RESEARCH REPORT

EJN European Journal of Neuroscience FENS WILEY

Expectancy changes the self-monitoring of voice identity

Joseph F. Johnson[1] (ID) | **Michel Belyk[2]** | **Michael Schwartze[1]** | **Ana P. Pinheiro[3]** | **Sonja A. Kotz[1,4]**

[1]Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, the Netherlands

[2]Division of Psychology and Language Sciences, University College London, London, UK

[3]Faculdade de Psicologia, Universidade de Lisboa, Lisbon, Portugal

[4]Department of Neuropsychology, Max Planck Institute for Human and Cognitive Sciences, Leipzig, Germany

Correspondence
Sonja A. Kotz, Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, the Netherlands.
Email: sonja.kotz@maastrichtuniverstiy.nl

Funding information
Fundação Bial, Grant/Award Number: BIAL 238/16; Fundação para a Ciência e a Tecnologia, Grant/Award Number: PTDC/MHC-PCN/0101/2014; Universiteit Maastricht, Grant/Award Number: F8000E14, F8000F14, F8042 and F8051

Abstract

Self-voice attribution can become difficult when voice characteristics are ambiguous, but functional magnetic resonance imaging (fMRI) investigations of such ambiguity are sparse. We utilized voice-morphing (self-other) to manipulate (un-)certainty in self-voice attribution in a button-press paradigm. This allowed investigating how levels of self-voice certainty alter brain activation in brain regions monitoring voice identity and unexpected changes in voice playback quality. FMRI results confirmed a self-voice suppression effect in the right anterior superior temporal gyrus (aSTG) when self-voice attribution was unambiguous. Although the right inferior frontal gyrus (IFG) was more active during a self-generated compared to a passively heard voice, the putative role of this region in detecting unexpected self-voice changes during the action was demonstrated only when hearing the voice of another speaker and not when attribution was uncertain. Further research on the link between right aSTG and IFG is required and may establish a threshold monitoring voice identity in action. The current results have implications for a better understanding of the altered experience of self-voice feedback in auditory verbal hallucinations.

KEYWORDS

auditory feedback, fMRI, motor-induced suppression, source attribution, voice morphing

1 | INTRODUCTION

The self-monitoring of the voice relies on comparing what we expect to hear and what we actually hear (Frith, 1992; Wolpert & Kawato, 1998). However, in a dynamic environment sensory feedback is often ambiguous, e.g., when listening to multiple speakers. Any judgment of the voice source

further depends on how much sensory feedback deviates from expectations (Feinberg, 1978). Minor deviations regarding one's own voice are typically self-attributed and used to compensate motor control. Major deviations may lead to source-attributing the voice to another person. The study of misattributed self-voice is often associated with auditory verbal hallucinations (AVH) in patients with psychotic disorders

(Kumari, Fannon, et al., 2010; Sapara et al., 2015). However, under conditions of ambiguous feedback, healthy individuals also display uncertainty in attributing the source of their own voice (Asai & Tanno, 2013; Pinheiro et al., 2019). Functional neuroimaging studies of self-voice monitoring have examined the neural substrates of self-other voice attribution but have so far not examined responses to uncertainty in ambiguous conditions (e.g., Allen et al., 2006; Fu et al., 2006). However, we need to better understand how the brain establishes correct self and other voice attribution and where and how the voice is processed in uncertain conditions to better understand the mechanisms underlying dysfunctional self-monitoring.

Previous research has reported that unaltered self-voice production leads to reduced functional brain activity in the auditory cortex (Christoffels et al., 2007). This motor-induced suppression (MIS) is compatible with the findings of numerous studies employing diverse methodology. It is similar to the N1 suppression effect, modulation of the event-related potential of the electroencephalogram (EEG) (e.g., Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005; Pinheiro et al., 2018; Sitek et al., 2013; Wang et al., 2014), or M1 suppression in magnetoencephalography (Houde et al., 2002; Numminen et al., 1999; Ventura et al., 2009), weakened activity in electrocorticography and at intracranial electrodes (Chang et al., 2013; Greenlee et al., 2011), or direct- and inter-cell recordings in non-human primates (Eliades & Wang, 2008; Müller-Preuss & Ploog, 1981).

In addition to suppressed activity in the auditory cortex, self-voice monitoring activates a widespread system of functionally connected brain regions, including cortical motor and speech planning areas as well as subcortical regions such as the thalamus and cerebellum (Behroozmand et al., 2015; Christoffels et al., 2007). Moreover, within the auditory cortex different regions contribute specialized roles in the processing of voice. Notably, the left lateral temporal cortex demonstrates a larger role in speech-related processing, while the right lateral temporal cortex plays an essential role in speaker-related features of voice (Belin et al., 2002; Ethofer et al., 2006, 2007; Formisano et al., 2008; Grandjean et al., 2005; Kotz et al., 2003; Moerel et al., 2012; Schirmer & Kotz, 2006; Wiethoff et al., 2008). In the current study, we focused on the perception of voice identity particularly ascribed to the right anterior superior temporal gyrus (aSTG) and the adjacent upper bank of the superior temporal sulcus (STS) (Belin et al., 2004; Belin & Zatorre, 2003; von Kriegstein et al., 2003; von Kriegstein & Giraud, 2004). Patient studies support this assumption as lesions or damage to the aSTG can lead to deficits in voice identity recognition (Gainotti et al., 2010; Gainotti & Marra, 2011; Hailstone et al., 2011; van Lancker & Canter, 1982; van Lancker & Kreiman, 1987).

