the conception of study, analyses, and writing of the manuscript

## Abstract

There is growing evidence showing that the representation of the human "self" recruits special systems across different functions and modalities. Compared to self-face and self-body representations, few studies have investigated neural underpinnings specific to self-voice. Moreover, self-voice stimuli in those studies were consistently presented through air and lacking bone conduction, rendering the sound of self-voice stimuli different to the self-voice heard during natural speech. Here, we combined psychophysics, voice-morphing technology, and high-density EEG in order to identify the spatiotemporal patterns underlying self-other voice discrimination (SOVD) in a population of 26 healthy participants, both with air- and bone-conducted stimuli. We identified a self-voice-specific EEG topographic map occurring around 345 ms post-stimulus and activating a network involving insula, cingulate cortex, and medial temporal lobe structures. Occurrence of this map was modulated both with SOVD task performance and bone conduction. Specifically, the better participants performed at SOVD task, the less frequently they activated this network. In addition, the same network was recruited less frequently with bone conduction, which, accordingly, increased the SOVD task performance. This work could have an important clinical impact. Indeed, it reveals neural correlates of SOVD impairments, believed to account for auditory-verbal hallucinations, a common and highly distressing psychiatric symptom.

**Key words:** bone conduction, high-density EEG, insula, limbic system, self-other voice discrimination

## Introduction

Human voice is one of the most important sounds we experience on a daily basis. Integrating different types of information (gender, emotional content, social setting, accents, and so on), it plays an important role in individuals' identification (Belin et al. 2004; McGettigan et al. 2013; Zarate et al. 2015). Neuroimaging studies demonstrate that, already in infancy, humans exhibit a selective activation of brain areas, predominantly in the right hemisphere (namely, right middle and superior temporal gyri) for voices compared to non-vocal stimuli (Grossmann et al. 2010). Although face is considered the main carrier of human identity (Blank et al. 2014), people can unequivocally be identified also through their voice (Belin et al. 2004). However, whether the process of identifying oneself is equivalent to identifying others is still a question of debate (Gillihan and Farah 2005).

The self could be defined as an entity that is distinct from the environment and other humans to which certain mental events and actions are ascribed (Kircher and David 2003). The special nature of the self is thought to arise from multisensory integration of bodily signals (Blanke 2012; Blanke et al. 2015; Tsakiris 2017; Park and Blanke 2019) and sensorimotor congruency resulting from the interactions of the body with the environment (Blakemore et al. 2000; Tsakiris and Haggard 2005; Kannape and Blanke 2012; Braun et al. 2018). Accordingly, self-specificity has been reported in many neuroimaging studies that investigated self-referential processes across different functional domains (e.g., emotional, spatial, memory) (Northoff et al. 2006). Compared to the extensive work done on self-face representation (for an overview see Uddin et al. 2005), self-voice representation has been investigated to a surprisingly lesser extent. However, acquiring a better understanding of neural mechanisms underlying self-voice perception is of utmost importance, as erroneous self-voice misattribution is thought to account for auditory-verbal hallucinations (Frith and Done 1989; Frith 1992; Ford and Mathalon 2005; Shergill et al. 2014), the most common hallucination in schizophrenia, associated with high degrees of distress in the affected population (Harkavy-Friedman et al. 2003).

Electroencephalographic (EEG) investigations of self-voice specificity have been confined to oddball paradigms and mostly focused on the analysis of the event-related potential (ERP) at predefined electrodes (e.g., located on central and parietal areas) and time windows following self- and other deviant vocal stimuli. The majority of these studies emphasizes the role of the late auditory ERP components, namely P3 (~300 ms from stimulus onset) and N4 (~400 ms from stimulus onset). For example, the amplitude of the P3a component in a passive oddball paradigm was observed to be lower for self-voice, compared to unfamiliar (Graux et al. 2013) and familiar voices (Graux et al. 2015), suggesting less pre-attentional processes being involved in the discrimination of one's own voice. (Conde et al. 2016) confirmed this finding and narrowed the P3 reduction down to experimental stimuli consisting of simple vocalizations, compared to words. However, in active oddball paradigms requiring attentional processes, the P3 amplitude increased for self-voices compared to other voices, indicating that one's own voice has a greater affective salience than an unfamiliar voice (Conde et al. 2015, 2018). However, Liu et al. (2019) failed to replicate the differences in P3 component between self- and other voices, while showing an increased parietal N400 amplitude for self-voices when uttering other names but not the own name.

It is important to note that in the above-described ERP studies, a few electrodes in pre-selected time windows were analyzed that differed between the studies, making a comparison difficult. Additionally, such approach does not allow for a comprehensive analysis of the spatio-temporal modulation of the ERPs to self- and other voices that might be seen at other electrodes or latencies than those pre-selected (Murray et al. 2008).

Concerning the brain areas associated to the self-voice, in a PET study of Nakamura et al. (2001), a contrast between self-voice and familiar voice activation revealed a peak in the right inferior frontal sulcus and parainsular cortex. Similarly, right inferior frontal gyrus produced greater signal to self- compared to familiar voice in a study of Kaplan et al. (2008). Allen et al. (2005) found that, when contrasted to unfamiliar voice, self-voice was associated to increased activation in left inferior frontal and right anterior cingulate cortex. Clinical investigations highlight the involvement of the right hemisphere in the self-other voice discrimination showing that patients with right lesions were worse in discriminating their (own) voice compared to left damaged patients and healthy participants (Candini et al. 2018). Together, these findings suggest that self-voice is represented differently from other voices both in neuronal activation and in the brain areas implicated. Importantly, however, no study related the behavioral performance (e.g., ability to recognize own voice or to discriminate self-voice from other voices) to the neural activations.

Behavioral investigations that compared recognition of self-versus-other voices mostly showed lower accuracy or slower response times for self-voice compared to other-voice stimuli (Gur and Sackeim 1979; Shuster 1998; Allen et al. 2005; Rosa et al. 2008; Hughes and Nicholson 2010; Conde et al. 2015; Schuerman et al. 2015), indicating an inability to recognize self-voice as well as other voices. This arguably reflects the lack of bone conduction in self-voice stimuli presented in corresponding studies, which is present while speaking (Békésy 1949; Reinfeldt et al. 2010). Namely, when we speak, we hear our voice also through bone conduction, which applies a physical transformation to the sound of our voice (Stenfelt 2016) and, besides auditory, often involves somatosensory (Tremblay et al. 2003; Ito et al. 2009) and vestibular (Todd et al. 2000; Emami et al. 2012) processing. In our recent work (Orepic 2020), we showed that multimodal presentation of self-voice stimuli facilitates self-voice recognition. Specifically, combining psychophysics with voice-morphing technology (Kawahara et al. 2013), we designed a sensitive self-other voice discrimination (SOVD) task that enables pinpointing of perceptual specificities underlying SOVD, and observed a better SOVD task performance with the stimuli presented through a commercial bone-conduction headset compared to traditional air conduction media (e.g., loudspeakers).

Here, we build up on those findings by investigating brain mechanism underlying SOVD both through air and bone conduction, with a high-density EEG setup, which allows to localize neuronal sources with high specificity, (Brodbeck et al. 2011; Mégevand et al. 2014; Michel and He 2019; Seeber et al. 2019). We recorded evoked responses of healthy individuals hearing ambiguous self-other voice morphs (e.g., a voice morph could be created such that it contains 40% of self-voice and 60% of stranger's voice) and related neural activity with the ability to determine a dominant voice in such voice morphs.

First of all, we investigated the EEG characteristics during SOVD task of the principal auditory ERP exogenous components P1 (~50 ms after stimulus onset), N1 (~100 ms after stimulus onset), P2 (~200 ms after stimulus onset), and for the late period corresponding to the complex P3/N4 (Winkler et al. 2015). The

late period of ERP demonstrated to play an important role in the discrimination between different voices (Titova and Näätänen 2001; Beauchemin et al. 2006; Graux et al. 2013, 2015; Conde et al. 2016; Liu et al. 2019).

For each component, we conducted an electrode-wise waveform analysis and examined the differences in global field power (GFP) as a measure of global synchronization (Lehmann and Skrandies 1980). The analysis was performed on all electrodes and not only on pre-selected channels to avoid a selection bias (Murray et al. 2008). While this conventional ERP analysis gives information about the "strength" of stimulus-induced activity, it does not provide information about potential differences in the topography of the evoked potential field induced by the different stimulus conditions. Differences in the topography would directly indicate modulation of the generating sources in the brain (Michel et al. 2009). Therefore, to study differences in the "distribution" of the scalp potential field, we performed the so-called topographic analysis of the variance, or TANOVA (Karniski et al. 1994; Koenig and Melie-García 2010; Habermann et al. 2018). In order to have a comprehensive description of the electrophysiological processes involved in our task, we further took advantages of a more complete spatio-temporal approach, known as EEG microstate segmentation (Brandeis and Lehmann 1986; Murray et al. 2008; Brunet et al. 2011; Schiller et al. 2016). This approach allows identifying the optimal set of EEG topographies that explain the averaged ERPs, both in space and in time. EEG microstate segmentation, indeed, relies on the observation that EEG topographic maps (both spontaneous EEG and ERPs) are not randomly distributed but exhibit a defined temporal sequence, which reflect the functionally stable (in time) associated brain states (Michel et al. 2001; Michel and Murray 2012; Michel and Koenig 2018). Importantly, this approach is completely reference-independent and thus renders statistically unambiguous results (Michel and Murray 2012).

We investigated dependency of EEG parameters describing each map obtained from the microstate segmentation (i.e., first onset, mean duration, occurrence, global explained variance) on hearing a varying degree of self-voice and its relationship to psychophysically quantified SOVD performance. Finally, we identified underlying neural sources accounting for the process of discriminating own from a stranger's voice. Based on previous works, we predicted that self-versus-other voice activates, in the late ERP component (after 300 ms) (Graux et al. 2015; Conde et al. 2016), limbic and cortical midline structures (Nakamura et al. 2001; Allen et al. 2005; Northoff et al. 2006) whose temporal characteristics might reflect in participant's task performance.

## Materials and Methods

### Participants

This study involved 26 healthy participants, 14 female, mean age ± SD: 37.4 ± 14.7 years old. All participants were right handed, reported no hearing deficits and no history of psychiatric or neurological disorders. They were instructed on the conduct of the task, gave informed consent in accordance with institutional guidelines (the Declaration of Helsinki and reference to the protocol PB_2016-01635, amendment 3 approved by the Commission Cantonal d' Ethique de la Recherche de Geneva) and received monetary compensation (CHF 20/h).

