



HAL
open science

Talking Heads: Bone conduction facilitates self-other voice discrimination

Pavo Orepic, Oliver Alan Kannape, Nathan Faivre, Olaf Blanke

► **To cite this version:**

Pavo Orepic, Oliver Alan Kannape, Nathan Faivre, Olaf Blanke. Talking Heads: Bone conduction facilitates self-other voice discrimination. 2022. hal-03718489v1

HAL Id: hal-03718489

<https://hal.science/hal-03718489v1>

Preprint submitted on 8 Jul 2022 (v1), last revised 21 Feb 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Talking Heads: Bone conduction facilitates self-other voice discrimination

Pavo Orepic¹, Oliver Alan Kannape², Nathan Faivre³, Olaf Blanke^{1, 4}

¹ Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Switzerland

² Virtual Medicine Center, University Hospital of Geneva, Geneva, Switzerland

³ University Grenoble Alpes, University Savoie Mont Blanc, CNRS, LPNC, Grenoble, France.

⁴ Department of Clinical Neurosciences, University Hospital of Geneva, Geneva, Switzerland

Corresponding author

Olaf Blanke

Bertarelli Chair in Cognitive Neuroprosthetics, Center for Neuroprosthetics & Brain Mind Institute, School of Life Sciences, Campus Biotech, Swiss Federal Institute of Technology (EPFL), 1012 Geneva, Switzerland

E-mail: olaf.blanke@epfl.ch

Tel: +41 (0)21 693 69 21

Fax: +41 (0)21 693 69 22

Abstract

One's own voice is one of the most important and most frequently heard voices and the sound we associate most with ourselves, and yet, it is perceived as strange when played back in a recording. One of the main reasons is the lack of bone conduction that is inevitably present when hearing own voice while speaking. The resulting discrepancy between experimental and natural self-voice stimuli has significantly impeded self-voice research, rendering it one of the least investigated aspects of self-consciousness. Accordingly, factors that contribute to self-voice perception remain largely unknown. In a series of three studies, we rectified this ecological discrepancy by augmenting experimental self-voice stimuli with bone-conducted vibrotactile stimulation that is present during natural self-voice perception. Combining voice-morphing with psychophysics, we demonstrate that specifically self-other but not familiar-other voice discrimination improved for stimuli presented using bone as compared to air conduction. Furthermore, our data outline independent contributions of familiarity and acoustic processing to separating the own from another's voice: although vocal differences increased general voice discrimination, self-voices were more confused with familiar than unfamiliar voices, regardless of their acoustic similarity. Collectively, our findings show that concomitant vibrotactile stimulation improves auditory self-identification, thereby portraying self-voice as a fundamentally multimodal construct.

Keywords

Self-voice, self-other voice discrimination, bone conduction, familiar voice, multisensory integration, self-other voice space

Introduction

We are all familiar with the strange sensation that occurs when we hear our voice in video or voice recordings¹⁻⁵. Considering the fundamental role our voice plays in our everyday communication, this should be quite surprising. We have a life-long daily exposure to our voice, higher than exposure even to the most familiar voices. Our own voice is the sound most intimately linked to our self. Although there is ample evidence showing that self-related stimuli are perceived differently and activate distinct cortical regions compared to other, non-self-associated stimuli⁶⁻¹⁴, the specific mechanisms of self-voice perception have been surprisingly underinvestigated, both in behavioral and neuroimaging studies¹⁵⁻¹⁷. For instance, the extent to which self-voice perception differs from that of other familiar voices remains poorly understood; as does the extent to which acoustic properties that enable discriminating voices of other people¹⁸ are involved in self-other voice discrimination (VD). A better understanding of self-voice perception is of immediate clinical relevance, as deficits in self-other VD have been related to auditory-verbal hallucinations¹⁹⁻²² (i.e., “hearing voices”), one of the most common^{23,24} and most distressing^{25,26} hallucinations in a major psychiatric disorder, schizophrenia. Investigating different perceptual factors underlying self-other VD, we here hypothesized that one key contribution would stem from bone-conduction and, based on our findings, propose a new experimental paradigm that improves the ecological validity for studying self-voice perception.

A crucial contribution for the perception of our own voice, and our own voice only, comes from bone conduction resulting from speech production/articulation. Under natural conditions, one’s spoken voice is transmitted not only through the air, but also, unfailingly through the skull^{27,28}, which alters self-voice perception in two ways. First, due to the different sound propagation, bone conduction transforms the sound of our voice – specifically, it is assumed to instantiate a low-pass filter^{29,30}. Because of the low-frequency emphasis, we hear our voice as lower²⁹ compared

to our voice's actual sound and how others hear us. Second, next to transforming the sound of our voice, bone conduction conveys additional sensory information, as not only auditory, but also vibrotactile signals are involved, resulting from the vibrations of the skull^{31–33}. Thus, self-voice, when heard under natural conditions, is not only an auditory but rather a multimodal percept.

One reason for the scarcity of self-voice studies likely lies in methodological obstacles faced when creating appropriate experimental stimuli. Without bone conduction, prior self-voice studies inevitably contain a perceptual mismatch between the experimental self-voice stimuli (e.g., presented through air-conducting loudspeakers) and the actual self-voice. In fact, the majority of studies that compared recognition of self- versus other voices reported lower accuracy rates and higher response times for self-voice compared to other-voices^{16,34–44}. Early self-voice studies suggested that this discrepancy between self- and other voices might result from a lower previous exposure to self-voice in voice recordings^{34,35,37}. However, similar behavioral differences still persist^{16,36–41} with a higher exposure to recorded self-voice through contemporary technology (e.g., voice messages and video recordings). Moreover, more recent self-voice paradigms often demonstrate ceiling effects^{37,39–41}, e.g., high accuracy rates in all experimental conditions, reflecting a need for more sensitive experimental paradigms. To account for the aforementioned ecological discrepancy, several studies investigated if acoustic transformations (e.g., low-pass or other types of filters) of air-conducted self-voice stimuli would render the self-voice more natural to the listeners. These attempts, however, yielded contradictory results^{45–49}, as they indicated preferences for different acoustic transformations. Crucially, these studies manipulated only one aspect related to bone-conduction effects on self-voice (i.e., acoustic transformations), and neglected the additional vibrotactile stimulation. In order to better approximate natural self-voice, experimental self-voice stimuli should be accompanied with the concomitant vibrotactile stimulation resulting from the vibrations of the skull. Here, we address this by providing vibrational

input through a bone-conduction headset and investigate whether it improves self-voice perception, as opposed to auditory input alone.

In a series of three behavioral studies in independent cohorts, and using a new self-voice perception paradigm, we investigated the following three main perceptual factors of self-other VD: (1) sound conduction type (air vs bone), (2) other-voice familiarity (familiar vs unfamiliar), and (3) acoustic voice parameters. Using voice-morphing technology⁵⁰ and bone-conduction headphones, we designed a psychophysical self-other VD task to investigate the nature of perceptual differences in self-other VD, while trying to avoid ceiling effects. Participants heard short voice morphs of their own and other people's vocalizations (phoneme /a/) and indicated whether the morphs more closely resembled their own or someone else's voice. In Study 1 (N = 16), we investigated differences in self-other VD as a function of sound conduction (air, bone) and how this is modulated by previous exposure to self-voice^{34,35,37}; in Study 2 (N = 16), we extended this to familiar-other voice discrimination in order to investigate whether the bone-conduction effects are specific for self-voice, or generalize to other familiar voices^{51,52}. In Study 3, we set out to replicate studies 1 and 2 within a single, larger cohort (N = 52). We furthermore included an additional self-familiar voice discrimination task and a control self-voice recognition task (without voice morphing) and investigated the acoustic parameters of all tested voices¹⁸. We hypothesized that bone conduction would facilitate self-voice perception in self-other VD (bias or increased sensitivity) (Study 1) but would not affect familiar-other VD task (Study 2). We further hypothesized that bone-conduction effects would be more prominent without exposure to the self-voice used in our experiment prior to the task – i.e., when the task difficulty is increased (Studies 1 and 2) – and that they would occur regardless of other-voice familiarity^{51,52} (Study 3).

Results

Study 1: Self-other voice discrimination (VD)

In Study 1, we morphed each participant's voice with the voice of a gender-matched unfamiliar person. For each voice morph, participants were instructed to indicate whether the voice they heard more closely resembled their own or someone else's voice by pressing on one of two buttons. Based on our previous work^{53,54}, six voice ratios (% self-voice: 15, 30, 45, 55, 70, 85;

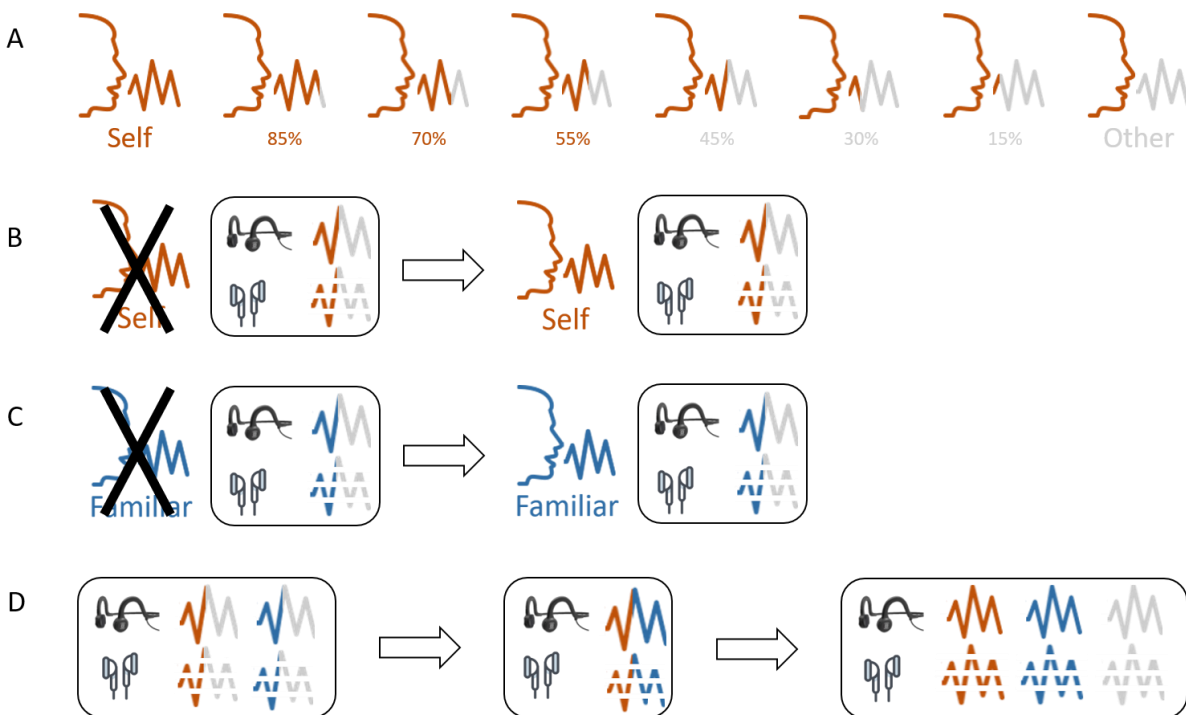


Fig. 1. Experimental design of Studies 1 – 3. Colors represent different types of voices: self (orange), unfamiliar other (gray), familiar other (blue). Blocks represent different types of auditory tasks (self-other, familiar-other, self-familiar, control). In all studies, all tasks were performed both with bone (solid line) and with air (dashed line) conduction (black and white headphone icons, respectively). **A)** Experimental stimuli for the self-other task in Studies 1 and 3. Six voice morphs were sampled from self-other voice continuum generated with voice-morphing technology. Equivalent voice morphs were used in other discrimination tasks (familiar-other, self-familiar). **B)** Study 1 design. Two blocks (with bone and air conduction) of self-other task were first performed without and then with self-voice shown prior to the task (previous exposure to self-voice). **C)** Study 2 design. Two blocks (with bone and air conduction) of familiar-other task were first performed without and then with familiar voice shown prior to the task (previous exposure to familiar voice). **D)** Study 3 design. Self-voice and familiar voice were first discriminated against the unfamiliar voice and then against each other. The control task in which self-voice was detected among the three unmorphed voices was conducted at the end of Study 3.