Motor-induced suppression in voice monitoring is not only effective in voice production but also in response to voice recordings activated via a button press (Ford et al., 2007; Knolle et al., 2019; Pinheiro et al., 2018; Whitford et al., 2011) as well as for non-verbal sounds including tones (e.g., Aliu et al., 2009; Baess et al., 2009; Knolle et al., 2013). Furthermore, MIS seems to operate across modalities of sensory feedback and arises from various motor effectors (e.g., Blakemore et al., 1998; Leube et al., 2003; Miall & Wolpert, 1996; Wolpert & Kawato, 1998). One explanation for MIS is that the internal model of an expected action outcome is fed-forward to the relevant cortical regions to cancel out impending activity to the anticipated stimulus (Jordan & Rumelhart, 1992; Miall & Wolpert, 1996; Wolpert, 1997). Studies that experimentally manipulated sensory feedback created a mismatch between expected and actual outcomes and indicated concomitant modulation or absence of MIS under such circumstances. EEG studies typically show decreased N1 suppression (e.g., Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005), while fMRI studies report a relative increase of STG activity when expected feedback is altered (Christoffels et al., 2007, 2011; Fu et al., 2006; McGuire et al., 1996; Zheng et al., 2010). With this approach, it is not only possible to make listeners uncertain about self- or other-voice attribution (Allen et al., 2004; Allen et al., 2005; Allen et al., 2006; Fu et al., 2006; Vermissen et al., 2007), but to also lead listeners to incorrectly attribute self-voice to another speaker (Allen et al., 2004, 2005, 2006; Fu et al., 2006; Johns et al., 2001, 2003, 2006; Kumari, Antonova, et al., 2010; Kumari, Fannon, et al., 2010; Sapara et al., 2015). STG suppression only persists when the voice is correctly judged as self-voice in distorted feedback conditions (Fu et al., 2006). Critically, data reflecting uncertain voice attribution are often removed from fMRI analyses (Allen et al., 2005; Fu et al., 2006). However, to gain a better understanding of voice attribution to internal or external sources, it is mandatory to specify such data and to define how the known voice attribution region of the STG reacts to uncertainty.

Next to the auditory cortex, activation in the right inferior frontal gyrus increases in response to distorted auditory feedback (Johnson et al., 2019). However, while attenuation of the right aSTG activation reflects expected voice quality, the right IFG is selectively responsive to unexpected sensory events (Aron et al., 2004). Increased right IFG activity has been reported when voice feedback is acoustically altered (Behroozmand et al., 2015; Fu et al., 2006; Guo et al., 2016; Tourville et al., 2008; Toyomura et al., 2007), delayed (Sakai et al., 2009; Watkins et al., 2005), replaced with the voice of another speaker (Fu et al., 2006), or physically perturbed during vocal production (Golfinopoulos et al., 2011). In response to unexpected sensory feedback in voice production, the right IFG produces a "salient signal," indicating the potential need to stop and respond to stimuli that may be affected by external influence.

In the current fMRI experiment, we investigated how cortical voice identity and auditory feedback monitoring brain

regions respond to (un)certain self-other voice attribution. Participants elicited their own voice that varied along a morphing continuum from self to other voices, including intermediate ambiguous voices. Region of interest (ROI) analyses motivated by a priori hypotheses focused on the right aSTG and the right IFG. The right aSTG ROI stems from a well-replicated temporal voice area (TVA) localizer task (Belin et al., 2000). The right IFG ROI conforms to a region responsive to the experimental manipulation of auditory feedback previously identified in an activation-likelihood estimation (ALE) analysis (Johnson et al., 2019). Due to possible individual variability in thresholds for self-other voice attribution (Asai & Tanno, 2013), each participant underwent psychometric testing to determine individualized points of maximum uncertainty on a continuum from self to other voice. The primary goal was to test if (a) MIS of self-voice in the right aSTG is present, and the degree of suppression is greater when self-voice attribution is certainly compared to uncertain, and (b) right IFG activation would increase in response to voice uncertainty or externalization. Confirming these results would further substantiate EEG findings regarding MIS for self-voice elicited via button-press (Ford et al., 2007; Knolle et al., 2019; Pinheiro et al., 2018; Whitford et al., 2011), indicating that suppressed activity in auditory cortex aligns with predicted self-voice quality and not only as a function of expected quality of voice feedback.

2 | METERIALS AND METHODS

2.1 | Participant recruitment

Twenty-seven participants took part in the study. The data of two participants were discarded due to scanning artifacts. Of the remaining 25 (17 female), the average age was 21.88 years ($SD = 4.37$; range 18 to 33). Inclusion criteria assured that participants had no diagnosis of psychological disorder, normal or corrected-to-normal vision, reported no hearing loss, and no evidence of phonagnosia. The latter was tested with an adapted version of a voice-name recognition test (Roswandowitz et al., 2014). All participants gave written informed consent and received university study participant credit. This study was approved by the Ethical Review Committee of the Faculty of Psychology and Neuroscience at Maastricht University (ERCPN-176_08_02_2017).

3 | PROCEDURES

3.1 | Phonagnosia screening

Phonagnosia is a disorder restricting individuals from perceiving speaker identity in the voice (Van Lancker et al., 1988).

We screened for phonagnosia using an adapted version of a phonagnosia screening task (see Roswandowitz et al., 2014). The task was composed of four rounds of successive learning and testing phases, in which participants initially listened to the voices of three speakers of the same gender. Identification of each speaker was subsequently tested 10 times with response accuracy feedback provided during the first half of the test trials. Finally, the task was repeated with stimuli of the gender not used in the first run. The presentation order of these runs was counterbalanced across participants.

3.2 | Psychometric task

In a voice attribution task (VAT), participants heard neutral samples of the vowels /a/ and /o/. These samples varied in voice identity, which was morphed along a continuum from "self-voice" to "other-voice" using the STRAIGHT voice morphing software package (Kawahara, 2003, 2006) running in MATLAB (R2019A, v9.6.0.1072779, MathWorks, Inc., Natick, MA). Samples of the self-voice (SV) and other-voice (OV), producing the two vowels were obtained from each participant and normalized for duration (500ms) and amplitude (70db), using the Praat software package (v6.0.28, http://www.praat.org/). The OV sample matched the gender of the participant. On this basis, 11 stimuli for each vowel were created along a morphing spectrum in steps of 10% morphing from SV to OV. In a two-alternative forced-choice (2AFC) task, participants listened to each stimulus 10 times presented in random order and responded to the question: Is the voice "more me" or "more other"? This procedure was repeated twice. In one run, stimuli were presented passively, while in the other run participants were visually cued to press a button that elicited the next stimulus (see Figure 1). A total of 440 trials were presented across both runs. The duration across both runs was 26.7 min. This task was used to identify an individualized point of maximum ambiguity (PMA) along the morphing spectrum for each participant. *The PMA was defined as the stimulus that was closest to chance level (50%) self-other judgment and used as the uncertain-voice (UV) to inform subsequent fMRI analyses.*