### Self-other Voice Discrimination Task

Prior to participating in the studies, participants' voices were recorded while vocalizing phoneme/a/for approximately 1 to 2 s (Zoom H6 Handy recorder). Each recording was standardized for average intensity (−12 dBFS) and duration (500 ms) and cleaned from background noise (Audacity software). Short vocalizations were chosen to control for other linguistic and paralinguistic accounts for speaker identification, such as accent or prosody, constricting the identification process to simple acoustical properties of the voice. Previous work has shown that short vocalizations suffice for speaker identification (Zarate et al. 2015). Each participant's preprocessed voice was mixed with a target voice of a gender-matched unfamiliar person in order to generate voice morphs spanning a voice identity continuum between the two voices by using TANDEM-STRAIGHT (Kawahara et al. 2013), a voice-morphing software package running in MATLAB. It decomposes recordings of two speakers into mutually independent acoustic parameters that can be precisely controlled, interpolated and extrapolated between the speakers. As a result, a new voice can be created, such that it contains a desired ratio of each speaker's voice (e.g., a voice containing 30% speaker A's and 70% of speaker B's voice). Six voice ratios (% self-voice: 15, 30, 45, 55, 70, 85) were chosen based on our previous work (Orepic 2020; Orepic et al. 2021).

During the experiment, participants were seated in comfortable armchair, in a silent and moderately lit room, and were asked to perform the task with eyes open, facing a white wall. Voice morphs were presented to participants either through bone-conducting headphones (Aftershokz Sports Titanium) or through laptop loudspeakers (air conduction). The order of air- and bone-conduction blocks was counterbalanced across participants. The study contained 10 experimental blocks, 5 of which were conducted consecutively with the same sound conduction type (air, bone). Each block contained 60 trials of morphs randomly selected from the 6 self-voice ratios (Fig. 1A), for a total of 50 trials per each voice-morph and sound conduction type. Intertrial intervals jittered between 1 and 1.5 s to avoid predictability of stimulus onset. For each voice morph, participants were asked to indicate whether the voice they heard resembled their own or someone else's voice by clicking on a mouse button. Importantly, participants were not presented with their unmorphed voice recordings prior to task execution, assuring that they performed the task by comparing voice norphs with the internal self-voice representation (Orepic 2020). The experimental paradigm was created in MATLAB 2017b with Psychtoolbox library (Kleiner et al. 2007).

During the task, electrophysiological (EEG) data were continuously recorded with a sampling rate of 1000 Hz using a 256-electrodes Hydrocel cap (Megstim, Electrical Geodesics Inc.), referenced to the vertex (Cz). The impedance of electrodes was monitored carefully and kept below 40 k$\Omega$. For the reference electrode, the impedance was kept below 10 k$\Omega$. Bone-conducting headphones were installed on participants' head underneath the EEG cap, by avoiding the overlap with any (namely temporal) electrodes. An illustration of the experimental paradigm and setup is given in Figure 1.

### Behavioral Performance

Performance in self-other voice discrimination task was analyzed with mixed-effects binomial regressions with Accuracy as dependent variable and two fixed effects—Conduction (Air, Bone) and Voice Morph (15, 30, 45, 55, 70, 85)—related with an interaction term. The Accuracy variable indicated the percentage to which participants correctly identified the dominant voice in the presented morph. The model further contained the polynomial expansion of the Voice Morph variable to level

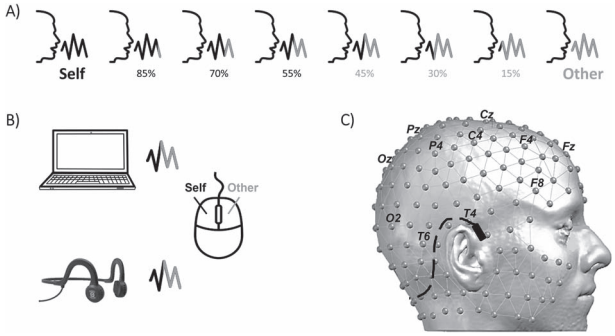


**Figure 1.** SOVD task. (*A*) Stimuli. Six self-other voice morphs between participant's voice Self, black and the voice of a gender-matched unfamiliar person (Other, gray) were randomly presented 50 times throughout the experiment. (*B*) Task. Voice morphs were presented either through air (laptop, above) or bone conduction (commercial headset, below). In every trial, participants responded whether the morph they hear resembles more to their or to someone else's voice by clicking on the corresponding mouse button. (*C*) EEG setup. Bone conduction headphones (black) placed under a high-density EEG cap (light gray spheres and connections) formed by 256 electrodes organized as an extension of the standard clinical 10–20 setup (electrode names indicated in black).

2. Random effects included a by-participant random intercept, whereas by-participant random slopes for the fixed effects were added following model selection based on maximum likelihood. Trials with reaction times greater or smaller than two interquartile ranges from the median for each subject were considered as outliers and excluded. Equivalent mixed-effects regression was run for Response Time as a dependent variable.

## EEG Data Pre-Processing And Selection of Epochs

EEG data were preprocessed with the free academic software CARTOOL (Brunet et al. 2011). The EEG was reduced to 204 channels, by eliminating the electrodes covering the cheeks and the lowest neck surfaces because they were often not attached to participants' skin (due to individual anatomical configurations) and contained motion artifacts (e.g., spontaneous chewing). First, the data were downsampled to 500 Hz and band-pass filtered between 1 and 40 Hz using non-causal Butterworth filters and a Notch filter of 50 Hz to eliminate (environmental) 50-Hz noise. Then, independent component analysis (ICA) was applied to remove eye-movement (eye blinks and saccades) and ballistocardiogram artifacts (Jung et al. 2000). A MATLAB script based on the "EEGlab runica" function that allows for component inspection both across time and scalp topographies was applied, while period of motion-artifact and noisy channels were temporarily excluded. After ICA artifact removal, noisy electrodes were interpolated (3-D spherical spline, Perrin et al. 1989) and a spatial filter was applied (Michel and Brunet 2019). In order to perform further analyses on the event-related potentials (ERPs), for each participant and for each type of Voice Morph, epochs between −50 ms pre-stimulus and 500 ms post-stimulus were visually inspected. Epochs characterized by residual artifacts, mostly due to motion were excluded.

## EEG Data Analysis

For each participant and each Conduction (Air, Bone), we grouped the ERPs belonging to each end of the self-other voice continuum, to increase the number of epochs from 50 to 100 and therefore the signal-to-noise ratio of the ERPs. Specifically, we averaged *Self-dominant* (containing 85% and 70% self-voice) and *Other-dominant* voice morphs (15% and 30% self-voice). This resulted in four group-averaged ERPs: Bone-Other, Bone-Self, Air-Other, Air-Self.

At the level of the scalp, the EEG differences associated to the SOVD task were characterized with four levels of analysis: 1) electrode-wise ANOVA, 2) GFP ANOVA, 3) TANOVA, and 4) microstate segmentation. Additionally, we interpreted our results in terms of brain regions by considering the signal differences in the inverse space.

## ERP Waveform Analysis And Global Topographic Measurement

In a first step, we conducted an exploratory waveform analysis on all electrodes (204) and on the GFP. To this aim, we defined the auditory ERP component peaks in the conventional time windows (Picton 2010; Winkler et al. 2015), adjusted by visually inspecting the four group-averaged ERPs: P1 = [25–75] ms; N2 = [75–175] ms; P2 = [175–270] ms; the late complex P3/N4 = [270–500] ms (Fig. 2). For each subject and for each ERP component, we collected the amplitude of the peak (with the appropriate positive or negative sign) and its latency, across all electrodes and for the GFP. Statistically significant differences for each peak and latency were, therefore, assessed by an electrode-wise and GFP two-way-ANOVA, with factors Conduction and Dominant Voice and respective levels of Bone/Air, Other/Self. A statistical threshold of $P < 0.05$ for the GFP and Bonferroni corrected for multiple comparisons ($P < 0.0002$) for the electrode-wise analyses were considered.

In a second step, we investigated the differences in the scalp potential distribution between the four experimental conditions by computing a TANOVA for each time point with the same factors and associated levels of the ERPs ([0–500] ms). The TANOVA is a non-parametric randomization test of global dissimilarity (GD) (Habermann et al. 2018). GD is an amplitude-independent measure which assesses the difference between the scalp topographies normalized by dividing the potential values at each electrode by the GFP. The TANOVA analysis was performed with the Ragu software (Koenig et al. 2011); 1000 randomizations were selected to assure a significant threshold of $P < 0.05$.

## Spatio-Temporal ERP Dynamics

As a third step, and in order to have a comprehensive information on the spatial and temporal EEG patterns associated to the SOVD task, we used the EEG segmentation approach (Fig. 6*A*). In details, we considered the group-averaged ERPs: Bone-Other, Bone-Self, Air-Other, Air-Self, and we applied a *k*-means cluster analysis (300 randomizations, 1–30 clusters) (Pascual-Marqui et al. 1995; Murray et al. 2008; Brunet et al. 2011). The values of a meta-criterion, that is a combination of seven independent optimization criteria (for details see Bréchet et al. 2019), allowed to define the optimal number of EEG topographic maps (or cluster maps) that best described the four group-averaged ERPs. To assess the statistical difference between the four experimental conditions, we projected back or "fitted back" the obtained cluster maps on the ERPs of each subject, for each dominant morph and each Conduction. The back-fitting of the cluster maps was applied in three non-overlapping time windows defined around the typical ERP components (N1, P2, and the late complex P3/N4) by taking into account the temporal subdivision revealed by the