Fig. 1A) were chosen and repeated 10 times within a block in a randomized order (total of 60 trials). The study contained four experimental blocks, which differed based on the sound conduction type (air, bone) and whether participants were exposed to the unmorphed self-voice immediately prior to the experiment. In the first two blocks, participants performed the task without having previously heard the unmorphed recording of their voice, once with each type of sound conduction, whereas before the remaining two blocks the unmorphed self-voice was presented to participants (**Fig. 1B**). The order of air- and bone-conduction blocks was counterbalanced across participants and for both parts of the experiment (with and without previous exposure to self-voice).

Performance in Study 1 was analyzed using mixed-effects binomial regressions with Response as dependent variable, indicating whether participants perceived the presented voice morph as resembling their voice; two fixed effects with an interaction term: Conduction (air, bone) and Previous Exposure (yes, no); a fixed effect of Voice Morph (15, 30, 45, 55, 70, 85%). As hypothesized, the analysis revealed main effects of Conduction (estimate=-0.47, $Z=-2.96$, $p=.003$), indicating a higher intercept of the psychometric curve fitted for bone conduction (**Fig. 2**), of Previous Exposure (estimate=-0.5, $Z=-4.64$, $p<.001$), indicating a higher intercept of the psychometric curve fitted for the blocks without previous exposure, and of Voice Morph (estimate=0.55, $Z=22.67$, $p<.001$), indicating that the ratio of 'self' response increased with increased amount of self-voice present in voice morphs (**Fig. 2**). Moreover, the analysis yielded a significant interaction between Conduction and Previous Exposure (estimate=0.43, $Z=2.85$, $p=.004$). In order to investigate the nature of the interaction, we ran a separate mixed-effects binomial regression for each type of Previous Exposure. The analysis for the blocks without previous exposure to self-voice revealed a significant interaction between Voice Morph and Conduction (estimate=-0.25, $Z=-3.46$, $p<.001$), indicating a steeper slope for the psychometric curve fitted for bone conduction, compared to the curve fitted for air conduction (**Fig. 2A**). On the

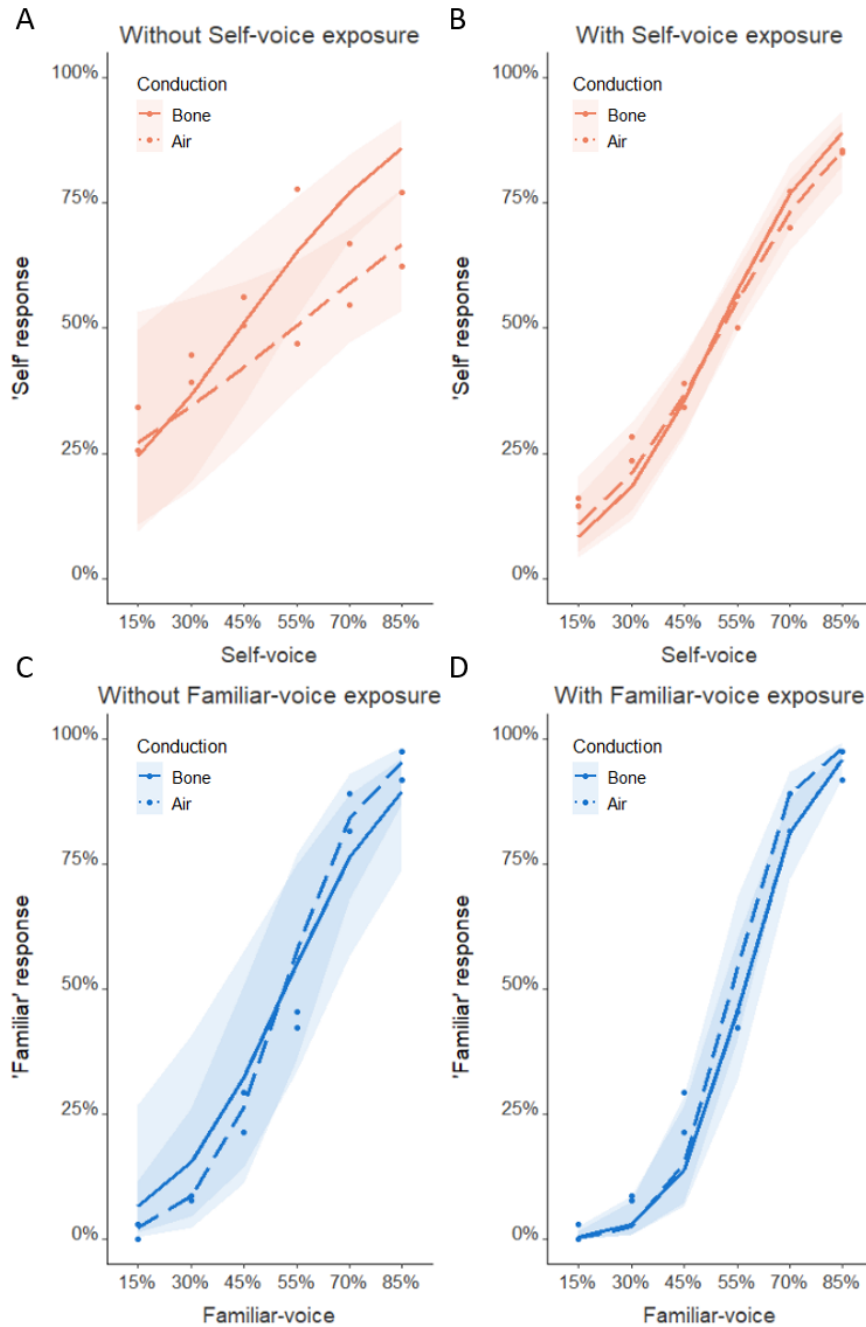


Fig 2. Studies 1 and 2. Psychometric curves fitted for two forms of sound conduction (bone – solid line; air – dashed) during studies 1 (self-other voice discrimination, **A** and **B**) and 2 (familiar-other voice discrimination, **C** and **D**). The abscissa indicates the percentage of the self/familiar voice present in a voice morph and the ordinate indicates the rate at which the corresponding voice morph was perceived as resembling the self/familiar voice. The shaded areas around each curve represent the 95% confidence intervals. Left plots (**A** and **C**) indicate perception for the blocks without and right plots (**B** and **D**) for the blocks with immediate previous exposure to the target voice prior to the task. Bone conduction improved self-unfamiliar discrimination only when participants were not previously exposed to their voice before the task (**A**). No such effects were observed for familiar-unfamiliar discrimination.

contrary, for the blocks with previous exposure to self-voice, the effect of Conduction did not interact with the effect of Voice Morph (estimate=-0.12, $Z=-1.63$, $p=.104$) (**Fig. 2B**). This shows that participants performed the self-other VD task better when stimuli were presented through bone conduction, compared to air conduction, but only without previous exposure to unmorphed self-voice recordings. In both post-hoc analyses, there was a main effect of Voice Morph (no previous exposure: estimate=0.59, $Z=4.39$, $p<.001$; previous exposure: estimate=0.9, $Z=7.28$, $p<.001$) but not of Conduction (no previous exposure: estimate=0.39, $Z=1.38$, $p=.17$; previous exposure: estimate=0.4, $Z=1.4$, $p=.16$). Collectively, the results of Study 1 indicate that participants are better at discriminating self- and other-voices (1) when voice morphs were presented through bone conduction, (2) that previous exposure makes the self-other VD task easier, and (3) that this bone-conduction-related enhanced self-perception disappears when subjects were exposed before the task to their own unmorphed voice stimuli.

Study 2: Familiar-other VD

The observed effect of enhanced perception of the self-voice in Study 1 may be caused by the familiarity with one's own voice and not one's own voice per se (i.e., the other voice was not familiar). Hence, in Study 2, the experimental design (**Fig. 1C**) and statistical analysis were equivalent to Study 1, except that the self-voice was substituted with the voice of a familiar other. Should the effect in Study 1 be caused by a familiarity effect, then we would expect similar performance in Study 2. Thus, participants heard voice morphs between a familiar other voice and an unfamiliar other voice, either via air or via bone conduction, and either without (first two blocks) or with previous exposure to a familiar voice. In each trial, they indicated whether the corresponding morph more closely resembled the familiar voice or someone else's. As hypothesized, the mixed-effects binomial regression in the familiar-other discrimination with Response as dependent variable only revealed a significant effect of Voice Morph (estimate=1.16,

$Z=30.53$, $p<.001$), indicating that the rate of 'familiar' response increased with increased amount of familiar voice present in voice morphs. Also as hypothesized, the effects of Conduction (estimate=-0.03, $Z=-0.22$, $p=.826$) and Previous Exposure (estimate=-0.54, $Z=-1.17$, $p=.242$) were not significant, nor was their interaction (estimate=0.27, $Z=1.41$, $p=.159$). Equivalent models relying on the Bayesian framework revealed evidence in favor of the null hypothesis according to which sound conduction did not affect familiar-other discrimination (Bayes factor = 0.17, see supplementary material). These data show that familiar-other discrimination was not significantly affected by the type of sound conduction or previous exposure to the familiar voice (**Fig. 2C** and **Fig. 2D**). Thus, they suggest that the effects observed in Study 1 involve self-related processes rather than those of familiarity.

Study 3: Self-other, familiar-other, and self-familiar VD

Results of Studies 1 and 2 showed that the bone-conduction effects are specific to self-voice and do not generalize to other familiar voices, respectively, albeit in different groups of participants. Hence, in Study 3, we contrasted self-other and familiar-other VD tasks in the same, independent cohort of participants. Moreover, in the same group of participants, we performed a self-familiar VD task, thereby investigating whether the bone-conduction effects persist irrespective of other-voice familiarity. Thus, Study 3 consisted of two parts (**Fig. 1D**). In the first part, participants performed two blocks of the self-other VD task (Air, Bone, cf. Study 1) and two blocks of the familiar-other task (Air, Bone, cf. Study 2), using the counterbalanced order. This was followed by two blocks of self-familiar voice discrimination task (Air, Bone) that were counterbalanced across participants. Self-familiar blocks were always conducted after self-other and familiar-other blocks to balance the exposure to self- and to familiar-voice for their discrimination from the unfamiliar voice, before they were tested against each other. The effect of sound conduction on each type of voice discrimination (self-other, familiar-other, self-familiar) was analyzed with mixed-effects

binomial regressions with Response as dependent variable and Conduction (air, bone) and Voice Morph (15, 30, 45, 55, 70, 85%), together with a two-way interaction, as fixed effects.