3.3 | FMRI tasks

Temporal Voice Area (TVA) Localizer: To identify voice-sensitive brain areas, participants were scanned during a voice localizer task (Belin et al., 2000). This task is widely used to reliably probe activity along the bilateral temporal cortices (e.g., Pernet et al., 2015) designated as anterior, middle, and posterior TVA regions. Stimuli consisted of 8-s auditory clips with 20 vocal and 20 non-vocal sounds. In a single run, participants passively listened to these sounds and 20 silent trials of the
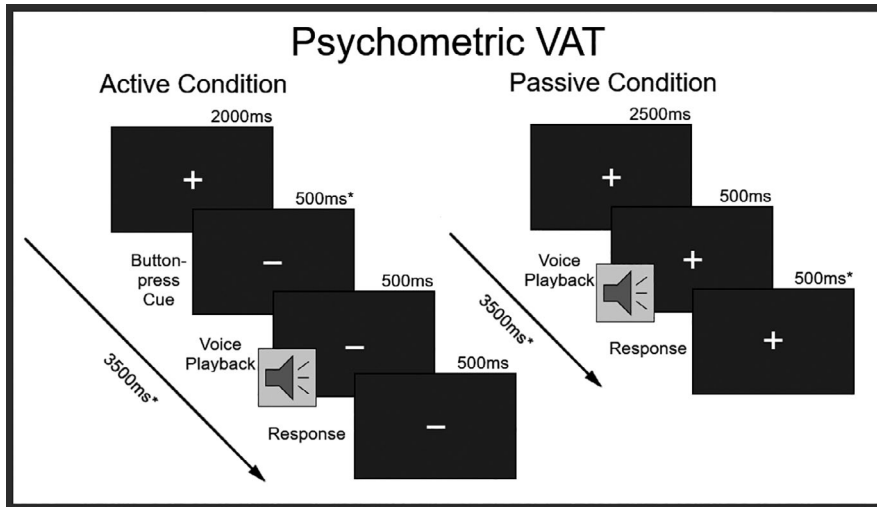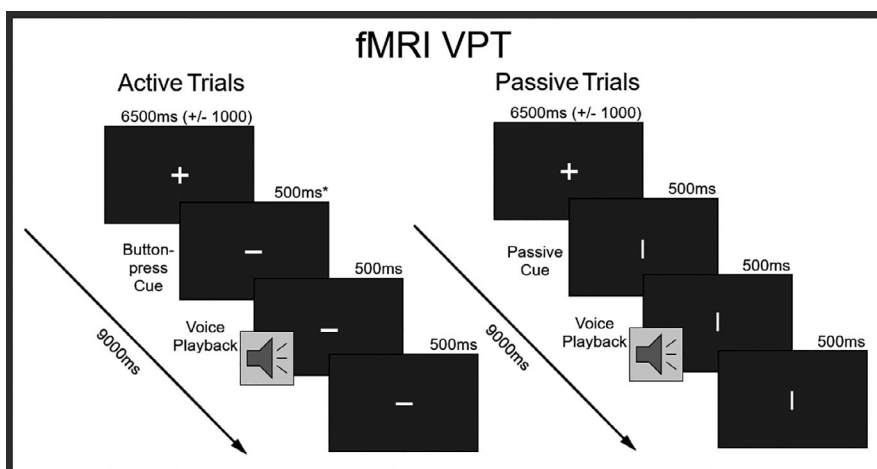
FIGURE 1 Psychometric Voice Attribution Task (VAT): Active = button-press condition; Passive = hearing conditions, * = affected by individual motor response-time variability; Response = two-alternate forcedchoice ("The voice sounded more like me." or "The voice sounded more like someone else.")

same duration in pseudorandom order. A 2-s inter-stimulus-interval separated each trial, resulting in a total task duration of 10 min. Contrasting responses of vocal and non-vocal sounds identified brain regions selectively sensitive to voice processing. The peak activation in the anterior TVA (aSTG) of the right hemisphere was then chosen as the voice-attribution ROIs in the subsequent empirical fMRI investigation.

Voice Perception Task (VPT): Participants listened to passively presented or self-generated voice stimuli. When shown a cue signifying the active button-press condition, participants pressed a button to elicit voice stimuli, and conversely when shown a cue signifying the passive condition were instructed to do nothing (Figure 2). In the active condition, half of the trials elicited a voice following the button press, while in the other half no voice was presented. In the passive condition, all trials involved the presentation of a voice. A subset of stimuli used in the VAT was selected for the VPT, specifically the 100%, 60%, 50%, 40%, and 0% self-voice morphs. Intermediate steps of 60%, 50%, and 40% were selected as pilot data had revealed that individual PMA fell within a range of 35%–65% morphing, while morphs

outside of this range produced high degrees of confidence in self versus other judgment. This ensured that every participant received the voice stimuli nearest to their subjective PMA. Trial onsets were 9 s ($\pm$500 ms) apart to allow the BOLD response to return to baseline before the presentation of the next stimulus started. To avoid the effects of adaptation suppression (Andics et al., 2010; Andics et al., 2013; Belin & Zatorre, 2003; Latinus & Belin, 2011; Wong et al., 2004), voice conditions were presented in a random order. Stimuli were presented via Sensimetrics S14 MR-compatible earphones, fitted with foam earplugs to reduce interference from scanner noise (Sensimetrics Corporation). While in the scanner participants were required to confirm the successful perception of the control stimulus at the standard 70db volume. Over two runs, a total of 100 trials were presented in each condition of Source (active and passive). Within each condition of Source, each voice stimulus (100%, 60%, 505, 405, and 0% morphs from self-to-other) was heard 20 times. Twenty null trials were included to provide a baseline comparison of activity in response to experimental trials. The total duration of this task over both runs was 33 min.