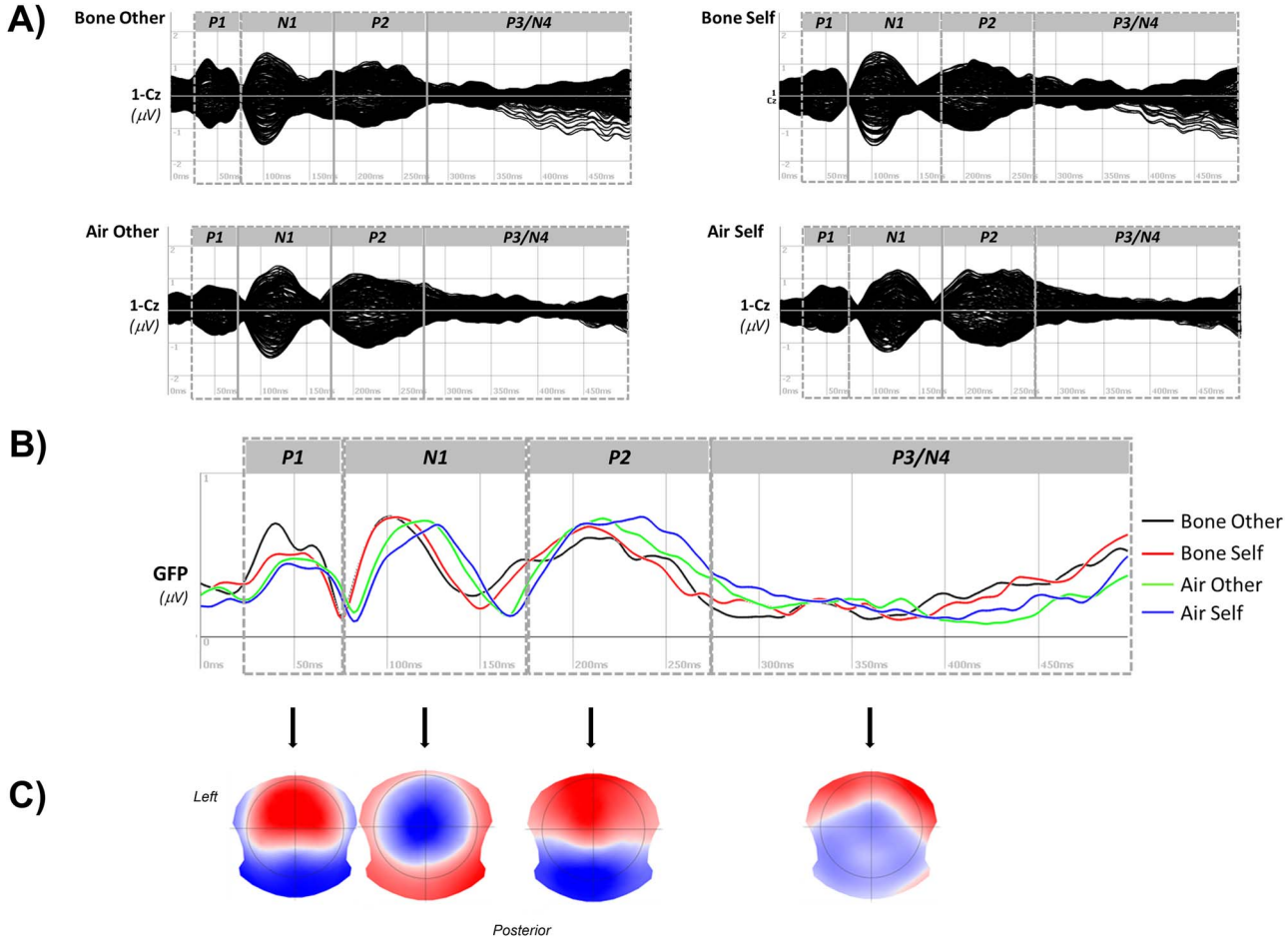


**Figure 2.** ERP components analysis. Butterfly plot of the ERP waveform of all electrodes overlapped (*A*) and of the GFP (*B*) for each of the four experimental conditions (Bone-Other, Bone-Self, Air-Other Air-Self) are represented. The gray dashed boxes in (*A*) and (*B*) indicate the time-windows selected for the auditory ERP components (P1, N1, P2, P3/N4). Peak and latency of the ERP components were defined in the associated time windows and then used for the electrode-wise and GFP ANOVA analyses. In (*C*) the averaged-topographies (across subjects, conditions, and time-window) associated to each ERP component are shown.

back-fitting of the maps to the four group-averaged data: 1) [20–140] ms for maps 1 and 2; 2) [140–270] ms for map 3; and 3) [270–500] ms for maps 4, 5, 6, and 7.

The back-fitting approach evaluates the spatial correlation between the cluster maps and the single-subject ERP at each time point, by labeling the latter with the cluster map that shows the highest spatial correlation after power normalizing each map. In our case, we used a competitive attribution of the maps, thereby taking into account the polarity of the maps. No label was assigned to a time point if the correlation between the ERP signal and any of all cluster maps was $<0.25$. In general, the output of the back-fitting consists of several parameters characterizing the time and goodness of fit of each cluster map, for each subject and each experimental condition. In particular, we investigated, the temporal parameters first onset (FO), occurrence (Occ) and the mean duration (MD), as well as the global explained variance (GEV). The description of these parameters is reported in Table 1.

A linear mixed-effects regression with the dependent variable parameter value (FO, MD, Occ, GEV) was performed with two fixed effects: Conduction (Air, Bone) and Dominant Voice (Self, Other), related with an interaction term, and with by-participant random effects. Mixed-effects regression was used to account for possible missing values in the data, as for some participants, back-fitting procedure could omit the attribution of certain maps. For the maps that indicated significant interactions involving the effect of Parameter, we ran additional mixed-effects regressions separately for each Parameter, directly assessing the effects of Conduction and Dominant Voice on Parameter Value of the corresponding parameter.

## EEG-Behavior Relationship

We further investigated the relationship between SOVD task performance and the significant fitting parameters (FO, Occ, MD, GEV). As task performance, or behavioral variables, we considered the subject's average Response Time and the Accuracy, i.e., the percentage of correct answers. We performed a linear mixed-effects regression with the dependent variable Parameter value (FO, Occ, MD, GEV) and with three fixed effects—Dominant Voice, Conduction, and the behavioral variable (respectively, Accuracy, and Response Time).

All the statistical tests (electrode-wise and GFP ANOVA for the peak and latency of the ERP components; link EEG-behavior) were performed with R (R Core Team 2020), using the lme4 (Bates et al. 2015), and lmerTest (Kuznetsova et al. 2018) packages. The

**Table 1** Back-fitting parameters

| Parameter | Description | Unit |
|---|---|---|
| First occurrence | The first time point a given map yields the highest spatial correlation value | ms |
| Occurrence | The total number of data points in which a given map yields the highest spatial correlation value | ms |
| Mean duration | Mean number of consecutive data points in which a given map yields the highest spatial correlation value | ms |
| GEV | Global Explained Variance, i.e., how well a given map "explains" the data | Scalar number in the range [0, 1] |

Notes: The definition and description of the four parameters selected from the back-fitting to assess the statistical difference between the experimental conditions (Bone/Air; Self-Other) are reported in the table.

results were illustrated using sjplot (Lüdecke 2021) and ggplot2 (Wickham 2016) packages.

### Source Localization of Cluster Maps

To estimate the brain networks underlying the topographies of the scalp EEG resulting from the clustering procedure, a distributed linear inverse solution matrix (needed to project the ERPs data from the scalp to the brain space) was computed.

For the forward model, we used the asymmetrical, non-linear Montreal Neurological Institute (MNI) template including the cerebellum, as head model with consideration of skull thickness (Locally Spherical Model with Anatomical Constraints (LSMAC); for a review see Michel and Brunet 2019) and a grid of around 6000 solution points distributed equally in the gray matter. The final inverse matrix was computed using the low-resolution brain electromagnetic tomography (LORETA) method (Pascual-Marqui et al. 2002).

Similarly to Bréchet et al. (2019), the cluster maps from the group-averaged ERPs segmentation were back-projected to each subjects' ERPs with a winner-takes-all method. For a given subject and for each map, data points assigned to a given map were concatenated in a separate cluster. The vectorial source localization of all the obtained clusters were then computed using the inverse matrix described above. Within each subject, the resulting inverse solutions were standardized across all solution points (Michel and Brunet 2019) using the whole subject's dataset. Finally, the vectorial mean across all subjects' standardized clusters were computed, then normalized between [0 1].

## Results

Two participants were excluded from the analysis due to the poor quality of the EEG and six because they had a task performance lower than 50% independent of conduction type (bone, air). Similar number of participants was unable to perform the same task in our previous study (Orepic 2020), where we introduced the task and compared it to other self-related tasks in three different behavioral experiments, thereby discussing the factors accounting for its high difficulty. Our sample resulted therefore in 17 subjects (9 females, mean age ± SD: 37.3 ± 15.1 years old).

### Behavioral Results

A mixed-effects binomial regression in the self-other discrimination with Accuracy as dependent variable revealed a main effect of the polynomial expansion of the Voice Morph variable (estimate = 72.85, $Z = 17.18$, $P < 0.001$), indicating a u-shaped task performance with the increase of self-voice present in the morph. It further revealed a main effect of Conduction (estimate = −0.18, $Z = 2.93$, $P = 0.003$), reflected as higher accuracy for bone conduction. A two-way interaction between the effects of Conduction and Voice Morph was also significant (estimate = 19.13, $Z = 3.05$, $P = 0.002$), observed as a difference in performance for bone conduction across Voice Morphs. Post hoc investigation of the interaction revealed that higher accuracy for bone conduction only in the morphs with least self-voice present or other voice (15%) (Estimate = 1.22, $Z = 2.57$, $P = 0.01$). Difference in accuracy between other Voice Morphs was not significant (all $P > 0.05$). Together, these findings indicated that participants discriminated own from a stranger's voice better with bone compared to air conduction, and this was most prominent for other-dominant morphs (Fig. 3, left).

Linear mixed-effects regression for response times as dependent variable similarly revealed the main effect of the polynomial expansion of Voice Morph variable (estimate = −6.05, $t(9535) = -11.93$, $P < 0.001$), indicating a reversed u-shape of response times dependent on self-voice ratio. There was the main effect of Conduction (estimate = −0.07, $t(9535) = -9.31$, $P < 0.001$), revealing faster response times for bone conduction. Interaction between the effects of Voice Morph and Conduction was not significant (estimate = −0.2, $t(9535) = 0.28$, $P = 0.783$). Response times are shown at the right of Figure 3.

### EEG Results

The visual inspection of ERPs resulted in an average of 45/50 (range [39/50, 50/50]) selected epochs, across subjects and per condition.

### Auditory ERPs Components And Topographical Differences

The electrode-wise ANOVA conducted for each peak amplitude and latency of the auditory ERP components revealed a significant difference ($P < 0.0002$) only for the latency of N1 ([75–175] ms). In more details, only a significant main effect of Conduction was observed for two right centro-parietal electrodes (131 and 155, close to CPz, C4, CP2, C2, Fig. 4).

No significant main effects for Conduction and Dominant Voice and no significant interaction between them were observed for the GFP ($P > 0.05$).