The results of the three discrimination tasks (self-other, familiar-other, self-familiar) are illustrated in **Fig. 3**. Replicating Studies 1 and 2, we observed a significant interaction between the effects of Conduction and Voice Morph, characterized by a steeper psychometric curve for bone compared to air conduction, but only in tasks involving self-voice (self-other, self-familiar) (self-other: estimate=-0.1, $Z=-2.26$, $p=.024$; familiar-other: estimate=-0.05, $Z=-1.12$, $p=.263$; self-familiar: estimate=-0.13, $Z=-2.71$, $p=.008$). Self-related tasks also had a significant effect of

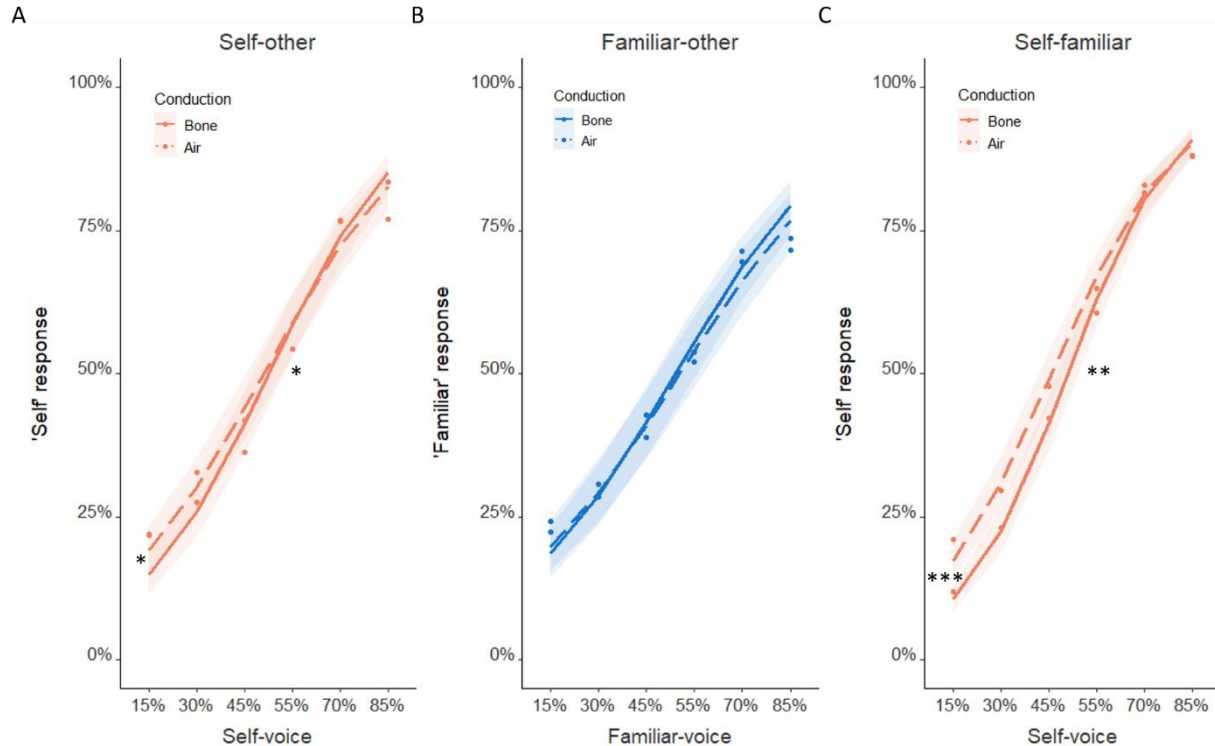


Fig. 3. Study 3: voice discrimination tasks. Psychometric curves fitted for two forms of sound conduction (bone – solid line; air - dashed) during all three voice-discrimination tasks (**A**: Self-other; **B**: Familiar-other; **C**: Self-familiar). The abscissa indicates the percentage of the self/familiar voice present in a voice morph and the ordinate indicates the rate at which the corresponding voice morph was perceived as more resembling the self/familiar voice. The shaded areas around each curve represent the 95% confidence intervals. Asterisks in the lower end of the curves indicate a significant difference in intercepts, whereas asterisks in the middle of the curves indicate a significant difference in slopes. Intercepts were lower and the slope was steeper for the curves fitted for bone conduction, but only in the self-related tasks (**A** and **C**). *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$

Conduction, showing a lower intercept for bone conduction (self-other: estimate=0.4, $Z=2.44$, $p=.015$; familiar-other: estimate=0.12, $Z=0.75$, $p=.452$; self-familiar: estimate=0.7, $Z=3.97$, $p<.001$). Equivalent models relying on the Bayesian framework revealed evidence in favor of the null hypothesis according to which sound conduction did not affect familiar-other discrimination (Bayes factor = 0.21, see supplementary material). As before, all 3 tasks indicated a main effect of Voice Morph, indicating an increase in 'self'/'familiar' answers with an increase with the amount of self/familiar voice present in voice morphs (self-other: estimate=0.7, $Z=22.13$, $p<.001$; familiar-other: estimate=0.56, $Z=19.09$, $p<.001$; self-familiar: estimate=0.89, $Z=25.15$, $p<.001$).

Overall, these data demonstrate that bone conduction improved the performance in voice discrimination tasks if the task involved self-voice morphs, regardless of other-voice familiarity (steeper psychometric curves in self-other and self-familiar, but not in familiar-other task; asterisks in the middle of plots in **Fig. 3**). Lower intercepts for bone conduction (asterisks in the left end of plots in **Fig. 3**) indicate that the accuracy in self-related tasks was more increased for other-dominant voice morphs (i.e., containing lower rate of self-voice present)⁵⁴.

Self-other VD acoustic analysis

We subsequently investigated whether the physical acoustic parameters that have been shown to account for the discrimination of other voices¹⁸ also impact VD for one's own voice. Participants' voices were placed in voice spaces as defined by Baumann and Belin¹⁸, whose axes represent different acoustic parameters of the voices (pitch and formant information, see Methods). In this space, similarly sounding voices are located close to each other and inter-voice distances have been correlated with other-voice discriminability¹⁸ and related to the activity in auditory cortex⁵⁵. A two-dimensional voice space was created^{18,56,57}, with the dimensions corresponding to contributions of source (pitch, larynx) and filter (formants, vocal tract) in voice production⁵⁸ (**Fig.**

4A). The voice spaces were normalized such that the origin of the spaces corresponds to the other voice in each self-other voice pair. The distance to the origin thus represents the acoustic difference between self and other voices. Correlation analysis indicated a positive association between self-other voice distances and self-other task performance ($r=0.2$, 95% CI = [0.01, 0.38]; $t(98)=2.06$, $p=.042$; **Fig. 4B**), indicating that the same acoustic parameters that have been linked to discrimination of other voices¹⁸ account for VD of the self-voice. Neither sound conduction (air, bone) nor the type of other voice (familiar, unfamiliar) affected the relationship between task performance and self-other distance (supplementary material). Further analyses related to acoustic properties – gender differences and contributions of source (larynx) and filter (vocal tract) – are reported in the Supplementary Material.

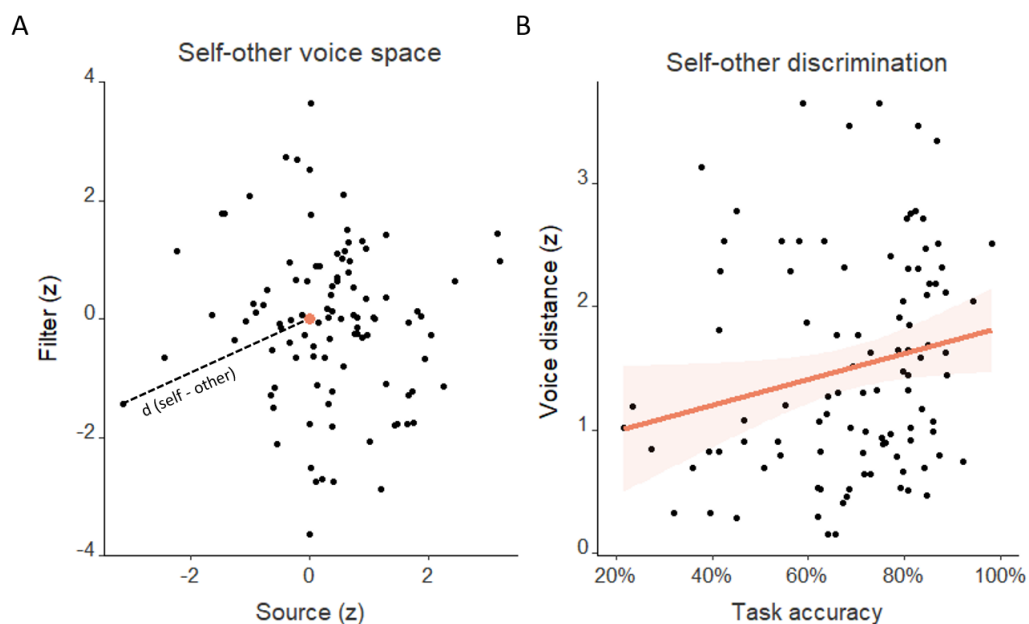


Fig. 4: Self-other voice discrimination – acoustic analysis. **A)** Voice space in which the origin (enlarged orange dot) represents the other voice in self-other discrimination tasks. Distance to the origin (dashed line) thus represents each participant's self-other voice distance in z-score units. **B)** Self-other distances were correlated to the self-other task performances. Shaded area around linear regression indicates 95% confidence interval.

Control Self-voice recognition task

Findings from studies 1-3 indicate that bone conduction facilitates self-other voice discrimination, regardless of other-voice familiarity. However, those findings could be misleading if participants are unable to recognize their voice without voice morphing. Moreover, it remains unclear whether the bone-conduction advantage generalizes to other self-voice tasks, similar to the ones reported in previous work, where no voice morphing was used.

Thus, at the end of Study 3, participants performed a control self-voice recognition task in which, unbeknown to participants, the stimuli consisted only of unmorphed voices (self, familiar, unfamiliar), and, as opposed to the VD tasks, all three voices were used as stimuli within the same experimental block. In each trial, participants were instructed to indicate whether the voice they hear resembled their voice by pressing a button. There were two control-task blocks, one for each form of sound conduction (air, bone), counterbalanced across participants. Each of the three unmorphed voices was randomly repeated 10 times within the block. As this task served as a control to identify whether participants were able to recognize their unmorphed recorded voice, it was always performed at the end of experiment, so as not to affect the performance in the discrimination tasks by previous exposure to unmorphed voice recordings.

The performance in the control task of Study 3 was also analyzed with mixed-effects binomial regressions with Response as dependent variable and two fixed effects with an interaction term: Conduction (air, bone) and Voice (self, familiar, unfamiliar). It showed that there were more 'self' responses in self-voice trials (i.e., hit rate, mean = 0.79, 95% CI = [0.72, 0.85]) both compared to familiar-voice (i.e., familiar false alarm (FA) rate, 0.17, [0.11, 0.23]; estimate=-3.23, Z=-17.48, $p < .001$) and unfamiliar-voice trials (i.e. unfamiliar FA rate, 0.13, [0.08, 0.18]; estimate=-3.44, Z=-17.93, $p < .001$) (**Fig. 5A**). There were no differences in 'self' responses between trials with familiar and unfamiliar voices (i.e., FA rates, estimate=-0.21, Z=-1.1, $p = .273$). The main effect of Conduction was not significant (estimate=0, Z=0.03, $p = .98$) nor was there a Conduction by Voice

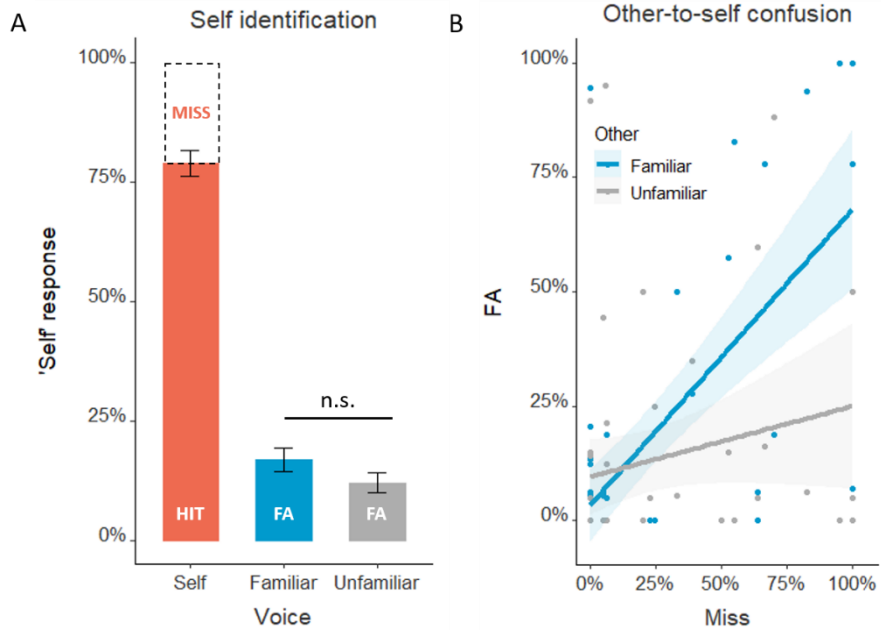


Fig. 5. Study 3: control task. The bar plot (A) indicates mean rates of ‘self’ responses occurring for each type of voice stimuli – hit rate for self and false alarm (FA) rates for familiar and unfamiliar voices – whereas the regression plots (B) indicate relationships between FA rates for familiar and unfamiliar voice with the miss rate for self-voice. Bar-plot whiskers and shaded areas around linear regressions indicate 95% confidence intervals. Although the absolute rate at which familiar and unfamiliar voices were misperceived as self-voice did not differ (A), only the familiar voice misperception was related to self-voice (B).

interaction (estimate=-0.13, Z=-0.46, p=.643). These data show that, although participants were mostly correct in identifying their own untransformed voice (79% hit rate), they also misinterpreted both familiar and unfamiliar other voices as their own during some trials (17% and 13% FAs), indicating that recognizing own voice in a recording even without additional transformations (e.g., morphing with other voices) is not as trivial as it might seem. Thus, nine out of 52 participants who could not recognize their voice in more than half of the self-voice trials (i.e., accuracy lower than 50%) in the control task were considered as outliers and excluded from the analysis in the discrimination tasks above.