FIGURE 2 fMRI Voice Perception Task (VPT): Active = button-press condition; Passive = hearing conditions, * = affected by individual motor response-time variability

FIGURE 3 fMRI Regions of Interest: Blue: right inferior frontal gyrus; MNI coordinates x 58, y 2, z −10; determined from ALE neuroimaging meta-analysis (Johnson et al., 2019). Red: right anterior superior temporal gyrus; MNI coordinates x 46, y 10, z 4; determined in our sample from fMRI temporal voice area localizer task
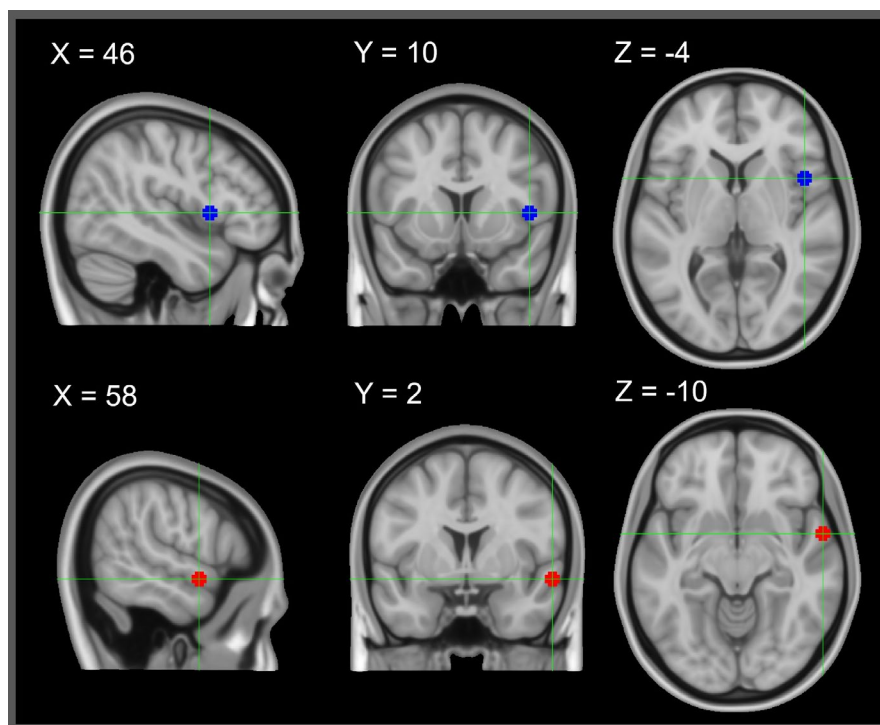
3.4 | FMRI data acquisition and analysis

Data acquisition was performed at a Siemens 3T Magnetom Prisma Fit Magnetic Resonance Imaging (MRI) scanner at Scannexus facilities (Maastricht, NE), equipped with a 32-channel head coil (Siemens Healthcare, Erlangen, Germany). A structural whole-brain T1-weighted single-shot echoplanar imaging (EPI) sequence was collected for each participant (field of view (FOV) 256 mm; 192 axial slices; 1mm slice thickness; 1 × 1 × 1mm voxel size; repetition time (TR) of 2250 ms; echo-time (TE) 2.21 ms). Two functional tasks were conducted with T2-weighted EPI scans (FOV 208mm; 60 axial slices; 2mm slice thickness; 2 × 2 × 2mm voxel size; TE 30 ms; flip angle = 77°). Both tasks applied a long inter-acquisition-interval where the time between consecutive image acquisition (2000 ms) was delayed, resulting in a TR of 10 and 9 s for the TVA localizer and VPT, respectively. This allowed auditory stimuli to be presented during a period of relative silence to reduce noise artifacts and for volume acquisition to proceed during a period of peak activation in the auditory cortex (Belin et al., 1999; Hall et al., 1999).

DICOM image data were converted to 4D NIFTI format using the Dcm2Nii converter provided in the MRIcron software package (https://www.nitrc.org/projects/mricron/). The topup tool (Smith et al., 2004) implemented in FSL (www.fmrib.ox.ac.uk/fsl) was used to estimate and correct for susceptibility-induced image distortions. Pre-processing was performed using SPM12 (Wellcome Department of Cognitive Neurology, London, UK). A pre-processing pipeline applied slice timing correction, realignment and

unwarping, segmentation, normalization to standard Montreal Neurological Institute (MNI) space (Fonov et al., 2009) as well as smoothing with a full width at half maximum (FWHM) 8mm isotropic Gaussian kernel.

General Linear Model (GLM) Analysis: The TVA localizer and experimental VPT fMRI data were analyzed with a standard two-level procedure in SPM12. For the TVA localizer, contrast images for Vocal > Non-Vocal and Vocal > Silent were estimated for each participant. To test for the main effect of interest, a conjunction analysis ((V > NV) ∩ (V > S)) was performed. A second-level random-effects analysis tested for group-level significance. A first-level fixed-effects GLM of the VPT data calculated contrast estimates for each participant. Contrast estimates were then used in the subsequent hypothesis-driven ROI analysis to investigate TVA activity.

Linear Mixed Model (LMM) ROI Analyses: Two spherical (5mm) ROIs were selected for analysis: the right aSTG/S in Brodmann Area (BA) 22 (MNI coordinates x 58, y 2, z −10) defined by the TVA fMRI localizer task, and the right IFG opercular region in BA 44 (MNI coordinates x 46, y 10, z 4) defined in a previous fMRI meta-analysis (Johnson et al., 2019) (See Figure 3). For both ROIs, the mean contrast estimates were produced for each SV, UV, and OV condition against the null trials. Using these values as input, a 2 × 3 factorial design was formulated using the factors of Source and Voice. The two-levelled factor Source included self-generated (A) and passively heard (P) playback of voice recordings. The three-levelled factor Voice included self-attributed (SV), other-attributed (OV), and ambiguous voice (UV).