The TANOVA across the whole ERP period ([0–500] ms) revealed a significant ($P = 0.02$) main effect of the Conduction

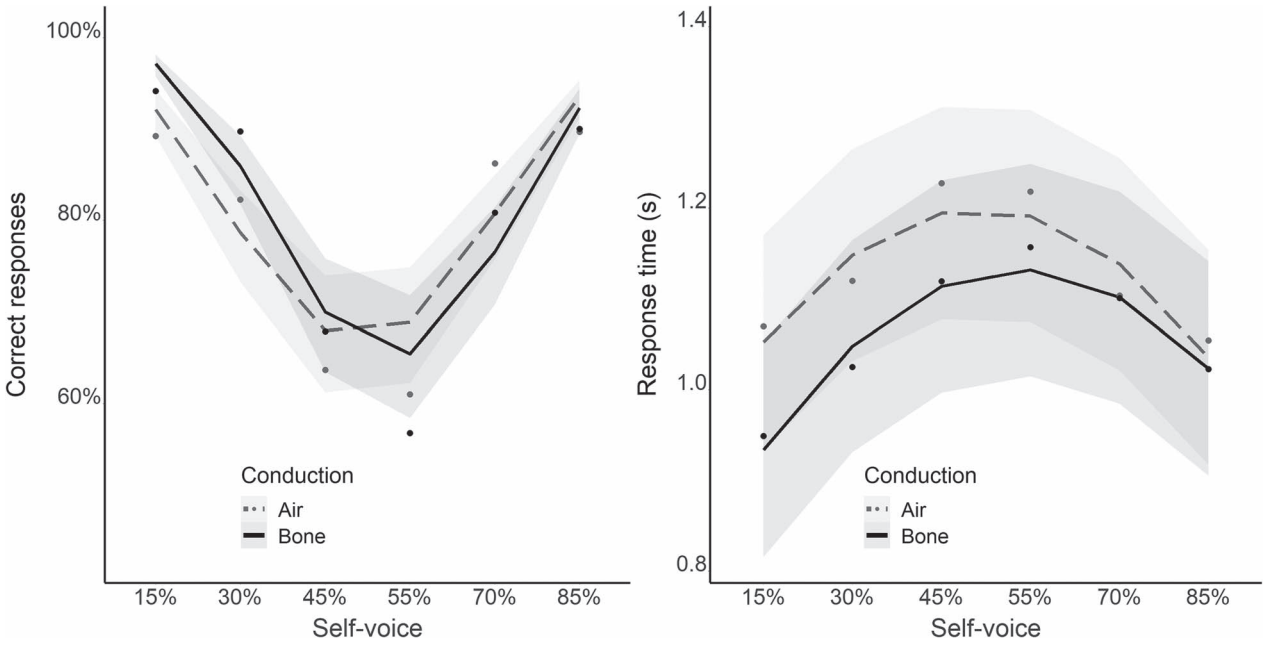


**Figure 3.** Behavioral results indicating the effects of the two forms of sound conduction (air, bone) on accuracy (left) and response times (right) in self-other voice discrimination task. The abscissa of both plots indicates the percentage of the self-voice present in a Voice Morph. The shaded areas around each curve represent the 95% confidence intervals. Accuracy was higher and response times were faster for bone conduction.

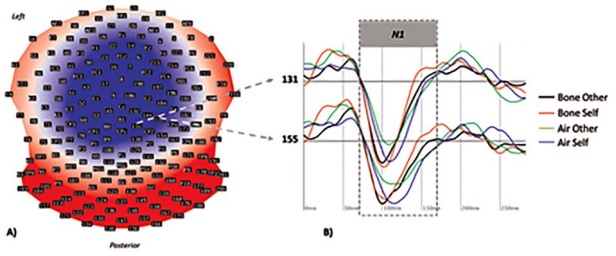


**Figure 4.** Electrode-wise ANOVA results. The ANOVA conducted for each peak and each latency of the auditory ERPs components, corrected for multiple comparisons, showed a main effect of Conduction ($P < 0.0002$) for the latency of N1 component over the electrodes 113 and 155. They are located over the right centro-parietal area (*A*) and their waveform is shown in (*B*), for each of the four experimental conditions.

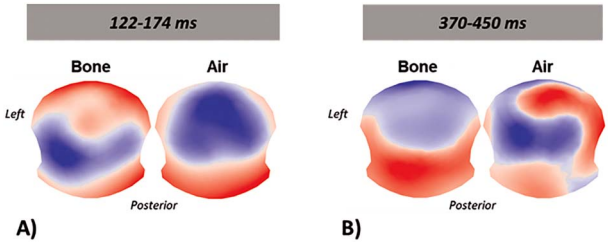


**Figure 5.** Topographical differences. The TANOVA across the period [0–500] ms revealed a significant main effect of the Conduction ($P < 0.05$) for N1 (*B*) and the late complex P3/N4 (*A*). The topographies associated to the Bone and Air and averaged in the periods of significance (gray dashed boxes) are shown (*B*).

for N1 and for the complex P3/N4, specifically in the periods: [122–174] ms and [370–450] ms. No main effect of the Dominant Voice ($P = 0.03$) and no interaction effect ($P = 0.8$) were observed. The difference in the topographies associated to Bone and Air in the first period (Fig. 5A) reflected a shift in latency of N1, which peaked later in Air than Bone. In the second period ([370–450] ms), differently to Bone, Air condition had more pronounced negativity over the right parietal regions which could be interpreted as a prolonged effect of P3 (in literature observed more over parietal sites) for Air, with respect to Bone (Fig. 5B).

## EEG Segmentation Results

The meta-criterion associated to the K-means clustering applied to the four group-averaged ERPs (Bone-Other, Bone-Self, Air-Other, Air-Self) established as optimal segmentation the set of seven cluster scalp EEG topographies, or maps that were dominating in seven different time segments (Fig. 6B). These set explained 93% of all ERP data. The investigation of missing values after the back-fitting procedure revealed a low number of missing maps in our cohort: 24%, 18%, and 6% and 0% for Bone-Other, Bone-Self, and both Air-Other and Air-Self, respectively.

## Statistical Differences of the Cluster EEG Topographies

Among the seven topographic maps derived from the EEG segmentation (Fig. 6B), we observed significant interactions of effects Dominant Voice and Conduction with the effect of Parameter only for map 4. This map appeared in the range [$292 \pm 46$; $380 \pm 68$] ms (average and standard deviation of first onset and last offset, across subjects and all the experimental conditions). Specifically, for map 4, there was a significant interaction between Parameter and Dominant Voice ($F(3, 227.04) = 4.51$, $P = 0.004$) and between Parameter and Conduction ($F(3, 227.62) = 5.47$, $P = 0.001$). A three-way interaction between the effects of Parameter, Dominant Voice, and Conduction was not significant ($F(3, 226.93) = 0.91$, $P = 0.437$).

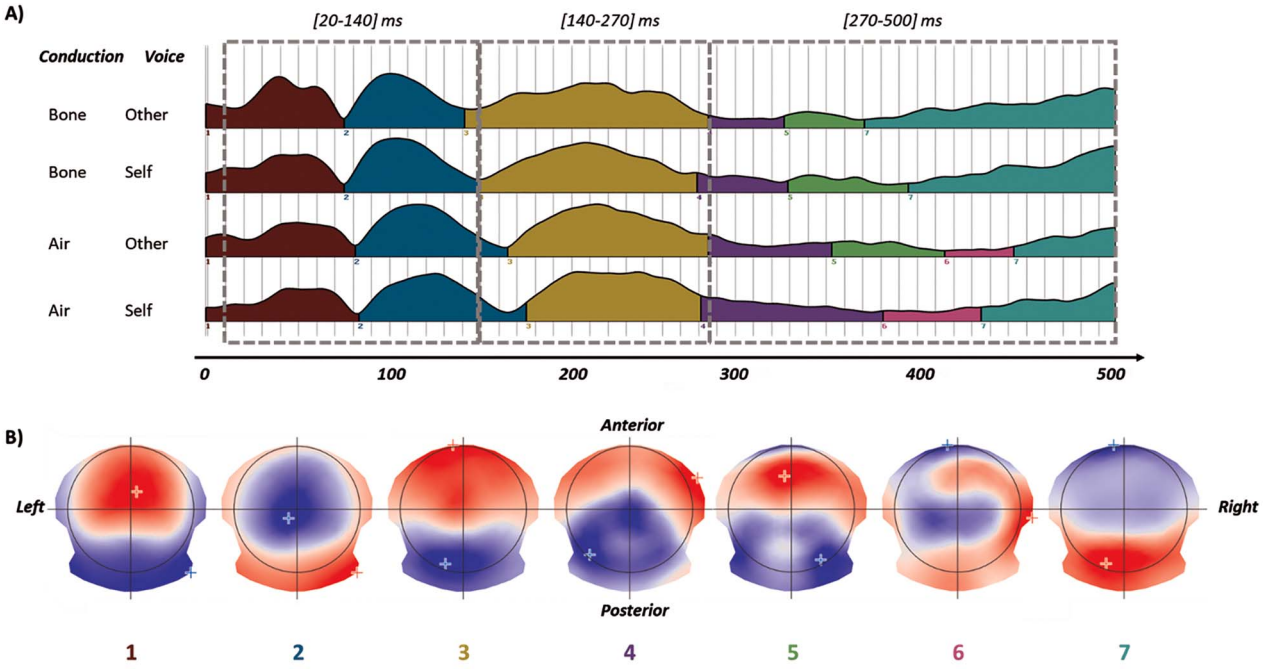


**Figure 6.** Results of the group-average ERP segmentation. (*A*) Different colors indicate different segments marked under the Global Field Power curves extracted by the *K*-means clustering on the group-averaged ERPs corresponding to the Other-dominant (15–30% self-voice) and Self-dominant (70–85%) Voice Morphs and to the two types of sound conduction (Bone, Air). The gray dashed boxes indicate the three time-windows considered for the back-fitting procedure. (*B*) Topographic maps associated to each segment. The "+" symbol indicates the position of the electrodes exhibiting the highest (positive, red) and the lower (negative, blue) amplitudes of the scalp voltage potential.

Thus, for each Parameter, we conducted a separate linear mixed-effects regression with Parameter Value as dependent variable, and Conduction and Dominant Voice as fixed effects. From all parameters, mixed-effects regression revealed significant effects only for the parameter Occurrence: there were both the main effect of Dominant Voice ($F(1, 51) = 6.64$, $P = 0.013$), indicating higher occurrence of map 4 for self-dominant morphs (Fig. 7A), and the main effect of Conduction ($F(1, 51) = 7.74$, $P = 0.01$), indicating higher occurrence during air conduction (Fig. 7B). There was no significant interaction between the effects of Dominant Voice and Conduction ($F(1, 51) = 1.7$, $P = 0.198$). For exploratory purposes, however, we additionally conducted the pairwise post hoc *t*-tests for all the combinations of the values of Conduction and Dominant Voice (Air-Self vs. Air-Other, Air-Self vs. Bone-Other, Air-Self vs. Bone-Self, Air-Other vs. Bone-Other, Air-Other vs. Bone-Self, Bone-Other vs. Bone-Self). This analysis indicated that map 4 occurred significantly more in Air-Self compared to all other conditions, with no significant differences in occurrence between other conditions (Supplementary Table 1 and Supplementary Fig. 2).