Other-to-self voice confusion

The self-voice recognition task indicated that in a non-negligible amount trials participants misperceived either familiar (17%) or unfamiliar (13%) voice as their own (i.e., responded “self” for “other” stimuli – false alarm), but it remains unknown whether the two false alarm rates (unfamiliar FA and familiar FA) are related to each other (i.e., indicating a general ownership over other voices), and to a decrease in recognition of their own voice (i.e., to the miss rate – responding “other” to “self” stimuli – indicating that other voices were confused with self-voice, a self-voice disownership). Thus, we ran correlation analyses between the two false-alarm rates and miss-rate. A significant correlation between the two FAs would indicate that participants who misperceived one type of voice as self-voice also misperceived another, suggesting a general tendency to misperceive other voices as self-voice, regardless of voice familiarity. A significant correlation between miss rate and the type of FA would indicate that another voice is confused as self-voice.

Pearson’s product-moment correlation did not show a significant relationship between the two FA rates ($r=-0.07$, 95% CI = [-0.34, 0.21]; $t(50)=-0.49$, $p=.624$), showing that participants either misperceived the familiar or the unfamiliar voice as their own, but independently. The correlation analysis identified a significant positive relationship ($r=0.67$, 95% CI = [0.48, 0.79]) between miss rate and familiar-FA rate ($t(50)=6.31$, $p<.001$), while there was no significant relationship ($t(50)=1.46$, $p=.151$) between miss rate and unfamiliar-FA rate ($r=0.2$, 95% CI = [-0.07, 0.45]) (**Fig. 5B**). Pearson and Filon’s z-test identified a stronger relationship between miss and familiar-FA compared to unfamiliar-FA rates ($z=2.86$, $p=.004$), indicating that participants were confusing self-voice more to the familiar compared to the unfamiliar voice. No correlations were affected by sound conduction type (supplementary material). This shows that, although familiar-to-self voice FA rate did not occur significantly more than unfamiliar-to-self FA rate (**Fig. 5A**), only familiar-to-

self FA rate was related to the miss rate of self-voice (**Fig. 5B**), indicating that (only) familiar other voice is confused as self-voice.

Discussion

As our main means of verbal communication, our voice, is an integral part of our identity and our self. Although traditionally thought of as a purely auditory signal, self-voice is a multimodal percept that also involves vibrotactile input^{32,33}, at least when we are actively speaking. Perceiving our voice passively, that is presented through air conduction loudspeakers, differs in two ways: 1) it sounds different, since it is not (low-pass) filtered as a result of passing through the skull and 2) it lacks multimodal input resulting from speech production. Conversely, this leads to a reduced ability to recognize ourselves in air-conducted recordings^{16,36,37,41–43}. This has significantly impeded self-voice-related research, rendering it one of the least investigated aspects of self-consciousness¹. While previous work has tried approximating the natural self-voice by applying acoustic transformations^{45–49}, we here focused on the multimodal aspect of the self-voice by presenting stimuli through a commercially available bone-conduction headset. This allowed us to pinpoint perceptual specificities related to self-voice, ranging from low-level acoustic to high-level cognitive aspects such as familiarity and previous exposure.

Vibrotactile stimulation

Studies 1 and 3, demonstrated that self-other VD improves with bone compared to air conduction. As we argue below, this improvement most likely results from vibrotactile signals generated by the bone conduction vibrations as opposed to its low-frequency filtering of the auditory signal or voice familiarity. Importantly, this improvement was specific to the involvement of the self-voice, as bone conduction did not improve familiar-other VD (Studies 2 and 3). This separates the effect of bone-conduction from the effect of voice familiarity: if familiarity were sufficient for the differences in perception in self-other task, then similar differences should have occurred in the familiar-other task. As this was not the case, it is likely that the bone-conduction effects are specific to self-voice, not because self-voice is a stimulus we are familiar with but because it

involves a dedicated neural system associated to the self. Moreover, a lack of perceptual differences between air and bone conduction in the familiar-other task further supports the importance of vibrotactile stimulation accompanying bone conduction, as opposed to physical transformations to self-voice stimuli (e.g., deeper sound due to filter properties of bone tissues). Namely, if only physical transformations were to account for the bone-conduction effect in the self-other task, the opposite effect should be observed in the familiar-other task (i.e., the disadvantage of bone conduction for familiar voices), given that we are not used to hearing familiar voices transformed in such a way (e.g., deeper voices of acquaintances). As equivalent acoustic transformations applied by the same bone conduction headset led to perceptual differences for self-voice and not familiar-voice stimuli, we suggest that concomitant vibrotactile signals, usually exclusive to one's own voice, play an important role in self-voice processing. This suggests that self-voice is a multimodal construct and consolidates evidence in favor of self-related processes in the perception of voices as it has been shown for the perception of faces^{14,59}.

Previous exposure

Our results further demonstrate better performance in self-voice tasks in people with higher previous exposure to self-voice, as has previously been demonstrated e.g., in radio announcers³⁴. By targeting the effect of previous exposure of self-voice in a more controlled way – in half of experimental blocks of Study 1, the unmorphed self-voice was shown to participants before the self-other VD task. This was done in the second half of the experiment as not to affect the performance in the first half of experiment. Confirming previous findings^{34,35}, we found that prior exposure facilitated self-other VD. Participants' performance improved when they heard their unmorphed voice recording immediately before the task: by hearing an unmorphed recording of their voice prior to task execution, participants could have created an arbitrary strategy of recognizing that specific voice recording in a voice morph, regardless of whether they associated the recording with themselves or not. For instance, when hearing two unmorphed voices before

the task, participants could focus on one acoustic property (e.g., a higher pitch in one of the recordings) and use that property as a reference against which voice morphs are compared. Without previous exposure, however, there was no pre-exposure based additional reference that participants could rely on to complete the task, and they had to rely on their internal self-voice representation. The fact that there was no effect of previous exposure on the familiar-other task in Study 2 also suggests that the effect of previous exposure in Study 1 did not occur because it was manipulated only in the second half of the experiment, controlling for task habituation effects.

The addition or omission of pre-exposure to the own voice is also important for understanding the contribution of bone-conduction in Study 1. Bone conduction only improved self-other VD when this discrimination was based on an internal representation of the own-voice as opposed to comparing it to the pre-exposure stimulus, which essentially rendered the task easier. This is supported by the familiar-other VD findings (Study 2), which showed that bone conduction did not affect performance in blocks with familiar versus unfamiliar voices (without and with previous exposure). This suggests that bone conduction facilitation is only found for the self-voice, but not for familiar voices that are mainly based on auditory cues.

Familiarity and acoustic parameters

We performed additional analyses demonstrating that both familiarity processing and acoustic differences contribute to self-other VD. On the one hand, the results of our correlation analysis between miss and false alarm rates in the control self-voice recognition task show that self-voice perception inevitably involves some familiarity processing. Thus, a failure to recognize own voice (miss rate) was correlated with familiar-to-self, and not with unfamiliar-to-self voice misattribution (false alarms), regardless of acoustic similarity between the three voices. This suggests that self-voice was more confused with a familiar than with an unfamiliar other voice and that familiarity mechanisms^{51,52} also bias self-voice perception. An analogy in the visual domain would be

observing that participants confuse their own face more to a familiar as opposed to an unfamiliar other face, despite unfamiliar face being physically more similar (e.g., based on eye colour or nose shape). On the other hand, our voice-space analysis indicates that, to a certain extent, also low-level acoustic properties have an impact on self-voice recognition (for detailed results and discussion, see Supplementary Material). Without “a priori” hypotheses, we placed our participants’ voices in other-centered voice spaces¹⁸ and observed a correlation between acoustic distances and discriminability ratings. This supports the involvement of a third factor, low-level acoustic processes, in self-other VD and shows that the acoustic differences accounting for discrimination of other voices extend to self-voice, that should be further explored (see Supplementary Material). In sum, these findings show that both familiarity mechanisms and acoustic processes contribute to self-voice perception and future studies should identify ways to delineate the corresponding contributions of these factors.

Task sensitivity

While previous self-voice tasks have been characterized by ceiling effects, the paradigm proposed here is able to capture inter-subject variability, which allows us to dissect perceptual specificities (e.g., a bias, general sensitivity, or effects specific to self- or other-voice perception) and personalize studies of self-other VD. Most participants in Studies 1 and 3 (as well as in our follow-up EEG study⁵⁴ spontaneously reported that they perceived the self-other VD task to be very difficult and showed poor metacognition, that is they misjudged their ability of successfully performing the task. Moreover, in Study 1 we observed large differences in performance across participants. That is why, in Study 3, in addition to increasing the sample size based on the power analysis of Study 1 (see Methods), we introduced a control task at the end of experiment, to narrow down the self-other VD analysis to include only those participants able to recognize their own unmorphed voice. To our surprise, nine out of 52 (17.3%) participants could not recognize their unmorphed voice in more than half of self-recognition task’s trials, and were thus excluded

from the analysis in the VD tasks. This shows that recognizing the own voice in short vocalizations (e.g., phoneme /a/ lasting for 500 ms) is not as trivial as it might seem (even without voice morphing), although it is shown to suffice for speaker identification⁶⁰. Moreover, a decrease in difficulty between the control self-voice recognition task and self-other VD task might account for a lack of differences between air and bone conduction in the control task. Namely, it is possible that the bone-conduction advantages for self-voice perception are detectable only in tasks that are sensitive enough to detect them, which was likely not the case for the control task not involving voice morphing.

High task sensitivity also enabled us to observe that the bone-conduction advantage is most advantageous for other-dominant self-other voice morphs (i.e., containing more other-voice features). This was observed in Study 3 in both self-other and self-familiar tasks (left side of psychometric curves on Fig 3), and replicated in a follow-up EEG study with an independent cohort of participants performing the same self-other task with five times more trials⁵⁴. This suggests that, rather than labelling an ambiguous voice as 'self', bone conduction facilitates discarding an ambiguous voice as being 'not self'. In other words, our data show that bone conduction specifically facilitates making a "not self" judgement in scenarios of vocal ambiguity.

Motor signals

It is important to note that during natural self-voice perception there are also motor signals involved, related to speech production, that were not tested here. The presence of such motor signals and the associated intraoral and pharyngeal sensory speech-related cues may also exert additional effects on self-other VD, as opposed to familiar-other voice discrimination. An equivalent study in which participants would hear self-other voice morphs triggered by vocalization onset could investigate the potential role of motor signals, in addition to or as opposed to vibrotactile effects. This would, however, be challenging to investigate, as speech-related motor

signals and the resulting vibrotactile sensations cannot be experimentally separated during natural speech.