TABLE 1 TVA localizer results

Cluster #	Peak Label	BA	Coordinates (x, y, x)			Cluster-Level *p*-FDR	Peak-Level *p*-FDR	Cluster Size (voxels)
1	L pSTG	22	−60	−24	0	2.67×10^{-14}	9.29×10^{-12}	4,551
	L aSTG	22	−58	−10	−2		9.04×10^{-11}	
	L mSTG	22	−66	−16	−2		3.51×10^{-8}	
2	R pSTG	22	58	−24	−2	2.05×10^{-14}	1.07×10^{-9}	4,565
	R aSTG	22	58	2	−10		2.00×10^{-9}	
	R mSTG	22	58	−8	−6		2.74×10^{-9}	
3	R preCG	6	52	52	0	0.007	3.44×10^{-4}	408
4	L IFG	44	−42	14	22	0.019	0.002	294

Note: Results from TVA localizer task: Coordinates listed in MNI space; L: left, R: right, (p/a/m)STG: posterior/anterior/middle superior temporal gyrus, preCG: precentral gyrus, IFG: inferior frontal gyrus; 7 peak-level activations in 4 clusters: 1. left STG, 2. right STG, 3. right preCG, 4. left IFG; All listed significant regions survived FDR-corrected threshold 0.05.

Data were analyzed in R v3.6.1 (R Core Team, 2019) running on OS v10.11.6. Data handling and visualization were supplemented with the tidyverse (Wickham, 2017). Linear Mixed Models (LMMs) were fit with lme4 (Bates et al., 2015). Separate LMMs were fitted for contrast estimates of the IFG and the aSTG ROIs with Source (A and P), Voice (SV, OV, and UV), and their interaction as fixed effects. Participant was modeled as a random intercept. Model residuals were examined for potential outliers. Five data points were removed from the IFG analysis and one was removed from the aSTG analysis.

The main effects of Voice, Source, and their interaction were tested with the afex package using Kenward-Rogers degrees of freedom (Singmann et al., 2015). Estimated marginal means and confidence intervals were computed with the emmeans package (Lenth et al., 2020) for visualization. All *p* values are corrected for multiple comparisons controlling at a false-discovery rate (FDR) of 0.05. Furthermore, to investigate the effect that (un)certainty has on the suppression of the right aSTG, we compared contrasts of A > P in each voice condition to see if it differs for SV stimuli as compared to OV or UV stimuli.

Finally, to provide clear effects of each treatment condition (UV and OV) compared to the control variable of one's own voice (SV), contrast estimates were reported via two-tailed paired sample *t* tests. These comparisons were performed within both active (A) and passive (P) conditions. Furthermore, we provide BOLD whole-brain activation maps for each Source condition (P and A) against null trials (see supplementary).

4 | RESULTS

4.1 | VAT results

Psychometric analysis of the VAT indicated little variability in the degree of morphing between SV and OV required to elicit responses at chance level (50%), which we identified

as the point of maximum ambiguity. For the A condition, 9 participants had PMAs at 40%, 8 at 50%, and 10 at 60% morphing. In the passive condition, eleven participants required 40%, seven 50%, and nine 60% morphing. There was no significant difference between the average morphing required to elicit PMA in A (μ 50%, *SD* 0.085) and P (μ 50%, *SD* 0.087) conditions. Although no participant matched criteria for phonagnosia as specified by the screening task, VAT data from one participant was excluded due to an inability to reliably differentiate between their own and other voices.

4.2 | TVA localizer results

The TVA fMRI localizer produced four significant cluster-level activations (see Table 1 for details). Within two large bilateral STG (BA 22) clusters, each included three peak-level significant activations. These peaks correspond to the posterior, middle, and anterior STG. Two smaller clusters were found in the right precentral gyrus (BA 6), the left IFG (BA 44), and the left inferior parietal lobule (BA 40). All significant cluster- and peak-level coordinates survived a FDR correction of 0.05. These results replicate the pattern of TVA regions of peak activity (e.g., Belin et al., 2000; Fecteau et al., 2004; Latinus et al., 2013; Pernet et al., 2015). The right aSTG peak was chosen for the ROI analysis of self-voice-attribution.

4.3 | LMM ROI results

Contrast estimates calculated for each Voice condition (SV, UV, and OV) within each Source condition (A and P) were used as input for the 2 × 3 factorial design (see supplementary). Linear mixed model analysis of the right aSTG (Figure 4a) produced an FDR-corrected significant main effect for the factor of Voice ($F_{2,118.94} = 4.90$, $p = .021$). No significant effect was observed for Source ($F_{1,118.92} = 0.53$,
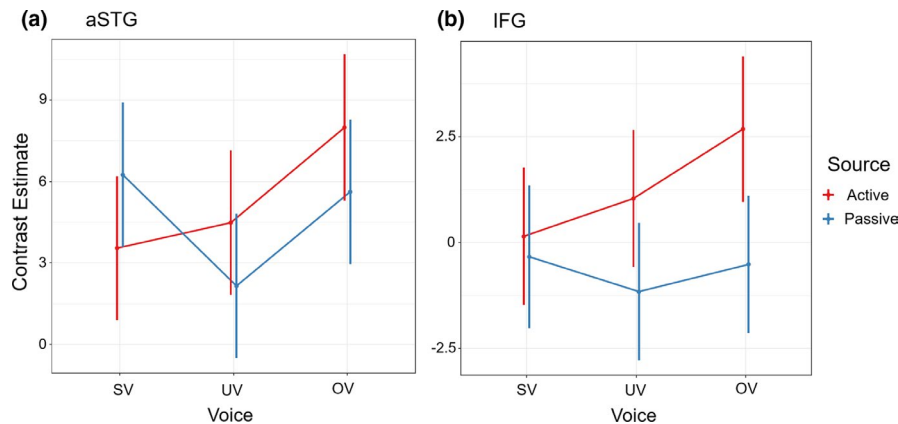
FIGURE 4 fMRI Voice Perception Task (VPT) LMM Results: Linear mixed model analysis on ROIs in A) right anterior superior temporal gyrus (aSTG) and B) right inferior frontal gyrus (IFG). Active: button-press condition, passive: hearing condition, SV: self-voice, UV: uncertain-voice, OV: other-voice. Hypothesis-driven analysis in right aSTG confirmed motor induced suppression (for contrast active > passive) for only SV as compared to UV or OV (t(119) = −2.7, p = .021)

$p = .47$). A trend for the expected interaction effect between Voice and Source was observed, although it did not survive FDR correction for multiple comparisons ($F_{2,118.94} = 3.40$, $p = .065$). Our a priori hypotheses regarding a difference in suppression effects (A > P) between voice conditions, however, was supported by the finding that MIS is observed preferentially for SV stimuli ($t_{119} = −2.7$, $p = .021$).