## Map4 Occurrence-Behavior Relationship

The linear mixed-effects regression considered for the link between EEG and behavioral patterns revealed the main effect of Accuracy ($F(1, 25.28) = 5.6$, $P = 0.026$), indicating a negative relationship with overall task accuracy and map occurrence. There was a three-way interaction between the effects of Dominant Voice, Conduction and Accuracy ($F(1, 39.42) = 5.29$, $P = 0.027$). To further investigate the nature of this interaction, we performed a separate analysis for each type of sound conduction. Whereas there were no significant effects in Bone conduction (all $P > 0.05$), in the trials with Air conduction there was a significant interaction between Dominant voice and Accuracy ($F(1, 26.27) = 5.29$, $P = 0.037$), indicating a stronger negative relationship between Accuracy and Occurrence for Self-dominant, compared to other-dominant morphs (Fig. 7C). For map 4, mixed-effects regression with Response Times as additional factor similarly revealed the main effect of Response Times ($F(1, 28.33) = 11.1$, $P = 0.002$), indicating an overall increase in Occurrence with the increase in Response Times. There was a significant interaction between Response Times and Dominant voice ($F(1, 21.86) = 15.87$, $P < 0.001$), indicating a stronger positive relationship between response times and map 4 occurrence for Self-dominant, compared to Other dominant morphs (Fig. 7D). Response Time did not interact with Conduction ($F(1, 28.52) = 1.34$, $P = 0.26$), and there was no significant three-way interaction between Response Times, Conduction and Response Times ($F(1, 19.09) = 1.2$, $P = 0.29$).

## Source Localization

In order to localize the brain networks associated to the EEG topographies resulting from the segmentation (Fig. 6B), we restricted the visualization of the current density values estimated from the inversion of the cluster maps to the top 5th percentile of the distribution of activation values across all solution points. Brain areas were assessed by the overlap with the automated anatomical labeling (AAL) brain atlas (Tzourio-Mazoyer et al. 2002).

Map 4 revealed a network more strongly activated in the right than in the left hemisphere (Supplementary Fig. 1) with

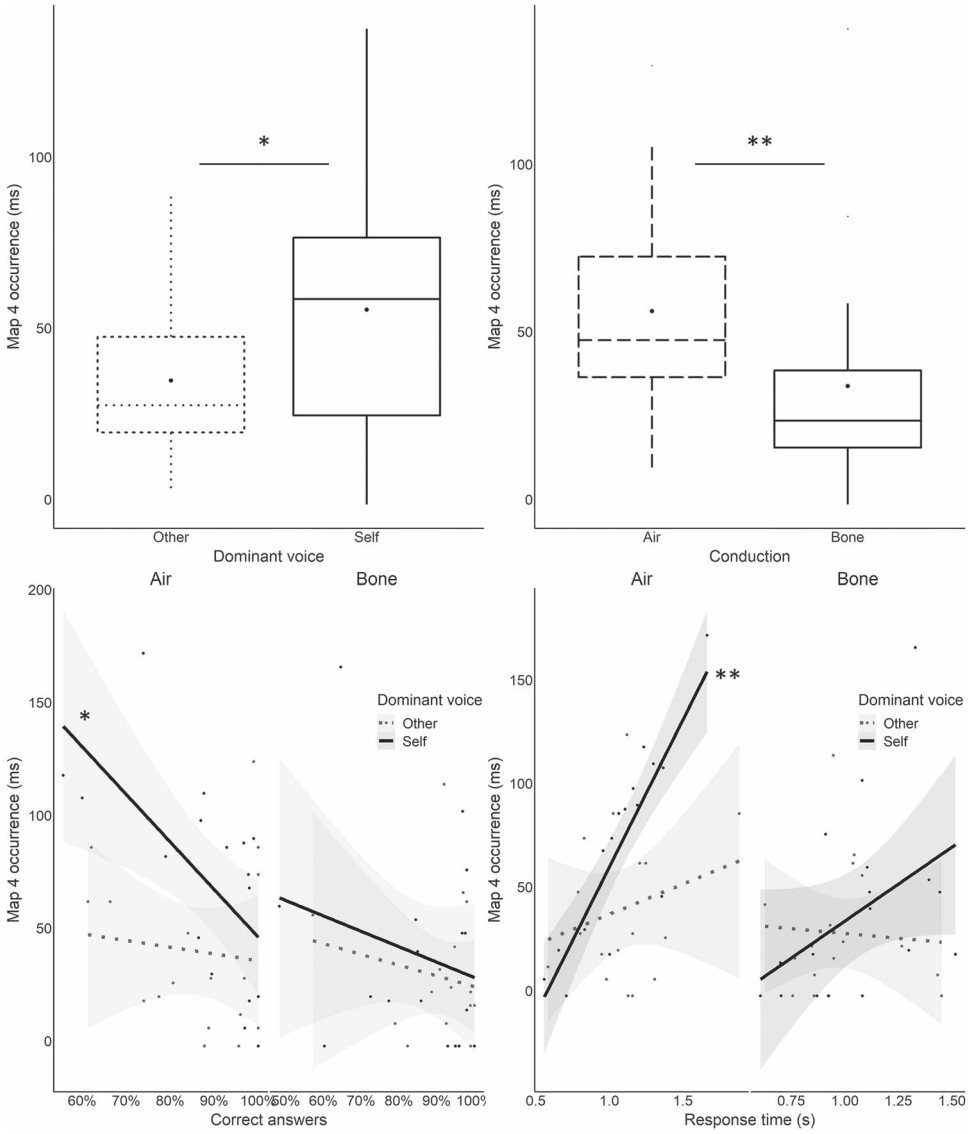


**Figure 7.** Experimental effects on map 4 occurrence. Map 4 occurred more for self-dominant morphs (*A*) and when stimuli were presented through air conduction (*B*). Horizontal lines in boxplots indicate median, whereas dots mean values. The whiskers extend to 1.5 interquartile range and the remaining dots are outliers. Map 4 occurrence was negatively correlated to task accuracy (*C*) and positively to response times (*D*), specifically for self-dominant stimuli presented through air conduction. Shaded areas around linear regressions represent 95% confidence intervals. *: $P < 0.05$; **: $P < 0.001$.

maximum of activation in the right insula. Concomitant activation was observed in: right hippocampus, right superior and inferior temporal pole, left insula, in the middle cingulum, in bilateral putamen and amygdalae (Fig. 8 and Table 2).

## Discussion

In this work, we analyzed the electrophysiological processes involved in the discrimination of self- versus other voice, by integrating the effect of hearing the voices through two sound conduction systems.

The electrode-wise waveform analysis showed a main effect of the conduction in the early stage of the auditory process, namely for the N1 ERP component, likely demonstrating an effect on the sound perception according to the conduction system used. Topographical analysis confirmed this finding and in addition revealed the main effect of conduction in a late stage, correspondent to the of the ERP components P3/N4. However, the conventional ERP analysis did not reveal any differences in amplitude or latencies between self- and other voice for P3/N4. In contrast, the EEG microstate segmentation allowed to define a topographic map associated with the ability to discriminate self-voice from other voices. This map (map 4) was activated more during self-dominant voice and more with air conduction. Map 4 occurred after 300 ms from stimuli onset had a fronto-temporal positivity and parieto-occipital negativity with gradients towards the right hemisphere. Source localization of this map identified a bilateral, but right-dominant network with the maximum activation in the insula and further involving bilaterally the putamen, the amygdalae,

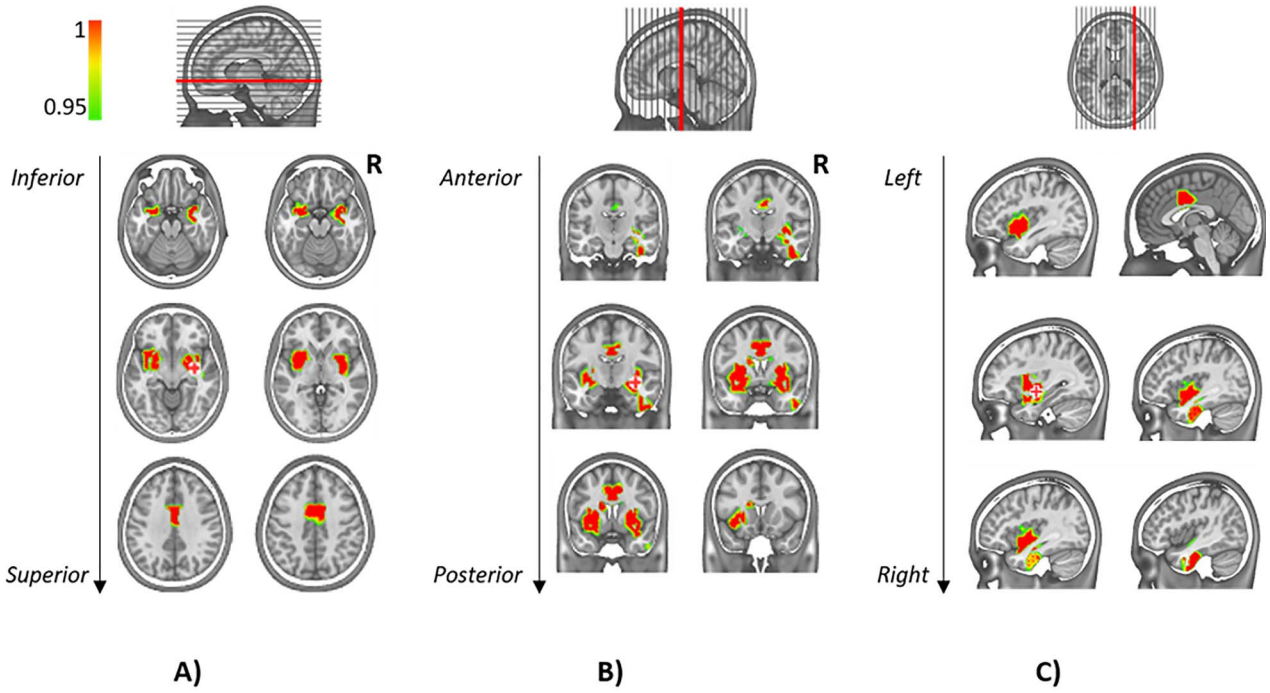


**Figure 8.** Brain network associated to map 4. The axial (A), coronal (B), and sagittal (C) view of the source current density estimated for map 4 are shown. Details on the anatomical regions forming the network are reported in Table 2. Green-to-red blobs indicate activation above the 95th percentile of the distribution of the total activation values (color-bar on the top left). Red lines indicate the region of maximal activation. R: right.