Impact and clinical relevance

The impact of this work is threefold. First, by shifting the classical perspective on self-voice away from purely auditory to multimodal, these findings incorporate self-voice into multisensory accounts of self-consciousness^{61–64}. According to these accounts, the sense of self is fundamentally based on the continuous integration of multisensory bodily signals, including tactile, proprioceptive, interoceptive, visual and auditory signals^{10,62,65–67}. Correspondingly, we show that integration of auditory and vibrotactile signals increases the recognition of our voice, that is an integral part of our self. Second, by introducing a method which improves auditory self-identification, we propose a new approach to addressing self-voice-related research questions. Based on these findings, future studies can avoid presenting self-voice stimuli through traditional air-conducting media, especially considering the increasing availability of bone-conduction headsets. Finally, this work could serve as a scaffold for clinical investigations of a very common^{23,24} and highly-distressing^{25,26} psychiatric symptom – auditory-verbal hallucinations (AVH), i.e., “hearing voices” – as they have been proposed to arise as a self-other VD deficit^{19,22,68–70}. Specifically, characterizing differences in self-other VD curves in voice-hearers compared to controls (e.g., quantifying a bias to hear other voice and relating it to clinical measures) could deepen the understanding as well as challenge this prominent account for AVH aetiology. Collectively, our findings demonstrate the importance of bone conduction with respect to self-voice perception and shed new light on the phenomenology of the self by portraying self-voice as a fundamentally multimodal composition in which both familiarity and acoustic properties play a significant role.

Methods

Participants

Studies 1 and 2 each involved 16 participants. In Study 1, seven participants were male (mean age \pm SD: 29.7 \pm 5.5 years old) whereas 8 were male in Study 2 (28.5 \pm 5.5 years old). For Study 3, participants were accompanied by an acquaintance (a friend) of the same gender and similar age, who also participated in the study, and it involved 52 participants (20 male, 26.5 \pm 4.6 years old). All participants were right-handed, reported no hearing deficits, and no history of psychiatric or neurological disorders. They were chosen from the general population and were naïve to the purpose of the study. Participants gave informed consent in accordance with institutional guidelines (protocol 2015-00092, approved by the Comité Cantonal d'Ethique de la Recherche of Geneva) and the Declaration of Helsinki, and received monetary compensation (CHF 20/h).

The sample size for Study 3 was selected based on power analysis of Study 1, which indicated that a sample size of $n = 47$ provides greater than 84% power (95% CI = [75.32, 90.57]) for the interaction between Conduction and Voice Morph with the effect size of -0.12 (100 simulations, $\alpha = 0.05$).

Stimuli

Prior to participating in the studies, participants' voices were recorded while vocalizing the phoneme /a/ for approximately 1 to 2 seconds (Zoom H6 Handy recorder). Each recording was normalized for average intensity (-12 dBFS) and duration (500 milliseconds) and cleaned from background noise (Audacity software). In Studies 1-3, such preprocessed voice recordings were used to generate voice morphs spanning a voice identity continuum between two voices by using TANDEM-STRAIGHT⁵⁰ (e.g. a voice morph can be generated such that it contains 40% of person A's, 60% of person B's voice).

In Study 1, the other voice was a voice of a gender-matched unfamiliar person. In Study 2, the familiar voice belonged to a male person to whom participants were acquainted. In Study 3, participants came with a gender-matched acquaintance who also participated in the study and whose voice served as familiar other voice. The gender-matched unfamiliar voices were the same in all studies.

In Studies 1 and 2, the unmorphed voices in blocks with previous exposure were presented to participants through the same sound conduction type used for that experimental block (Air, Bone).

In Study 3, as air-conduction medium, we used headphones (Bose QC20) instead of laptop loudspeakers (GIGABYTE AORUS x5, Studies 1 and 2). Both air- and bone-conducting headphones were installed on participants' heads before the beginning of the experiment and matched for loudness at lower sound intensities, such that vibrotactile sensations resulting from bone conduction could not be perceived, resulting in participants being unable to determine the source of the auditory stimuli throughout the experiment. This served as a stricter methodology, as it enables a better concordance in sound intensity and spatial location between the bone- and air-conducted stimuli. Despite this difference in air-conduction medium, i.e., loudspeakers (Study 1) vs headphones (Study 3), we observed similar effects in comparison to bone conduction in both studies. In all studies, we used the same Aftershokz Sports Titanium headphones as bone-conducting medium.

In all studies, inter-trial intervals jittered between 1 and 1.5 seconds to avoid predictability of stimulus onset.

All studies were performed in MATLAB 2017b with Psychtoolbox library⁷¹.

Statistical analysis

Statistical tests were performed with R⁷², using the lme4⁷³, lmerTest⁷⁴, and cocor⁷⁵ packages. The results were illustrated using sjplot⁷⁶ and ggplot2⁷⁷ packages. Power analysis was performed with simr⁷⁸ package. Bayesian models were created in Stan computational framework (<http://mc-stan.org/>) accessed with the brms package⁷⁹.

Voice discrimination tasks

In all studies, the data were analyzed with binomial mixed-effects analysis described in the Results section separately for each study. For all mixed-effects regressions in all studies, random effects included a by-participant random intercept and by-participant random slopes for the main 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 participant were considered as outliers and excluded. Additionally, a linear mixed-effects regression with Reaction Times as a dependent variable and the same fixed and random effects was performed for all studies, with the polynomial expansion of the Voice Morph variable to level 2 (supplementary material).

Other-to-self voice confusion

For the control task of Study 3, we additionally explored whether self-voice was more misperceived as the familiar or with the unfamiliar voice. For that purpose, we correlated the rate of 'other' response in the self-voice trials (i.e., miss rate) with the rates of 'self' response in both familiar and unfamiliar voice trials (i.e., false alarm (FA) rate). Pearson and Filon's z-test for comparing two correlations based on dependent groups with overlapping variables⁸⁰ was used to compare these two correlations (miss rate with two types of false alarm rates – familiar-as-self and unfamiliar-as-self misperception). The two FA rates were also correlated with each other.

Where significant, separate correlations were then conducted for and compared between the two forms of sound conduction (air, bone; supplementary material).

Self-other VD acoustic analysis

To investigate acoustical properties related to self-other VD, we correlated the performances in self-related tasks (both self-other and self-familiar) with the self-other distances in the two-dimensional voice space identified by Baumann and Belin¹⁸. Thus, for each voice recording, we extracted the fundamental frequency (F0) and five formants (F1-F5) using Praat software⁸¹ and computed its voice-space coordinates, corresponding to source (x coordinate) and filter (y components of voice production⁵⁸ (males: $x = \log(F0)$, $y = \log(F5 - F4)$; females: $x = \log(F0)$, $y = \log(F1)$). The coordinates were first transformed into z-scores, after which the voice spaces were normalized for the other voice, such that other-voice coordinates were subtracted from self-voice coordinates in each self-other voice pair. This resulted in a coordinate system where Euclidean distance to the origin represented self-other voice distance in z-score units. Z-scoring coordinates enabled us to place all participants (male and female) in the same voice-space. Distances to the origin (self-other voice distances) were then correlated with the percentage of correct responses in self-other VD task. In the same way, we created familiar-other voice space and compared familiar-other distances with familiar-other task performances. Significant correlations were run again for and compared between the two forms of sound conduction (air, bone; supplementary material).

Author contribution

Study concept and design: PO, OK, NF, OB. Acquisition of data: PO. Analysis and interpretation of data: PO, OK, NF, OB. Drafting of the manuscript: PO, OK, OB. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: PO. Obtained funding: OB. Administrative, technical, or material support: All authors. Study supervision: OB.

References

1. Holzman, P. S. & Rousey, C. The voice as a percept. *Journal of Personality and Social Psychology* **4**, 79–86 (1966).
2. Holzman, P. S., Rousey, C. & Snyder, C. On listening to one's own voice: Effects on psychophysiological responses and free associations. *Journal of Personality and Social Psychology* **4**, 432–441 (1966).
3. Lee, M., Drinnan, M. & Carding, P. The reliability and validity of patient self-rating of their own voice quality. *Clinical Otolaryngology* **30**, 357–361 (2005).
4. Daryadar, M. & Raghobi, M. The Effect of Listening to Recordings of One's Voice on Attentional Bias and Auditory Verbal Learning. *International Journal of Psychological Studies* **7**, (2015).
5. Holzman, P. S., Berger, A. & Rousey, C. Voice confrontation: A bilingual study. *Journal of Personality and Social Psychology* **7**, 423–428 (1967).
6. Qin, P., Wang, M. & Northoff, G. Linking bodily, environmental and mental states in the self — A three-level model based on a meta-analysis. *Neuroscience and Biobehavioral Reviews* **115**, 77–95 (2020).
7. Northoff, G. *et al.* Self-referential processing in our brain-A meta-analysis of imaging studies on the self. *Neuroimage* **31**, 440–457 (2006).
8. Sui, J. & Humphreys, G. W. The ubiquitous self: What the properties of self-bias tell us about the self. *Ann N Y Acad Sci* **1396**, 222–235 (2017).
9. Kircher, T. & David, A. S. Self-consciousness: an integrative approach from philosophy, psychopathology and the neurosciences. in *The Self in Neuroscience and Psychiatry* (Cambridge University Press, 2003). doi:10.1017/cbo9780511543708.023.
10. Blanke, O., Slater, M. & Serino, A. Behavioral, Neural, and Computational Principles of Bodily Self-Consciousness. *Neuron* **88**, 145–166 (2015).
11. Uddin, L. Q., Kaplan, J. T., Molnar-Szakacs, I., Zaidel, E. & Iacoboni, M. Self-face recognition activates a frontoparietal “mirror” network in the right hemisphere: An event-related fMRI study. *Neuroimage* **25**, 926–935 (2005).
12. Frassinetti, F., Maini, M., Romualdi, S., Galante, E. & Avanzi, S. Is it Mine? Hemispheric Asymmetries in Corporeal Self-recognition. *Journal of Cognitive Neuroscience* **20**, 1507–16 (2008).
13. Ma, Y. & Han, S. Why we respond faster to the self than to others? An implicit positive association theory of self-advantage during implicit face recognition. *Journal of Experimental Psychology: Human Perception and Performance* **36**, 619 (2010).
14. Platek, S. M., Wathne, K., Tierney, N. G. & Thomson, J. W. Neural correlates of self-face recognition: An effect-location meta-analysis. *Brain Research* **1232**, 173–184 (2008).

15. Nakamura, K. *et al.* Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia* **39**, 1047–1054 (2001).
16. Allen, P. P. *et al.* Neural correlates of the misattribution of self-generated speech. *Human Brain Mapping* **26**, 44–53 (2005).
17. Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q. & Iacoboni, M. The self across the senses: an fMRI study of self-face and self-voice recognition. *Soc Cogn Affect Neurosci* **3**, 218–23 (2008).
18. Baumann, O. & Belin, P. Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research* **74**, 110–120 (2010).
19. Frith, C. D. & Done, D. J. Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychol Med* **19**, 359–363. (1989).
20. Frith, C. D. *The cognitive neuropsychology of schizophrenia*. (Erlbaum, 1992). doi:10.1016/0028-3932(93)90120-O.
21. Allen, P., Aleman, A. & McGuire, P. K. Inner speech models of auditory verbal hallucinations: Evidence from behavioural and neuroimaging studies. *International Review of Psychiatry* **19**, 407–415 (2007).
22. Ford, J. M. & Mathalon, D. H. Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations? *International Journal of Psychophysiology* **58**, 179–189 (2005).
23. Bauer, S. M. *et al.* Culture and the prevalence of hallucinations in schizophrenia. *Comprehensive Psychiatry* **52**, 319–325 (2011).
24. Nayani, T. H. & David, A. S. The auditory hallucination: A phenomenological survey. *Psychological Medicine* **26**, 177–189 (1996).
25. Harkavy-Friedman, J. M. *et al.* Suicide attempts in schizophrenia: the role of command auditory hallucinations for suicide. *The Journal of clinical psychiatry*. **64**, 871–874 (2003).
26. Kelleher, I. *et al.* Psychotic symptoms in adolescence index risk for suicidal behavior: findings from 2 population-based case-control clinical interview studies. *Arch Gen Psychiatry* **69**, 1277–1283 (2012).
27. Békésy, G. V. The Structure of the Middle Ear and the Hearing of One's Own Voice by Bone Conduction. *Journal of the Acoustical Society of America* **21**, 217–232 (1949).
28. Reinfeldt, S., Östli, P., Håkansson, B. & Stenfelt, S. Hearing one's own voice during phoneme vocalization—Transmission by air and bone conduction. *J Acoust Soc Am* **128**, 751–762 (2010).
29. Tonndorf, J. Bone conduction. in *Auditory System* 37–84 (Springer Berlin Heidelberg, 1976). doi:10.1007/978-3-642-66082-5_13.
30. Wheatstone, C. Experiments on audition. *Quarterly Journal of Science, Literature and Art* **24**, 67–72 (1827).