The LMM analysis was repeated for the right IFG ROI (Figure 4b). A significant FDR-corrected main effect of Source was observed ($F_{1,116.04} = 9.93$, $p = .002$). No main effect was found for the factor of Voice ($F_{2,115.95} = 1.52$, $p = .26$), and no interaction between Voice and Source were observed ($F_{2,115.81} = 1.60$, $p = .26$).

Finally, for both ROIs, we report direct comparisons of treatment (UV and OV) and control (SV) conditions within each source conditions (A and P). Activation of the right aSTG was reduced when participants actively produced SV compared to OV ($t_{121} = −2.773$, $p = .0064$), but not compared to UV ($t_{121} = −0.591$, $p = .5553$). Conversely, activation of the right aSTG was no different when participants were passively exposed to SV compared to the OV ($t_{118} = 0.401$, $p = .6891$), but increased compared to UV ($t_{118} = 2.612$, $p = .0102$). Therefore, the right aSTG was less active when self-generating a voice certain to be one's compared to an external source, and more active when passively hearing a voice certain to be one's own compared to when uncertain. In the right IFG, activation was reduced when participants actively produced SV compared to OV ($t_{116} = −2.316$, $p = .0223$), but not compared to UV ($t_{115} = −0.846$, $p = .3993$). When passively exposed, activation of the right IFG was no different when presented with SV compared to OV ($t_{117} = 1.66$, $p = .8681$), or to UV ($t_{117} = 0.758$, $p = .4499$). In summary, in the right IFG a greater activation relative to the self-voice control was reported in only the externalized self-generated voice.

5 | DISCUSSION

The current study investigated how unexpected sensory feedback affects certainty in self-voice attribution. We report first fMRI evidence that aligns with EEG reports, namely self-voice MIS is observed in the anterior region of the STG even when one's own voice is elicited by a button press. Expected self-voice quality, learned through long-term experience with self-voice feedback sufficiently modulates MIS. Importantly, this effect was specific to vocal properties matching the producer's own voice and was not observed when hearing another voice or being uncertain about the voice of a speaker. The right IFG showed increased activation in response to the self-generated voice compared to listening to the same voice. It is possible that this response is driven by voice trials not attributed to oneself. This region is known to be more active when perceived stimuli conflict with expected sensory feedback. Together, these findings suggest a differentiation between and a potential interplay of right IFG and anterior STG in voice processing, and more specifically feedback monitoring of self-generated voice and voice attribution.

5.1 | Voice identity and motor-induced suppression in the anterior STG

Our results confirm right anterior STG/S involvement for voice identity and indicate that this region plays a specific role in segregating the speaker's voice from other voices in the monitoring of auditory feedback. We replicate previous TVA findings that the STG and upper bank of the STS contain three bilateral voice patches (Table 1) (Belin et al., 2000; Pernet et al., 2015). The processing of speech-related linguistic ("what") features have been attributed predominantly to the left hemisphere, while speaker-related paralinguistic

("who") features have been attributed predominantly to the right hemisphere (Belin et al., 2002; Ethofer et al., 2006, 2007; Formisano et al., 2008; Grandjean et al., 2005; Kotz et al., 2003; Moerel et al., 2012; Schirmer & Kotz, 2006; Wiethoff et al., 2008). Furthermore, regions of the right lateral temporal cortex are specialized for different speaker-related information. Identity attribution is localized to the anterior region of the STG/S (Belin & Zatorre, 2003; Fecteau et al., 2004; von Kriegstein et al., 2003; von Kriegstein & Giraud, 2004; Latinus et al., 2013; Schelinski et al., 2016). Considering voice-identity processing as a multi-stage process, low-level acoustics features are evaluated in the posterior STG for cues relevant to speaker identification; the extracted cues are then further processed and compared to prototypes for deviance detection in the middle STG, and finally voice identity recognition occurs in the anterior STG (Maguinness et al., 2018).

We conducted ROI analyses to test voice identity in the right anterior STG due to its responsiveness to variation in voice identity but did not include other TVA regions in our analysis. This allowed us to detect fine-grain differences in activation patterns influenced only by voice identity in a region that is related to the perception of one's own voice. To provide sufficient information for the extraction of paralinguistic speaker-related features, steady 500 ms vowel excerpts were chosen as voice samples (Pinheiro et al., 2018; Schweinberger et al., 1997, 2011; Van Berkum et al., 2008). Although vowels provide fundamental cues that allow differentiating between speakers (Belin et al., 2004; Kreiman & Sidtis, 2011; Latinus & Belin, 2011; Schweinberger et al., 2014), to the best of our knowledge, no study has yet confirmed whether such basic stimuli carry enough identity cues to allow for explicit self-recognition (Conde et al., 2018). Our results confirm that the use of short vowels is sufficient to accurately recognize self versus other voices.

Suppression of self-generated relative to passively heard voice in the right anterior STG occurred only within SV (see Figure 4a). One possible interpretation for this selective finding is that participants are most familiar with their own voice and they can, therefore, predict the features of their own voice more efficiently. In the right anterior STG, voice identity is defined by the extent that speaker-related cues deviate from prototypes of expected voice qualities (Andics et al., 2010, 2013; Bruckert et al., 2010; Latinus & Belin, 2011; Latinus et al., 2013; Mullennix et al., 2011; Petkov & Vuong, 2013; Schweinberger et al., 2014). These referential prototypes are learned through mean-based coding (Hoffman & Logothesis, 2009), and differ for male and female voices (Charest et al., 2013; Latinus et al., 2013). While it is clear that low-level acoustic processing is involved in recognizing the identity of a speaker (Baumann & Belin, 2010; Gaudrain et al., 2009; Kreitewolf et al., 2014; Nolan et al., 2011; Smith & Patterson, 2005; Smith et al., 2007; Zheng et al., 2011), the specific features that drive voice identification vary from voice to voice (Kreiman et al., 1992; Latinus & Belin, 2012;

Lavner et al., 2000, 2001; Xu et al., 2013). Furthermore, variable acoustic features of the voice do not only exist between speakers, but also within individual speakers (Lavan et al., 2019). Therefore, increased experience with the voice of a specific speaker facilitates more efficient recognition of voice identity. As speakers are most experienced with their own voice, little divergence from mean-based coding is expected.