**Table 2** Brain network associated to EEG topography 4

| Brain region | AAL region number | MNI coordinates |
|---|---|---|
| *R* insula | 30 | [43,−1,0] |
| *R* hippocampus | 38 | [26,−5,−17] |
| *R* superior temporal pole | 84 | [40,−1,11] |
| *R* inferior temporal pole | 90 | [38,−10,−36] |
| *L* insula | 29 | [−40,9,4] |
| Middle cingulum | 33, 34 | [0,−6,40] |
| *R* and *L* putamen | 73,74 | [±31,−3,−4] |
| *R* and *L* amygdala | 41, 42 | [±24,−5,−11] |

Notes: AAL atlas was used to define the anatomical name of the regions defining the brain network resulting from the inversion of the cluster map 4. The index of the AAL atlas and the MNI coordinates are also indicated. R: right; L: left.

middle cingulum, and the right superior and inferior temporal pole.

Taken together, our results emphasize the role of the late auditory ERP components, P3/N4 in the voice discrimination task, as reported in the literature (Graux et al. 2013, 2015; Conde et al. 2016, 2018; Liu et al. 2019). However, in contrast to previous work using oddball paradigms, our task did not evoke differences in terms of amplitudes of late components, but in terms of topography of the evoked potential field.

Crucially, we link the self-voice related neural activations to behavioral performance: the occurrence of this map negatively correlated with accuracy in self-voice trials, which was more prominent during air conduction. Similarly, we observed a positive correlation between map occurrence and response times in self-voice trials, which was, again, stronger when self-voice stimuli were presented through air conduction.

We propose that map 4 reflects an auditory self-referencing mechanism that compares the heard voice with the internal self-voice representation. The map occurred more during self-dominant voice morphs, which indicates its specificity for self-voice, and, crucially, it was related to behavioral performance only in self-dominant voice morphs. Specifically, the better participants were in labeling the ambiguous morph as self-dominant and the faster they responded, the less they activated the map. This finding corresponds to the interpretation of earlier ERP studies that less pre-attentional processes need to be recruited in the discrimination of one's own voice (Graux et al. 2015). Seen from a different perspective, the worse participants were at recognizing their voice, the more they recruited the network generating this map. This is reminiscent of a scenario in which a given resource is used only when needed. For example, imagine writing an exam in physics while being allowed to consult a relevant textbook. The better the hypothetical examinee is in physics, the less she/he will use the textbook. Accordingly, our data shows that the better participants are in recognizing our voice, the less they recruit

this network. Thus, we suggest that this map represents a neural resource that is used in scenarios of vocal ambiguity, thereby exploited proportionally to the ability to recognize self-voice. This interpretation is further confirmed in the observation that the same map is activated more with air conduction. Namely, as we have previously shown (Orepic 2020) and replicated here, participants exhibit a poorer performance at the SOVD task with air, compared to bone conduction. It is thus possible that, since the task is more difficult when conducted with air conduction, the map is recruited more.

The spatial organization of the brain network activating during the SOVD task further supports our proposal. First, right-hemisphere dominance for self-recognition has been consistently reported in various paradigms (Decety and Sommerville 2003; Feinberg and Keenan 2005; Uddin et al. 2005; Frassinetti et al. 2008; Candini et al. 2018). The insula, where we observe the maximal activation, is considered a hub for multisensory integration of exteroceptive and interoceptive signals that serves as a basis for maintaining a coherent representation of our bodily self (Craig 2009; Seth 2013; Ionta et al. 2014; Babo-Rebelo et al. 2016; Park et al. 2018; Park and Blanke 2019). Insula activation specific to self-dominant morphs further supports our proposal that self-voice is fundamentally a multimodal construct (Orepic 2020), integrating auditory (i.e., the sound of our voice) and somatosensory (i.e., vibrotactile excitation resulting from bone conduction) signals. Integration of auditory and somatosensory information has been observed in rat's posterior insula (Rodgers et al. 2008); in monkeys insula selectively activates for conspecific vocalization compared to other sounds (environmental or other animals vocalizations) (Remedios et al. 2009) and in humans bilateral insulae has been related to auditory temporal processing of nonlinguistic auditory stimuli (Steinbrink et al. 2009). Moreover, (Nakamura et al. 2001) observed activations in parainsular cortex specific to self-voice stimuli and (Shergill et al. 2001) associated insula with inner speech generation. The cingulate gyrus has been associated to self-referential processing across different functional domains (for a review, see Northoff et al. 2006) and specific to self-voice in the study of Allen et al. (2005). Thereby, and due to its connections with all cerebral lobes via association fibers a primordial role for the cingulate for emotional awareness and for the cognitive default mode network has been postulated for the cingulate gyrus (Maldonado et al. 2020). Moreover, from anatomical point of view, it is believed that insula and the anterior part of the cingulate cortex evolved by developing same neuronal characteristics (the spindle-shaped neurons or von Economo neurons) to support specific functions, such as self-awareness (Allman et al. 2010).The hippocampus is well known to play a role in autobiographical memory retrieval (Greenberg et al. 2005; Cabeza and St Jacques 2007) as well as in self-referential processing (Kurczek et al. 2015); thus, its recruitment in SOVD task might reflect the retrieval of the internal self-voice representation, that is compared to the voice morph. Finally, the amygdala is recognized as a detector of several self-relevant processes, such as fear conditioning in both animals and humans (Markram et al. 2008; Rodgers et al. 2008; Feinstein et al. 2011), or for the evaluation of the emotional content in human voices (Sander et al. 2003). The amygdaloid complex consists of several nuclei with reciprocal (para-) limbic connections with the olfactory system, the anterior nucleus of the thalamus, and the basal part of the insula, (insula of Reil; Mesulam and Mufson 1985), which may explain the functional association of these areas in integrating the extra-personal experiences and (self-)motivational states (Marafioti et al. 2018). This notion has been confirmed by various clinical and basic neuro-anatomic observations in the context of the salience-network research (Seeley 2019). We hypothesize that in our experiment the amygdala could have contributed to the "affective" response known to occur after hearing the own voice (Holzman and Rousey 1966).

We believe that this work would have relevant clinical impact, in particular to characterize the still poorly understood neuronal mechanisms underlying auditory-verbal hallucinations (AVH) (Wilkinson and Alderson-Day 2016). Being predominantly negative in content and in affect, AVH represent a major source of distress in patients with psychosis and their presence alone significantly increases risk of suicide in this group (Harkavy-Friedman et al. 2003). Thus, understanding the etiology of AVH is a critical next step towards the development of new diagnostic tools and treatments. One of the most prominent accounts for AVH suggests that they arise as a self-to-other misattribution of inner speech, i.e., as an impairment in (internal) SOVD (Frith and Done 1989; Frith 1992; Ford and Mathalon 2005; Ford et al. 2007; Moseley et al. 2013). Crucially, to date, no study has directly investigated neural underpinnings of SOVD in voice hearers and the empirical support for this account mainly comes from studies (reviewed by Whitford 2019) in which patients with schizophrenia, compared to healthy participants, exhibited differences in amplitudes of certain ERP components following self-voice stimuli. In this work, however, we identified a brain mechanism directly related to SOVD; thus, an investigation of the corresponding network in people with AVH could provide evidence that would either significantly support or challenge this long-standing and prominent account.

## Methodological Consideration

In this work, we investigated the electrophysiological effect linked the SOVD task, in a period lasting 500 ms from the stimulus onset. Although the literature is rich with documentations and guidelines for an efficient analysis of ERPs, a standard duration for the analysis is not established and rather depends on the type of stimulation (auditory, motor, visual, and so on) and on the hypothesis beyond the study (Handy 2005; Woodman 2010; Luck 2014). In our case, we investigated the periods associated to the auditory ERP components with particular regard to the late components, P3 and N4 which in the literature are considered as the main "candidates" for the self-voice discrimination (Titova and Näätänen 2001; Beauchemin et al. 2006; Graux et al. 2013, 2015; Conde et al. 2018), extending the window length to 500 ms to account for participants' variability.

The EEG segmentation was selected as a comprehensive analysis of the changes of the scalp voltage potential in strength, topography, and time independent of the appearance of dominant ERP components. This technique is complemented by the back-fitting, which establishes the "role" of each topographic map in each subject, allowing the statistical inference. Therefore, segmentation results on the grand mean ERP should not been interpreted without the back-fitting procedure and statistical analysis of the fitting parameters. For example, in our study, map 6 could be considered as a peculiarity of the air conduction according to the segmentation (Fig. 6A). However, when tested across subjects, the statistics did not show any significance for this map, probably due to the low GFP already visible in the grand-averaged ERPs and even lower at the subject level.

Although the recognized limitation of the spatial resolution of EEG and MEG source imaging compared to other neuroimaging techniques (e.g. fMRI), the accuracy and

precision of the estimated sources from high-density EEG (>200 channels) is well established through experimental, clinical, and simulation studies (Michel and Murray 2012; Michel and He 2019; Seeber et al. 2019). In this work, we applied an innovative and validated approach to localize sources of EEG microstates by temporal normalization approaches (Bréchet et al. 2019, 2020, 2021; Michel and Brunet 2019) to define the brain sources responsible of the activity recorded on the scalp and interpret our findings in terms of brain network.

First of all, the solutions points were constrained to belong to the gray matter only. Secondly, the approach attributes an identical background distribution ("weight") to each solution point and significantly improves the accuracy (not to be confused with the lower resolution) of the localization. In fact, after the inversion step, each solution point returns an activation value that explains the input EEG. The distribution of such values, which can be normalized between [0,1] (as in this work), demonstrates that the majority of solution points are characterized by low values and can be, therefore, ascribed to noise/background activity. Therefore, only the right tail of the distribution (i.e., solution points with normalized activation values close to 1) is meaningful to detect significant activations.