31. Stenfelt, S. Acoustic and physiologic aspects of bone conduction hearing. *Implantable bone conduction hearing aids* 10–21 (2011) doi:10.1159/000323574.
32. Tremblay, S., Shiller, D. M. & Ostry, D. J. Somatosensory basis of speech production. *Nature* **423**, 866–869 (2003).
33. Ito, T., Tiede, M. & Ostry, D. J. Somatosensory function in speech perception. *Proceedings of the National Academy of Sciences* **106**, 1245–1248 (2009).
34. Rousey, C. & Holzman, P. S. Recognition of One's Own Voice. *Journal of Personality and Social Psychology* **6**, 464–466 (1967).
35. Olivos, G. Response delay, psychophysiologic activation, and recognition of one's own voice. *Psychosom Med* **29**, 433–440 (1967).
36. Shuster, L. I. The perception of correctly and incorrectly produced /r/. *Journal of Speech, Language, and Hearing Research* **41**, 941–950 (1998).
37. Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P. & Belin, P. Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition* **68**, 204–214 (2008).
38. Allen, P. P. *et al.* Misattribution of external speech in patients with hallucinations and delusions. *Schizophrenia Research* **69**, 277–287 (2004).
39. Candini, M. *et al.* The lost ability to distinguish between self and other voice following a brain lesion. *NeuroImage: Clinical* **18**, 903–911 (2018).
40. Candini, M. *et al.* Who is speaking? Implicit and explicit self and other voice recognition. *Brain and Cognition* **92**, 112–117 (2014).
41. Hughes, S. M. & Nicholson, S. E. The processing of auditory and visual recognition of self-stimuli. *Consciousness and Cognition* **19**, 1124–1134 (2010).
42. Schuerman, W. L., Meyer, A. & McQueen, J. M. Do we perceive others better than ourselves? A perceptual benefit for noise-vocoded speech produced by an average speaker. *PLoS ONE* **10**, 1–18 (2015).
43. Gur, R. C. & Sackeim, H. A. Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology* **37**, 147–169 (1979).
44. Douglas, W. & Gibbins, K. Inadequacy of voice recognition as a demonstration of self-deception. *Journal of Personality and Social Psychology* **44**, 589–592 (1983).
45. Vurma, A. The timbre of the voice as perceived by the singer him-/herself. *Logopedics Phoniatrics Vocology* **39**, 1–10 (2014).
46. Won, S. Y., Berger, J. & Slaney, M. Simulation of One's Own Voice in a Two-parameter Model. *Proceedings of the International Conference on Music Perception and Cognition (ICMPC)* (2014).
47. Shuster, L. I. & Durrant, J. D. Toward a better understanding of the perception of self-produced speech. *Journal of Communication Disorders* **36**, 1–11 (2003).

48. Kimura, M. & Yotsumoto, Y. Auditory traits of “own voice.” *PLoS ONE* **13**, 1–16 (2018).
49. Maurer, D. & Landis, T. Role of bone conduction in the self-perception of speech. *Folia Phoniatrica et Logopaedica* **42**, 226–229 (1990).
50. Kawahara, H., Morise, M., Banno, H. & Skuk, V. G. 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* 1–10 (IEEE, 2013). doi:10.1109/APSIPA.2013.6694355.
51. Sidtis, D. & Kreiman, J. In the Beginning Was the Familiar Voice: Personally Familiar Voices in the Evolutionary and Contemporary Biology of Communication. *Integrative Psychological and Behavioral Science* **2011** 46:2 **46**, 146–159 (2011).
52. Stevenage, S. V. Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. *Neuropsychologia* **116**, 162–178 (2018).
53. Orepic, P., Rognini, G., Kannape, O. A., Faivre, N. & Blanke, O. Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners. *Schizophrenia Research* **231**, (2021).
54. Iannotti, G. R. *et al.* EEG Spatiotemporal Patterns Underlying Self-other Voice Discrimination. *Cerebral cortex* bhab329 (2021) doi:10.1093/CERCOR/BHAB329.
55. Latinus, M., McAleer, P., Bestelmeyer, P. E. G. & Belin, P. Norm-based coding of voice identity in human auditory cortex. *Current Biology* **23**, 1075–1080 (2013).
56. López, S. *et al.* Vocal caricatures reveal signatures of speaker identity. *Scientific Reports* **3**, 1–7 (2013).
57. Skuk, V. G. & Schweinberger, S. R. Gender differences in familiar voice identification. *Hearing Research* **296**, 131–140 (2013).
58. Ghazanfar, A. A. & Rendall, D. Evolution of human vocal production. *Current Biology* **18**, R457–R460 (2008).
59. Alzueta, E., Melcón, M., Poch, C. & Capilla, A. Is your own face more than a highly familiar face? *Biological Psychology* **142**, 100–107 (2019).
60. Zarate, J. M., Tian, X., Woods, K. J. P. & Poeppel, D. Multiple levels of linguistic and paralinguistic features contribute to voice recognition. *Scientific Reports* **5**, 11475 (2015).
61. Blanke, O. Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience* **13**, 556–571 (2012).
62. Park, H. D. & Blanke, O. Coupling Inner and Outer Body for Self-Consciousness. *Trends in Cognitive Sciences* **23**, 377–388 (2019).
63. Tsakiris, M. The multisensory basis of the self: From body to identity to others. *The Quarterly Journal of Experimental Psychology* **70**, 597–609 (2017).
64. Petkova, V. I. *et al.* From part- to whole-body ownership in the multisensory brain. *Current Biology* **21**, 1118–1122 (2011).

65. Blanke, O. & Metzinger, T. Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences* **13**, 7–13 (2009).
66. Gallagher, S. Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences* **4**, 14–21 (2000).
67. Jeannerod, M. The mechanism of self-recognition in humans. *Behavioural Brain Research* **142**, 1–15 (2003).
68. Blakemore, S. J., Smith, J., Steel, R., Johnstone, E. C. & Frith, C. D. The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine* **30**, 1131–1139 (2000).
69. Moseley, P., Fernyhough, C. & Ellison, A. Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neuroscience and Biobehavioral Reviews* **37**, 2794–2805 (2013).
70. Whitford, T. J. Speaking-Induced Suppression of the Auditory Cortex in Humans and Its Relevance to Schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **4**, 791–804 (2019).
71. Kleiner, M. *et al.* What's new in Psychtoolbox-3? *Perception* **36**, 1–16 (2007).
72. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing* Preprint at (2020).
73. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* (2015) doi:10.18637/jss.v067.i01.
74. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**, 1–26 (2018).
75. Diedenhofen, B. & Musch, J. Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoS ONE* **10**, 1–12 (2015).
76. Lüdtke, D. sjPlot: Data Visualization for Statistics in Social Science. R package version 2.6.2. Preprint at <https://doi.org/10.5281/zenodo.1308157> (2018).
77. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York* Preprint at (2016).
78. Green, P. & Macleod, C. J. SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution* **7**, 493–498 (2016).
79. Bürkner, P. C. brms: An R package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software* **80**, (2017).
80. Pearson, K. & Filon, L. N. G. Mathematical Contributions to Theory of Evolution: IV. On the Probable Errors of Frequency Constants and on the Influence of Random Selection and Correlation. *Proceedings of the Royal Society of London* **62**, 229–311 (1898).
81. Boersma, P. & Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 6.1.12. <http://www.praat.org/> (2020).

Supplementary material:

Talking Heads: Bone conduction facilitates self-other voice discrimination

Pavo Orepic¹, Oliver Alan Kannape², Nathan Faivre³, Olaf Blanke^{1, 4}

¹ Laboratory of Cognitive Neuroscience, Center for Neuroprosthetics and Brain Mind Institute, Faculty of Life Sciences, Swiss Federal Institute of Technology (EPFL), Switzerland

² Virtual Medicine Center, University Hospital of Geneva, Geneva, Switzerland

³ University Grenoble Alpes, University Savoie Mont Blanc, CNRS, LPNC, Grenoble, France.

⁴ Department of Clinical Neurosciences, University Hospital of Geneva, Geneva, Switzerland

Corresponding author

Olaf Blanke

Bertarelli Chair in Cognitive Neuroprosthetics, Center for Neuroprosthetics & Brain Mind Institute, School of Life Sciences, Campus Biotech, Swiss Federal Institute of Technology (EPFL), 1012 Geneva, Switzerland

E-mail: olaf.blanke@epfl.ch

Tel: +41 (0)21 693 69 21

Fax: +41 (0)21 693 69 22

Bayesian analyses of the null effects

To further validate null findings of the mixed-effects binomial regressions reported in the main text, we ran two main regressions in a Bayesian framework for the absence of the sound conduction effect in familiar-other task (1) for Study 2, and (3) for Study 3. All Bayesian models were created in Stan computational framework (<http://mc-stan.org/>) accessed with the brms package¹, based on four chains of 10000 iterations including 2000 warmup samples. We report the highest density probability for all estimates, which specifies the range covering the 95% most credible values of the posterior estimates.

For the null effect of Conduction in the familiar-other task of Study 2 we had a prior assumption of a better performance with bone conduction (i.e., prior on the main effect Conduction with Gaussian distribution of mean = -0.47 and SD = 0.16, based on the same main effect observed in Study 1 for the self-other task). No main effect of Conduction was found (estimate = -0.05, highest posterior density interval = [-0.31 0.21], Bayes factor = 0.17). For the null effect of Conduction in the familiar-other task of Study 3 we had a prior assumption of a better performance with bone conduction (i.e., prior on the main effect Conduction with Gaussian distribution of mean = 0.4 and SD = 0.16, based on the same main effect observed in Study 3 for the self-other task). No main effect of Conduction was found (estimate = 0.14, highest posterior density interval = [-0.16 0.44], Bayes factor = 0.21). Bayes factors smaller than 1 support the null hypothesis, according to which sound conduction did not familiar-other voice discrimination in Studies 2 and 3.

As exploratory analysis, both for Studies 2 and 3, we ran equivalent regressions in a Bayesian framework with a prior assumption of better performance with bone conduction, but represented as a steeper slope (i.e., prior on the interaction between the effects of Conduction and Voice Morph). For the regression in Study 2, the prior had a Gaussian distribution of mean = 0.43 and

SD = 0.15 (based on the same interaction observed in the self-other task of Study 1). No interaction between Conduction and Voice Morph was found (estimate = 0.29, highest posterior density interval = [-0.06 0.63], Bayes factor = 0.85). For the regression in Study 3, the prior had a Gaussian distribution of mean = -0.1 and SD = 0.04 (based on the same interaction observed in the self-other task of Study 3). No interaction between Conduction and Voice Morph was found (estimate = -0.05, highest posterior density interval = [-0.13 0.03], Bayes factor = 0.3).