Alternatively, MIS of the self-voice in a dynamic multi-speaker environment is important for the segregation of internally and externally controlled voice stimuli. During vocalization, an efference copy of the motor command is sent from motor planning areas to auditory and sensorimotor cortical regions to notify of impending feedback (Hickok, 2012; Hickok et al., 2011; Kearney & Guenther, 2019; Rauschecker, 2011; Rauschecker & Scott, 2009; Tourville & Guenther, 2011). Error-cells in the posterior STG (planum temporale) receive these signals from Broca's area to remain inactive in response to the expected self-voice and to engage when perceiving voice feedback outside the control of the speaker (Guenther et al., 2006). To date, fMRI research using vocal feedback paradigms, has provided evidence for this form of MIS dependent on vocal production. For example, MIS has been reported for unaltered vocal production relative to hearing a recording of self-voice or in a noisy environmental (Christoffels et al., 2007), when acoustically distorted (Christoffels et al., 2011; Fu et al., 2006; McGuire et al., 1996; Zheng et al., 2010), or replaced with the voice of another speaker (Fu et al., 2006; McGuire et al., 1996). However, as these paradigms all rely on vocal production, they could not isolate how voice identity engages the anterior STG in voice production. EEG research has provided evidence for MIS in the auditory cortex that does not depend on vocal speech production as it is observed even when sounds are elicited by a button press. For example, MIS of the N1 response was reported for both, vocal (Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005; Sitek et al., 2013; Wang et al., 2014) and button-press elicited self-voice (Ford et al., 2007; Knolle et al., 2019; Pinheiro et al., 2018; Whitford et al., 2011). In line with previous EEG evidence, the current findings confirm self-voice suppression as a marker of voice identity in the right anterior STG. The reported MIS is specific to self-voice processing, providing further evidence of voice identity suppression independent of previously reported cortical suppression during unperturbed speech. Importantly, this pattern was observed only for own voice attribution and was not present when the voice was distorted to an extent that self-attribution was uncertain.

5.2 | Expected feedback and the IFG

The right IFG was more strongly activated when participants generated their own voice with a button press as compared to passive listening to their own voice. This finding confirms that

this region is more responsive to sounds triggered by oneself, potentially as part of an auditory feedback loop. Increased activity in this region has been observed in response to acoustically altered (Behroozmand et al., 2015; Fu et al., 2006; Guo et al., 2016; Tourville et al., 2008; Toyomura et al., 2007), physically perturbed (Golfinopoulos et al., 2011), and externalized voice feedback (Fu et al., 2006).

In response to unexpected sensory information, the right IFG plays a significant role in relaying salient signals to attention networks. Moreover, the right IFG is part of a prediction network, which forms expectations and detects unexpected sensory outcomes (Siman-Tov et al., 2019). When prediction errors are detected, an inferior frontal network produces a salience response (Cai et al., 2014; Chang et al., 2013; Power et al., 2011; Seeley, 2010). Salience signals engage ventral and dorsal attention networks, overlapping the right inferior frontal cortex. The ventral attention network responds with bottom-up inhibition of ongoing action (Aron, Robbins, & Poldrack, 2004, 2014), such as halting manual or speech movement (Aron, 2007; Aron & Poldrack, 2006; Chevrier et al., 2007; Xue et al., 2008). Correspondingly, damage to prefrontal regions affects the ability to stop one's own actions (Aron et al., 2003), and is similarly diminished when the IFG is deactivated with TMS (Chambers et al., 2006). The salience response may also engage the dorsal attention network to facilitate a top-down response (Corbetta & Shulman, 2002; Dosenbach et al., 2007; Eckert et al., 2009; Fox et al., 2006), for example, in goal-directed vocal compensation to pitch-shift (Riecker et al., 2000; Toyomura et al., 2007; Zarate & Zatorre, 2005) or somatosensory perturbation (Golfinopoulos et al., 2011). The right IFG in the current study maps with a region determined by an ALE meta-analysis of neuroimaging studies that experimentally manipulated auditory feedback in vocal and manual production (Johnson et al., 2019). As the current experiment required no explicit response to a change in stimulus quality, we hypothesized that increased activity in the right IFG may represent the initial salience response to unexpected voice quality. However, the effect of voice identity in the right IFG did not reach significance, and there was no significant interaction between stimulus source and voice identity in this region. We note that the main effect of the source appears most strongly driven by unfamiliar or ambiguous voices, with an intermediate level increase in the uncertain condition (see Figure 4b). It is possible that substantial variability in the data limiting these results was due to the passive nature of the task with no overt attention to stimulus quality. As activity in this region is associated with attention and subsequent inhibition/adaptation responses, the degree to which each participant attended to the change in stimulus quality is unclear. However, simple contrast analyses relative to self-voice demonstrated a significant increase in activation for only self-generated other-voice. It is possible that this region is solely activated when voice feedback is manipulated to the extent that it is externalized. Conversely, although

psychometric testing confirmed the subjective ability of participants to correctly recognize their own and other voices at a behavior level, it is possible that the brief vowel stimuli did not provide sufficient information to signal a strong response to unexpected changes in self-voice leading to uncertainty. Further research is, therefore, needed to clarify whether the right IFG is responsive to voice identity, and to which extent this may be driven by the degree of salience elicited in divergence from expected qualities of self-voice.