Of course, the results strongly depend on the threshold value, which at the same time is a necessary parameter to fix (in all types of analysis). Here, we considered as threshold a set of percentiles of the distribution of activation values. In details, we explored the results obtained by selecting the values from the 20th to the 1th percentiles of the distribution which allowed to detect a consistency of the localization of the activated brain areas (Supplementary Fig. 1) across the different threshold used.

### General Conclusions

In conclusion, this work encompasses novel scientific and methodological aspects. First, to the best of our knowledge, we were the first to correlate self-voice-related neural activity with behavioral task performance. This work advances the understanding of the self-voice phenomenon as it provides an exhaustive characterization of spatio-temporal activity related to SOVD, relying on the improved source localization of high-density EEG and the topographic segmentation approach. Second, we consolidate the importance of multisensory self-voice presentation by demonstrating a reduced processing of the map associated to SOVD task performance with bone-compared to air-conducted self-voices. Finally, this work has clinical importance as it sheds new light on the very mechanism believed to account for auditory–verbal hallucinations—a major source of distress in mental disorders (Harkavy-Friedman et al. 2003) whose underlying principles are still unknown (Wilkinson and Alderson-Day 2016).

## Supplementary Material

Supplementary material can be found at *Cerebral Cortex* online.

## Notes

The authors thank the Swiss Foundation for Innovation and Training in Surgery (SFITS) in Geneva who hosted the experimental platform for this research. We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), Ecole polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG). *Conflict of Interest:* The corresponding author (C.M.M.) and co-authors have no conflict of interest to declare.

## Funding

The Swiss National Science Foundation (grant no. 320030_182497 to K.S.), with co-partners C.M.M., O.B.; the Pictet Foundation to K.S., O.B., and C.M.M; the Swiss National Science Foundation (grant no. 320030_184677 to C.M.M.). This research was supported by two donors advised by CARIGEST SA (Fondazione Teofilo Rossi di Montelera e di Premuda and a second one wishing to remain anonymous) to O.B.

## References

Allen PP, Amaro E, Fu CHY, Williams SCR, Brammer M, Johns LC, McGuire PK. 2005. Neural correlates of the misattribution of self-generated speech. *Hum Brain Mapp*. 26:44–53.

Allman JM, Tetreault NA, Hakeem AY, Manaye KF, Semendeferi K, Erwin JM, Park S, Goubert V, Hof PR. 2010. The von Economo neurons in frontoinsular and anterior cingulate cortex in great apes and humans. *Brain Struct Funct*. 214(5):495–517.

Babo-Rebelo M, Wolpert N, Adam C, Hasboun D, Tallon-Baudry C. 2016. Is the cardiac monitoring function related to the self in both the default network and right anterior insula? *Philos Trans R Soc B Biol Sci*. 371(1708):20160004.

Bates D, Mächler M, Bolker BM, Walker SC. 2015. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 67:1–48.

Beauchemin M, De Beaumont L, Vannasing P, Turcotte A, Arcand C, Belin P, Lassonde M. 2006. Electrophysiological markers of voice familiarity. *Eur J Neurosci*. 23:3081–3086.

Békésy GV. 1949. The structure of the middle ear and the hearing of one's own voice by bone conduction. *J Acoust Soc Am*. 21:217–232.

Belin P, Fecteau S, Bédard C. 2004. Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci*. 8: 129–135.

Blakemore SJ, Wolpert D, Frith C. 2000. Why can't you tickle yourself? *Neuroreport*. 11:R11–R16.

Blank H, Wieland N, von Kriegstein K. 2014. Person recognition and the brain: merging evidence from patients and healthy individuals. *Neurosci Biobehav Rev*. 47:717–734.

Blanke O. 2012. Multisensory brain mechanisms of bodily self-consciousness. *Nat Rev Neurosci*. 13:556–571.

Blanke O, Slater M, Serino A. 2015. Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*. 88:145–166.

Brandeis D, Lehmann D. 1986. Event-related potentials of the brain and cognitive processes: approaches and applications. *Neuropsychologia*. 24:151–168.

Braun N, Debener S, Spychala N, Bongartz E, Sörös P, Müller HHO, Philipsen A. 2018. The senses of agency and ownership: a review. *Front Psychol*. 9:1–17.

Bréchet L, Brunet D, Birot G, Gruetter R, Michel CM, Jorge J. 2019. Capturing the spatiotemporal dynamics of self-generated, task-initiated thoughts with EEG and fMRI. *Neuroimage*. 194:82–92.

Bréchet L, Brunet D, Perogamvros L, Tononi G, Michel CM. 2020. EEG microstates of dreams. *Sci Rep*. 10:17069.

Bréchet L, Ziegler DA, Simon AJ, Brunet D, Gazzaley A, Michel CM. 2021. Reconfiguration of electroencephalography microstate

networks after breath-focused, digital meditation training. *Brain Connect*. 11:146–155.
Brodbeck V, Spinelli L, Lascano AM, Wissmeier M, Vargas M-I, Vulliemoz S, Pollo C, Schaller K, Michel CM, Seeck M. 2011. Electroencephalographic source imaging: a prospective study of 152 operated epileptic patients. *Brain*. 134:2887–2897.
Brunet D, Murray MM, Michel CM. 2011. Spatiotemporal analysis of multichannel EEG: CARTOOL. *Comput Intell Neurosci*. 2011:1–15.
Cabeza R, St Jacques P. 2007. Functional neuroimaging of autobiographical memory. *Trends Cogn Sci*. 11:219–227.
Candini M, Avanzi S, Cantagallo A, Zangoli MG, Benassi M, Querzani P, Lotti EM, Iachini T, Frassinetti F. 2018. The lost ability to distinguish between self and other voice following a brain lesion. *NeuroImage Clin*. 18:903–911.
Conde T, Gonçalves ÓF, Pinheiro AP. 2015. Paying attention to my voice or yours: An ERP study with words. *Biol Psychol*. 111:40–52.
Conde T, Gonçalves ÓF, Pinheiro AP. 2016. The effects of stimulus complexity on the preattentive processing of self-generated and nonself voices: an ERP study. *Cogn Affect Behav Neurosci*. 16:106–123.
Conde T, Gonçalves ÓF, Pinheiro AP. 2018. Stimulus complexity matters when you hear your own voice: attention effects on self-generated voice processing. *Int J Psychophysiol*. 133:66–78.
Craig AD. 2009. How do you feel - now? The anterior insula and human awareness. *Nat Rev Neurosci*. 10(1):59–70. doi: 10.1038/nrn2555.
Decety J, Sommerville JA. 2003. Shared representations between self and other: a social cognitive neuroscience view. *Trends Cogn Sci*. 7:527–533.
Emami SF, Pourbakht A, Sheykholeslami K, Kamali M, Behnoud F, Daneshi A. 2012. Vestibular hearing and speech processing. *ISRN Otolaryngol*. 2012:1–7.
Feinberg TE, Keenan JP. 2005. Where in the brain is the self? *Conscious Cogn*. 14:661–678.
Feinstein JS, Adolphs R, Damasio A, Tranel D. 2011. The human amygdala and the induction and experience of fear. *Curr Biol*. 21(1):34–38.
Ford JM, Mathalon DH. 2005. Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *Int J Psychophysiol*. 58:179–189.
Ford JM, Roach BJ, Faustman WO, Mathalon DH. 2007. Synch before you speak: auditory hallucinations in schizophrenia. *Am J Psychiatry*. 164:458–466.
Frassinetti F, Maini M, Romualdi S, Galante E, Avanzi S. 2008. Is it mine? Hemispheric asymmetries in corporeal self-recognition. *J Cogn Neurosci*. 20:1507–1516.
Frith CD. 1992. *The cognitive neuropsychology ofschizophrenia*. Hove, UK: L. Erlbaum Associates. https://www.worldcat.org/title/cognitive-neuropsychology-of-schizophrenia/oclc/26896101?page=citation.
Frith CD, Done DJ. 1989. Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychol Med*. 19:359–363.
Gillihan SJ, Farah MJ. 2005. Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychol Bull*. 131:76–97.
Graux J, Gomot M, Roux S, Bonnet-Brilhault F, Bruneau N. 2015. Is my voice just a familiar voice? An electrophysiological study. *Soc Cogn Affect Neurosci*. 10:101–105.
Graux J, Gomot M, Roux S, Bonnet-Brilhault F, Camus V, Bruneau N. 2013. My voice or yours? An electrophysiological study. *Brain Topogr*. 26:72–82.
Greenberg DL, Rice HJ, Cooper JJ, Cabeza R, Rubin DC, LaBar KS. 2005. Co-activation of the amygdala, hippocampus and inferior frontal gyrus during autobiographical memory retrieval. *Neuropsychologia*. 43:659–674.
Grossmann T, Oberecker R, Koch SP, Friederici AD. 2010. The developmental origins of voice processing in the human brain. *Neuron*. 65(6):852–858.
Gur RC, Sackeim HA. 1979. Self-deception: a concept in search of a phenomenon. *J Pers Soc Psychol*. 37:147–169.
Habermann M, Weusmann D, Stein M, Koenig T. 2018. A student's guide to randomization statistics for multichannel event-related potentials using Ragu. *Front Neurosci*. 12:355.
Handy TC, editor. 2005. *Event-related potentials: a methods handbook*. MIT press.
Harkavy-Friedman JM, Kimhy D, Nelson EA, Venarde DF, Malaspina D, Mann JJ. 2003. Suicide attempts in schizophrenia: the role of command auditory hallucinations for suicide. *J Clin psychiatry*. 64:871–874.
Holzman PS, Rousey C. 1966. The voice as a percept. *J Pers Soc Psychol*. 4:79–86.
Hughes SM, Nicholson SE. 2010. The processing of auditory and visual recognition of self-stimuli. *Conscious Cogn*. 19:1124–1134.
Ionta S, Martuzzi R, Salomon R, Blanke O. 2014. The brain network reflecting bodily self-consciousness: a functional connectivity study. *Soc Cogn Affect Neurosci*. 9:1904–1913.
Ito T, Tiede M, Ostry DJ. 2009. Somatosensory function in speech perception. *Proc Natl Acad Sci*. 106:1245–1248.
Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. 2000. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol*. 111:1745–1758.
Kannape OA, Blanke O. 2012. Agency, gait and self-consciousness. *Int J Psychophysiol*. 83:191–199.
Kaplan JT, Aziz-Zadeh L, Uddin LQ, Iacoboni M. 2008. The self across the senses: an fMRI study of self-face and self-voice recognition. *Soc Cogn Affect Neurosci*. 3:218–223.
Karniski W, Blair RC, Snider AD. 1994. An exact statistical method for comparing topographic maps, with any number of subjects and electrodes. *Brain Topogr 1994 63*. 6: 203–210.
Kawahara H, Morise M, Banno H, Skuk VG. 2013. Temporally variable multi-aspect N-way morphing based on interference-free speech representations. In: *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013*. IEEE, pp. 1–10.
Kircher T, David AS. 2003. Self-consciousness: an integrative approach from philosophy, psychopathology and the neurosciences. In: Kircher T, David A, editors. *The self in neuroscience and psychiatry*. Cambridge University Press, pp. 445–473. https://doi.org/10.1017/CBO9780511543708.023.
Kleiner M, Brainard DH, Pelli DG, Ingling A, Murray R, Broussard AI, R Murray C. 2007. What's new in psychtoolbox-3? *Perception*. 36:1–16.
Koenig T, Kottlow M, Stein M, Melie-García L. 2011. Ragu: a free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Comput Intell Neurosci*. 2011.