Reaction times

Both in studies 1 and 2, the linear mixed-effect regression on response times identified the main effect of Previous Exposure (1: estimate=-2.03, $t(3574)=-14.4$, $p<.001$; 2: estimate=-0.38, $t(15.98)=-2.99$, $p=.009$). Participants responded faster when they previously heard the target voice. In none of the four tasks of Study 3 (self-other, familiar-other, self-familiar, control) were

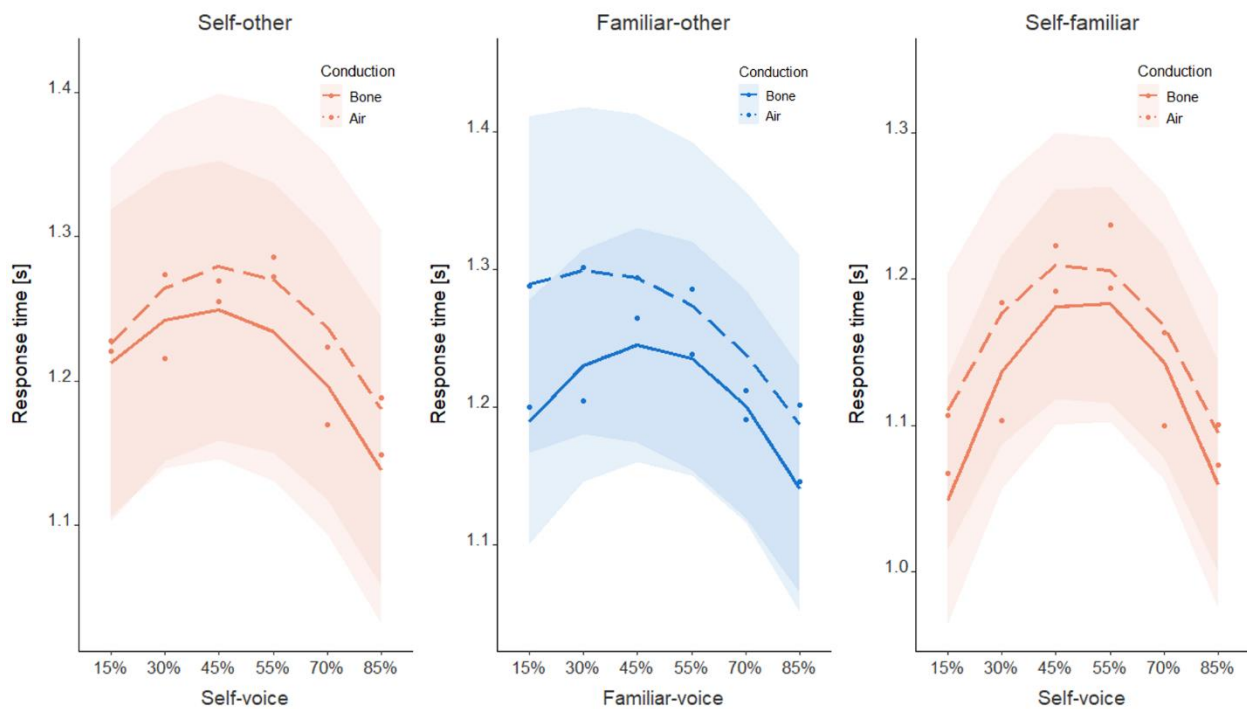


Fig. S1. Reaction times for the three tasks of Study 3. Shaded areas represent 95% confidence intervals. Sound conduction did not affect reaction times in any task.

reaction times significantly affected by the effect of Conduction, nor was there a significant interaction with the effect of Voice Morph. Polynomial expansion of Voice Morph was significant for all voice discrimination tasks, indicating the ‘inverse-u’ shape of reaction times with the increasing levels of Voice Morph. The details of the models are given in Supplementary Table 1 and the results of the discrimination tasks are illustrated on Supplementary **Figure S1**.

Table S1. An overview of the linear mixed-effects models with Reaction Times as dependent variable for all tasks of Study 3.

	Voice Morph		-1.92	4718.88	-3.38	< 0.001	
	Conduction	*	Voice	-0.15	4718.68	-0.18	0.854
	Morph						
Familiar-other	Conduction		0.04	42.98	1.34	0.188	
	Voice Morph		-3.73	4825.44	-6.49	< 0.001	
	Conduction	*	Voice	0.66	4825.6	0.81	0.417
	Morph						
Self-familiar	Conduction		0.05	43	1.19	0.239	
	Voice Morph		-2.17	4698.25	-3.75	< 0.001	
	Conduction	*	Voice	0.87	4697.89	1.07	0.287
	Morph						
Control	Conduction		0.02	2719.91	0.94	0.345	
	Voice		-0.02	104.81	-0.72	0.476	
	Conduction * Voice		-0.02	2717.07	-0.78	0.434	

Other-to-self voice confusion

Pearson and Filon's z-test did not indicate significant differences in correlations between miss and familiar-FA rates with bone and air conduction ($z=-0.74$, $p=.462$, **Figure S2**, left). No such differences were observed for unfamiliar-FAs ($z=1$, $p=.314$, Supplementary **Figure S2**, right).

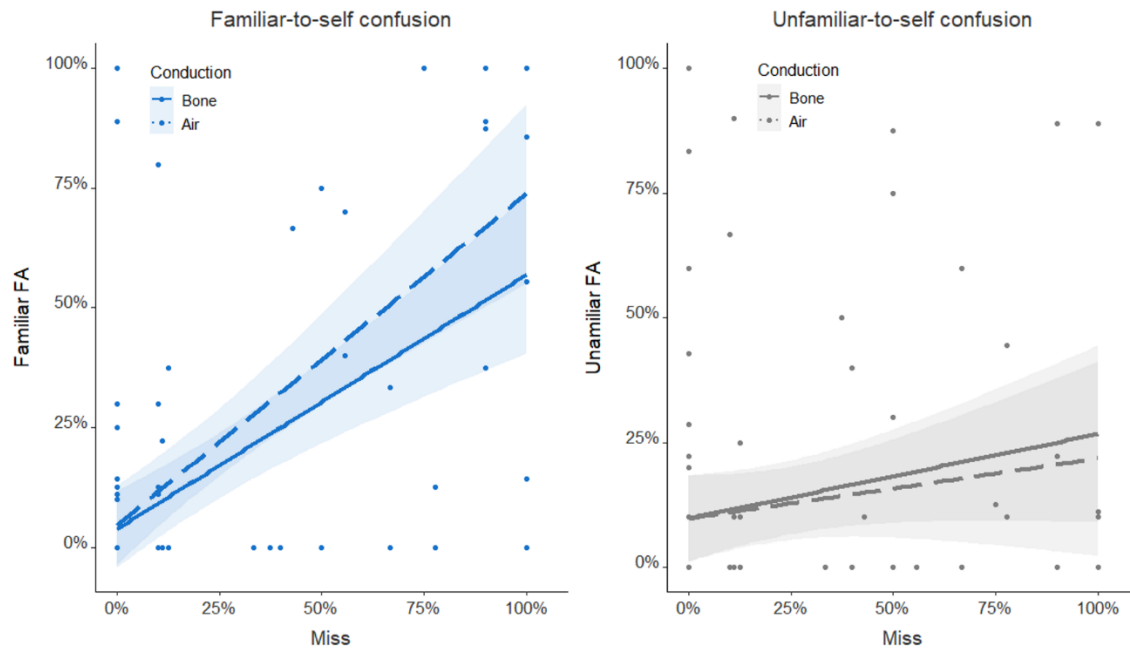


Fig. S2. Self-confusion effects were unaffected by sound conduction. Regression plots indicate relationships between false alarm rates for familiar (left) and unfamiliar (right) voice with the miss rate for self-voice, separately for air and bone conduction. Shaded areas around linear regressions indicate 95% confidence intervals. Neither familiar-FA nor unfamiliar-FA rates were affected by sound conduction.

Self-other VD acoustic analysis

As the voice-space differs for male and female voices², we placed our voice recordings in the corresponding gender-based space (**Figure S3A** and **S3C**) and correlated voice distances with

task performance separately for each gender (**Figure S3B** and **S3D**). (In the main text we report the same analysis with all participants – male and female – placed in the same voice-space).

Correlation analysis indicated a positive association between self-other voice distances and self-other task performance for male participants ($r=0.41$, 95% CI = [0.11, 0.64]; $t(38)=2.76$, $p=.009$; **Figure S3B**). This was not found for female voices ($r=-0.02$, 95% CI = [-0.27, 0.23]; $t(58)=-0.17$, $p=.869$; **Figure S3D**).

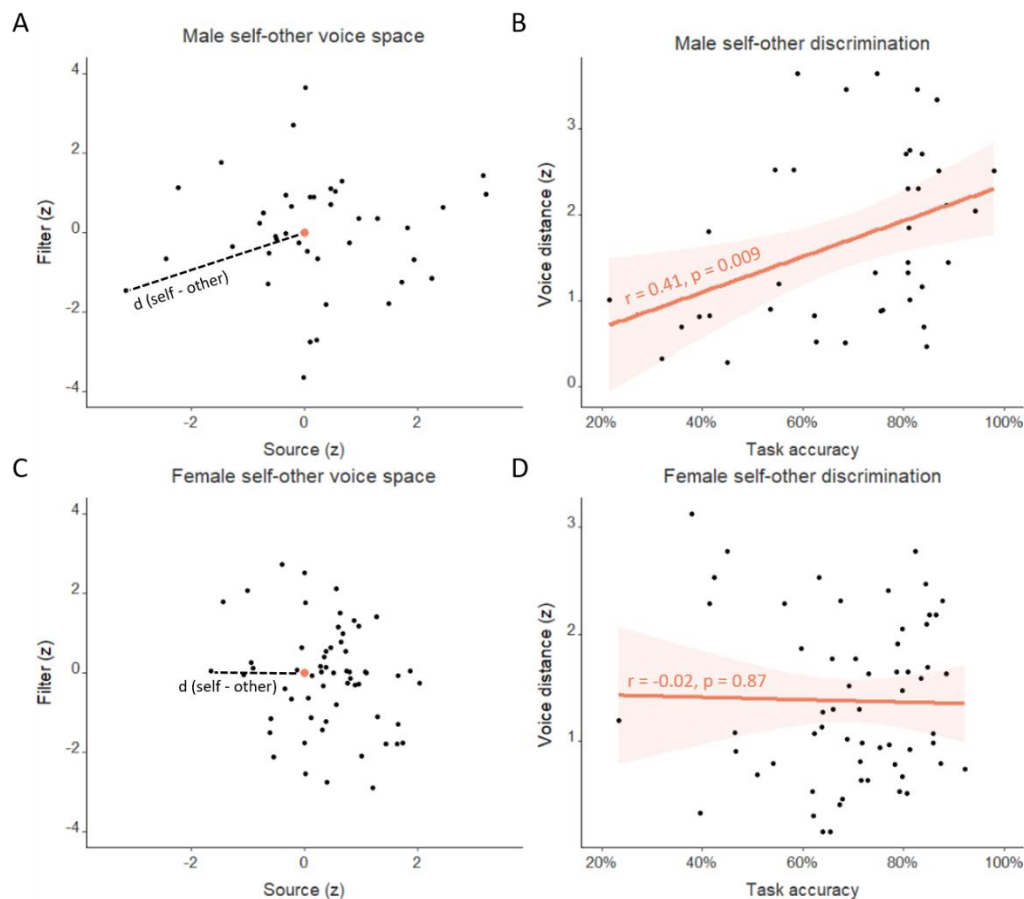


Fig. S3: Self-other voice discrimination – acoustic analysis. Upper plots (**A**, **B**) refer to male and lower plots (**C**, **D**) to female participants. Left plots (**A**, **C**) indicate voice spaces in which the origin (enlarged orange dot) represents the other voice in self-other discrimination tasks. Distance to the origin (dashed line) thus represents each participant's self-other voice distance in z-score units. On the right plots (**B**, **D**), self-other distances were correlated to the self-other task performances. Shaded areas around linear regressions indicate 95% confidence intervals. Positive relationship between voice distances and task performances was found only for male participants (**B**).

We conducted two exploratory analyses investigating the effects of sound conduction and familiarity of other voice on the relationship between voice-space distances and task performance. None were significant (all $p > 0.05$ in Pearson and Filon's z-test). (Supplementary **Figure S4**).