Koenig T, Melie-García L. 2010. A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topogr*. 23:233–242.

Kurczek J, Wechsler E, Ahuja S, Jensen U, Cohen NJ, Tranel D, Duff M. 2015. Differential contributions of hippocampus and medial prefrontal cortex to self-projection and self-referential processing. *Neuropsychologia*. 73:116–126.

Kuznetsova A, Brockhoff PB, Christensen RHB. 2018. lmerTest package: tests in linear mixed effects models. *J Stat Softw*. 82:1–26.

Lehmann D, Skrandies W. 1980. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol*. 48:609–621.

Liu L, Li W, Li J, Lou L, Chen J. 2019. Temporal features of psychological and physical self-representation: an ERP study. *Front Psychol*. 10:785.

Luck S. 2014. *An introduction to the event-related potential technique*. MIT press.

Lüdecke D. 2021. sjPlot: data visualization for statistics in social science. *R package version 2.8.9*, https://CRAN.R-project.org/package=sjPlot. https://cran.r-project.org/web/packages/sjPlot/citation.html.

Maldonado IL, Parente de Matos V, Castro Cuesta TA, Herbet G, Destrieux C. 2020. The human cingulum: from the limbic tract to the connectionist paradigm. *Neuropsychologia*.. 144:107487.

Marafioti V, Turri G, Carbone V, Monaco S. 2018. Association of prolonged QTc interval with Takotsubo cardiomyopathy: a neurocardiac syndrome inside the mystery of the insula of Reil. *Clin Cardiol*. 41(4):551–555.

Markram K, Rinaldi T, La Mendola D, Sandi C, Markram H. 2008. Abnormal fear conditioning and amygdala processing in an animal model of autism. *Neuropsychopharmacology*. 33(4): 901–912.

McGettigan C, Eisner F, Agnew ZK, Manly T, Wisbey D, Scott SK. 2013. T'ain't what you say, it's the way that you say it–left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. *J Cogn Neurosci*. 25(11):1875–1886.

Mégevand P, Spinelli L, Genetti M, Brodbeck V, Momjian S, Schaller K, Michel CM, Vulliemoz S, Seeck M. 2014. Electric source imaging of interictal activity accurately localises the seizure onset zone. *J Neurol Neurosurg Psychiatry*. 85: 38–43.

Mesulam M-M, Mufson EJ. 1985. *The insula of Reil in man and monkey*. In *Association and auditory cortices*. Boston, MA: Springer, pp. 179–226.

Michel CM, Brunet D. 2019. EEG source imaging: a practical review of the analysis steps. *Front Neurol*. 10:325.

Michel CM, He B. 2019. EEG source localization. *Handb Clin Neurol*. 160:85–101.

Michel CM, Koenig T. 2018. EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: a review. *Neuroimage*. 180:577–593.

Michel CM, Koenig T, Brandeis D, Gianotti LR, Wackermann J. 2009. *Electrical neuroimaging*. Cambridge University Press.

Michel CM, Murray MM. 2012. Towards the utilization of EEG as a brain imaging tool. *Neuroimage*. 61:371–385.

Michel CM, Thut G, Morand S, Khateb A, Pegna AJ, Grave de Peralta R, Gonzalez S, Seeck M, Landis T. 2001. Electric source imaging of human brain functions. *Brain Res Brain Res Rev*. 36:108–118.

Moseley P, Fernyhough C, Ellison A. 2013. Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neurosci Biobehav Rev*. 37:2794–2805.

Murray MM, Brunet D, Michel CM. 2008. Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr*. 20: 249–264.

Nakamura K, Kawashima R, Sugiura M, Kato T, Nakamura A, Hatano K, Nagumo S, Kubota K, Fukuda H, Ito K, et al. 2001. Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia*. 39:1047–1054.

Northoff G, Heinzel A, de Greck M, Bermpohl F, Dobrowolny H, Panksepp J. 2006. Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *Neuroimage*. 31:440–457.

Orepic P. 2020. *Dissecting self-voice perception: from bone conduction to robotically-induced self-other voice misattribution in healthy listeners (No. 8468)*. EPFL.

Orepic P, Rognini G, Kannape OA, Faivre N, Blanke O. 2021. Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners. *Schizophr Res*. 231: 170–177.

Park HD, Bernasconi F, Salomon R, Tallon-Baudry C, Spinelli L, Seeck M, Schaller K, Blanke O. 2018. Neural sources and underlying mechanisms of neural responses to heartbeats, and their role in bodily self-consciousness: an intracranial EEG study. *Cereb cortex*. 28:2351–2364.

Park HD, Blanke O. 2019. Coupling inner and outer body for self-consciousness. *Trends Cogn Sci*. 23:377–388.

Pascual-Marqui RD, Esslen M, Kochi K, Lehmann D. 2002. Functional imaging with low-resolution brain electromagnetic tomography (LORETA): a review. *Meth Find Exp Clin Pharmacol*. 24(Suppl C):91–95.

Pascual-Marqui RD, Michel CM, Lehmann D. 1995. Segmentation of brain electrical activity into microstates: model estimation and validation. *IEEE Trans Biomed Eng*. 42:658–665.

Perrin F, Pernier J, Bertrand O, Echallier JF. 1989. Spherical splines for scalp potential and current density mapping. *Electroencephalogr Clin Neurophysiol*. 72:184–187.

Picton TW. 2010. *Human auditory evoked potentials*. Plural Publishing.

R Core Team. 2020. R: a language and environment for statistical computing. *R Found Stat Comput*.

Reinfeldt S, Östli P, Håkansson B, Stenfelt S. 2010. Hearing one's own voice during phoneme vocalization—transmission by air and bone conduction. *J Acoust Soc Am*. 128:751–762.

Remedios R, Logothetis NK, Kayser C. 2009. An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. *J Neurosci*. 29:1034–1045.

Rodgers KM, Benison AM, Klein A, Barth DS. 2008. Auditory, somatosensory, and multisensory insular cortex in the rat. *Cereb Cortex*. 18:2941–2951.

Rosa C, Lassonde M, Pinard C, Keenan JP, Belin P. 2008. Investigations of hemispheric specialization of self-voice recognition. *Brain Cogn*. 68:204–214.

Sander D, Grafman J, Zalla T. 2003. The human amygdala: an evolved system for relevance detection. *Rev Neurosci*. 14: 303–316.

Schiller B, Gianotti L, Baumgartner T, Nash K, Koenig K, Knoch D. 2016. Clocking the social mind by identifying mental processes in the IAT with electrical neuroimaging. *Proc Natl Acad Sci USA*. 113:2786–2791.

Schuerman WL, Meyer A, McQueen JM. 2015. Do we perceive others better than ourselves? A perceptual benefit for noise-vocoded speech produced by an average speaker. *PLoS One*. 10:1–18.

Seeber M, Cantonas L-M, Hoevels M, Sesia T, Visser-Vandewalle V, Michel CM. 2019. Subcortical electrophysiological activity is detectable with high-density EEG source imaging. *Nat Commun*. 10:1–7.

Seeley WW. 2019. The salience network: a neural system for perceiving and responding to homeostatic demands. *J Neurosci*. 39(50):9878–9882.

Seth AK. 2013. Interoceptive inference, emotion, and the embodied self. *Trends Cogn Sci*. 17:565–573.

Shergill SS, Bullmore ET, Brammer MJ, Williams SCR, Murray RM, McGuire PK. 2001. A functional study of auditory verbal imagery. *Psychol Med*. 31:241–253.

Shergill SS, White TP, Joyce DW, Bays PM, Wolpert DM, Frith CD. 2014. Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. *JAMA Psychiatry*. 71:28–35.

Shuster LI. 1998. The perception of correctly and incorrectly produced /r/. *J Speech, Lang Hear Res*. 41:941–950.

Steinbrink C, Ackermann H, Lachmann T, Riecker A. 2009. Contribution of the anterior insula to temporal auditory processing deficits in developmental dyslexia. *Hum Brain Mapp*. 30:2401–2411.

Stenfelt S. 2016. Model predictions for bone conduction perception in the human. *Hear Res*. 340:135–143.

Titova N, Näätänen R. 2001. Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neurosci Lett*. 308:63–65.

Todd NPMA, Cody FWJ, Banks JR. 2000. A saccular origin of frequency tuning in myogenic vestibular evoked potentials? Implications for human responses to loud sounds. *Hear Res*. 141:180–188.

Tremblay S, Shiller DM, Ostry DJ. 2003. Somatosensory basis of speech production. *Nature*. 423:866–869.

Tsakiris M. 2017. The multisensory basis of the self: from body to identity to others. *Q J Exp Psychol*. 70:597–609.

Tsakiris M, Haggard P. 2005. Experimenting with the acting self. *Cogn Neuropsychol*. 22:387–407.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 15:273–289.

Uddin LQ, Kaplan JT, Molnar-Szakacs I, Zaidel E, Iacoboni M. 2005. Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: an event-related fMRI study. *Neuroimage*. 25:926–935.

Whitford TJ. 2019. Speaking-induced suppression of the auditory cortex in humans and its relevance to schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 4:791–804.

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag. https://ggplot2.tidyverse.org.

Wilkinson S, Alderson-Day B. 2016. Voices and thoughts in psychosis: an introduction. *Rev Philos Psychol*. 7:529–540.

Winkler I, Denham S, Escera C. 2015. Auditory event-related potentials. In: Jaeger D, Jung R, editors. *Encyclopedia of computational neuroscience*. New York (NY): Springer.

Woodman GF. 2010. A brief introduction to the use of event-related potentials in studies of perception and attention. *Atten Percept Psychophys*. 72:2031–2046.

Zarate JM, Tian X, Woods KJP, Poeppel D. 2015. Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Sci Rep*. 5:11475.