We further split the voice distances to source and filter components as, although both carry identity information^{3,4}, the filter component has explicitly been related to speaker identification⁵ and further been shown to contribute to self-voice recognition⁶. We thus investigated whether the filter component would account more strongly for self-other, as compared to familiar-other discrimination. Pearson and Filon's z-test revealed a difference in relationships ($z = 2.41$, $p = .016$) between filter-distances and self-other task ($r = 0.35$, $p = .132$) and between filter-distances and familiar-other task ($r = -0.35$, $p = .164$) (**Figure S5B**), which did not occur for source-distances (self-other: $r = 0.33$, $p = .156$; familiar-other: $r = 0.39$, $p = .086$; $z = -0.22$, $p = .819$) (**Figure S5A**). In addition, the observed difference between self-other and familiar-other tasks in filter-component correlations was more pronounced for bone conduction (self-other: $r = 0.36$, $p = .115$; familiar-other: $r = -0.41$, $p = .07$; $z = 2.89$, $p = 0.004$), as compared to air conduction (self-other: $r = 0.29$, $p = .216$; familiar-other: $r = -0.19$, $p = .416$; $z = 1.64$, $p = 0.1$) (**Figure S5C**).

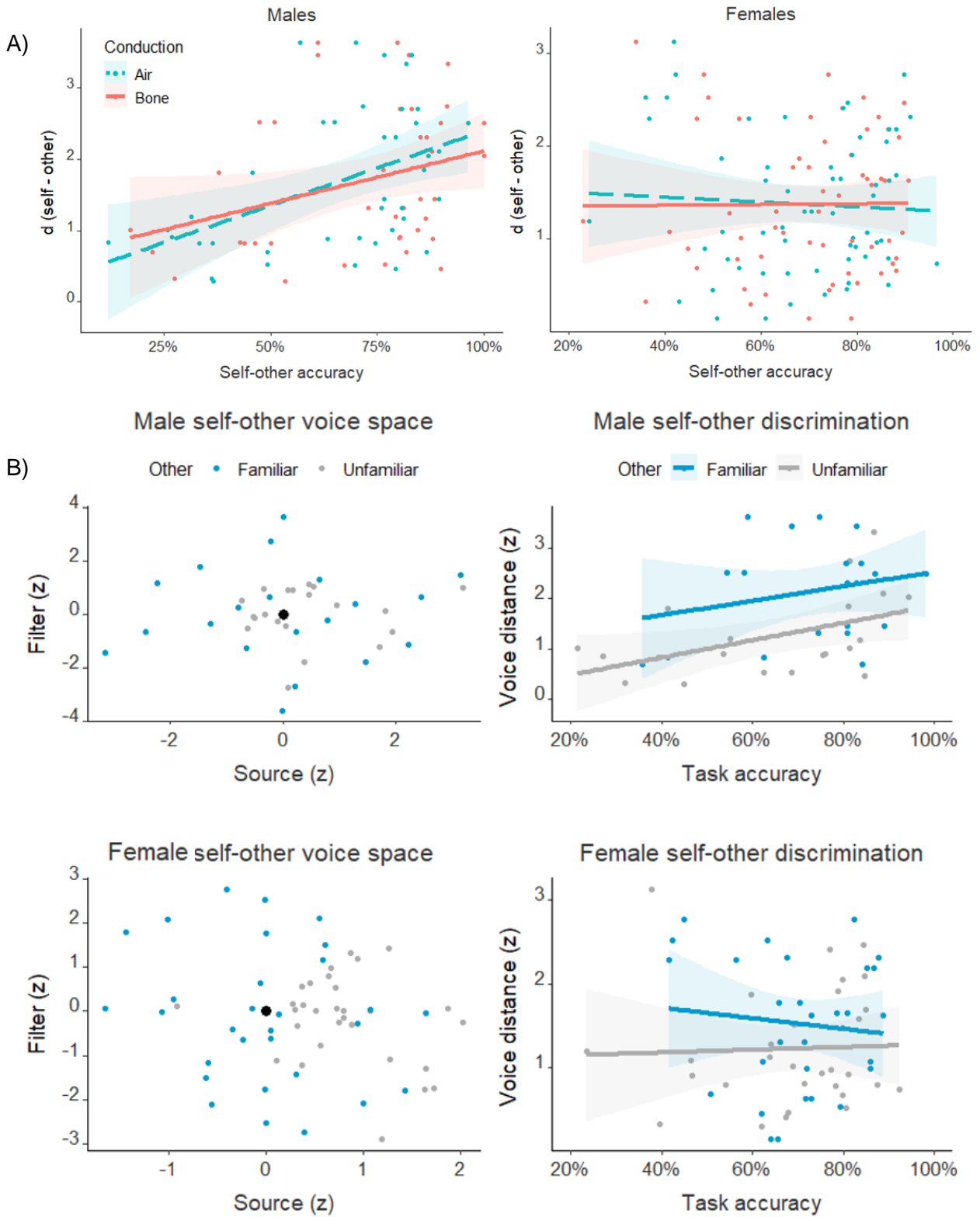


Fig. S4. Neither sound conduction (A) nor other-voice familiarity (B) did affect the relationship between task performance and voice-space distances.

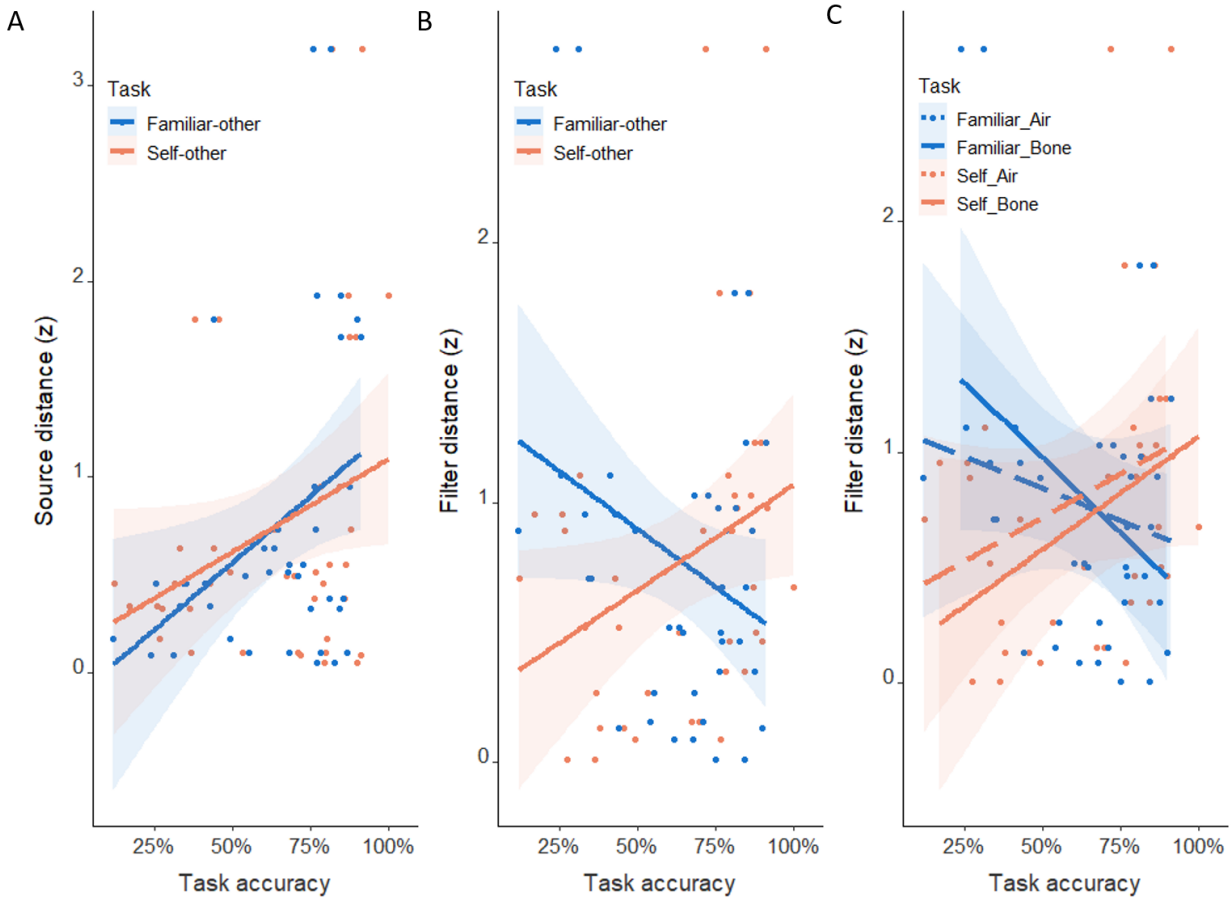


Fig. S5. Different contributions of source and filter components of the source-filter voice space to self-other and familiar-other discriminations. Accuracy of the two discriminations was correlated with source (A) and filter (B) components of voice distances. Shaded areas represent 95% confidence intervals. The difference in relationships was significant only for the filter component (B), which was more prominent for bone, as compared to air conduction (C).

Together, these results indicate that the same acoustic parameters that have been linked to discrimination of other voices² account for VD of the self-voice, and that acoustic parameters related to the filter component of voice production⁷ are related to self-other VD differently than to familiar-other discrimination. However, this was shown only for male voices, possibly because male-voice distances had a higher standard deviation (1.02, z-score) compared to female-voice distances (0.75), although mean voice distance was similar for both genders (male = 1.69, female

= 1.37). Female voices have been shown to be less distinctive in terms of auditory properties⁸ and a lack of correlation for female voices might thus simply arise from a reduced variability in female voice distances (which was indeed present in our data; see **Figure S3**).

Our acoustic analysis further supports the importance of higher formant information for self-voice and extend the understanding of the different roles acoustic properties of voice signals play for speaker identification. By indicating a stronger contribution of the second as compared to the first voice-space dimension to self-other VD, we further extend previous reports⁶ of higher formant information being specific for self-voice. Namely, the two dimensions in Baumann and Belin's voice space^{2,5,8} correspond to contributions of source (pitch and lower formants, larynx) and filter (higher formants, vocal tract) in the source-filter model of voice production⁷, respectively. Although both are related to speaker's identity^{3,4}, lower formants can be altered by the speaker⁹, whereas higher formants cannot, as they reflect the morphology of individual vocal cavities¹⁰. Higher formants were thus more closely related to speaker's identity⁵ and were further shown to play an important role in self-voice recognition⁶. In sum, these findings show that both familiarity mechanisms and acoustic processes contribute to self-voice perception and future studies should identify ways to delineate the corresponding contributions of these factors.

Supplementary references

1. Bürkner, P. C. brms: An R package for Bayesian Multilevel Models using Stan. *Journal of Statistical Software* **80**, (2017).
2. Baumann, O. & Belin, P. Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research* **74**, 110–120 (2010).
3. Latinus, M. & Belin, P. Perceptual auditory aftereffects on voice identity using brief vowel stimuli. *PLoS ONE* **7**, 1–7 (2012).
4. Schweinberger, S. R. *et al.* Auditory Adaptation in Voice Perception. *Current Biology* **18**, 684–688 (2008).
5. López, S. *et al.* Vocal caricatures reveal signatures of speaker identity. *Scientific Reports* **3**, 1–7 (2013).
6. Xu, M., Homae, F., Hashimoto, R. & Hagiwara, H. Acoustic cues for the recognition of self-voice and other-voice. *Frontiers in Psychology* **4**, 1–7 (2013).
7. Ghazanfar, A. A. & Rendall, D. Evolution of human vocal production. *Current Biology* **18**, R457–R460 (2008).
8. Skuk, V. G. & Schweinberger, S. R. Gender differences in familiar voice identification. *Hearing Research* **296**, 131–140 (2013).
9. Maeda, S. Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. in *Speech production and speech modelling* 131–149 (Springer, Dordrecht, 1990).

10. Kitamura, T. & Akagi, M. Speaker individualities in speech spectral envelopes. *Journal of the Acoustical Society of Japan (E) (English translation of Nippon Onkyo Gakkaishi)* **16**, 283–289 (1995).