5.3 | Variability in self-monitoring thresholds

Although recordings of self-voice can produce a feeling of eeriness for listeners as compared to when spoken (Kimura & Yotsumoto, 2018), people nevertheless recognize recorded voice samples as their own (Candini et al., 2014; Hughes & Nicholson, 2010; Kaplan et al., 2008; Nakamura et al., 2001; Pinheiro et al., 2016, 2019; Pinheiro, Rezaii, Nestor, et al., 2016; Rosa et al., 2008; Xu et al., 2013). However, in ambiguous conditions (i.e., acoustic distortion), the ability to accurately attribute a voice to oneself is diminished (Allen et al., 2004, 2005, 2006; Allen et al., 2007; Fu et al., 2006; Kumari, Fannon, et al., 2010; Kumari, Antonova, et al., 2010). As ambiguity increases, an attribution threshold is passed, initiating a transition from uncertainty to externalization (Johns et al., 2001, 2003, 2006; Vermissen et al., 2007). This threshold, however, varies from person to person (Asai & Tanno, 2013). It was, therefore, necessary to determine the degree of morphing required to elicit uncertainty in the attribution of voice identity via separate 2AFC psychometric analysis for each participant. In doing so, we could confirm that fMRI responses in the PMA condition were specific to the experience of maximum uncertainty, regardless of any variability in the individual thresholds. These results confirmed that participants were able to discriminate their self-voice from an unfamiliar voice, with relatively little variation regarding the point of maximum ambiguity.

An externalization bias is particularly prominent in schizophrenia patients who experience AVH (Allen et al., 2004, Allen et al., 2007; Costafreda et al., 2008; Heinks-Maldonado et al., 2007; Johns et al., 2001, 2006; Pinheiro, Rezaii, Rauber, et al., 2016). It has been hypothesized that the processing of salient stimuli with minimal divergence from expectations leads to an externalization bias that may manifest in the experience of AVH (Sommer et al., 2008). Correspondingly, as the severity of AVH symptoms increase, accuracy in self-attribution voice diminishes (Allen et al., 2004, 2006; Pinheiro, Rezaii, Rauber, et al., 2016). Notably, this symptomology does not only exist within patient groups. Individuals who present sub-clinical symptoms but are at a high risk to develop psychosis, display levels of self-monitoring performance similar

to patients who meet a clinical diagnosis of schizophrenia (Johns et al., 2010; Vermissen et al., 2007). Indeed, proneness to hallucinate is a continuum and AVH is experienced in the general populations as well, although at lower rates (Baumeister et al., 2017). Even in non-clinical populations, AVH are associated with a bias toward external voice attributions (Asai & Tanno, 2013; Pinheiro et al., 2019). The current findings may be of value in the understanding of the neural substrates underlying dysfunctional self-other voice attribution. In light of our observation that the anterior STG displays a qualitatively different activation for self-voice relative to an unfamiliar voice and the hypothesized influence of right IFG overactivity in salience detection in AVH, we suggest future research in high-risk groups to assess a possibly altered interaction between these two regions. Structural and functional connectivity MRI analysis may help explain whether aberrant communication between these two regions, or individual changes in either or both regions lead to this symptomatology.

6 | CONCLUSION

The goal of the current experiment was to investigate how levels of self-voice certainty alter brain activity in voice identity and feedback quality monitoring regions of the brain. By replicating earlier findings using a voice area localizer task, we isolated a putative voice identity region in the right anterior STG. Our results indicate activity in this TVA is suppressed only for the self-generated voice. Although the involvement of the right IFG was not confirmed in processing unexpected features of uncertain voice identity, increased IFG activity in response to the self-generated other-voice was observed, indicating a possible role of feedback-monitoring for externalized voice. Using a novel self-monitoring paradigm, we provide the first fMRI evidence for the effectiveness of button-press voice-elicitation in modulating identity-related MIS in the auditory cortex. Further, we present novel findings on the effectiveness of brief vowel excerpts to provide sufficient paralinguistic information to explicitly identify one's own voice. Finally, we suggest a potential dynamic interaction of the right anterior STG and IFG in self-voice monitoring. The feedback monitoring frontal region may inform the temporal voice identity region whenever a salience threshold has been passed and voice feedback is influenced by or under the control of an external actor. The implications of the current results may be particularly relevant to the externalization of self-generated voice in AVH.

7 | COMPETING INTERESTS

All authors disclose no potential sources of conflict of interest.

AUTHOR CONTRIBUTIONS

JFJ, MB, MS, APP, & SAK designed the experiment. JFJ collected the data. JFJ analyzed the data with methodological feedback from MB, MS, and SAK. JFJ wrote the manuscript and MB, MS, APP, and SAK provided feedback and edits. APP, MS, and SAK secured funding.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.15162.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Joseph F. Johnson https://orcid.org/0000-0001-6239-0380

REFERENCES

Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience*, *21*(4), 791–802. https://doi.org/10.1162/jocn.2009.21055

Allen, P. P., Amaro, E., Fu, C. H., Williams, S. C., Brammer, M., Johns, L. C., & McGuire, P. K. (2005). Neural correlates of the misattribution of self-generated speech. *Human Brain Mapping*, *26*(1), 44–53. https://doi.org/10.1002/hbm.20120

Allen, P., Freeman, D., Johns, L., & McGuire, P. (2006). Misattribution of self-generated speech in relation to hallucinatory proneness and delusional ideation in healthy volunteers. *Schizophrenia Research*, *84*(2–3), 281–288. https://doi.org/10.1016/j.schres.2006.01.021

Allen, P. P., Johns, L. C., Fu, C. H., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. *Schizophrenia Research*, *69*(2–3), 277–287. https://doi.org/10.1016/j.schres.2003.09.008

Andics, A., McQueen, J. M., & Petersson, K. M. (2013). Mean-based neural coding of voices. *NeuroImage*, *79*, 351–360. https://doi.org/10.1016/j.neuroimage.2013.05.002

Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z. (2010). Neural mechanisms for voice recognition. *NeuroImage*, *52*(4), 1528–1540. https://doi.org/10.1016/j.neuroimage.2010.05.048

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The Neuroscientist*, *13*(3), 214–228. https://doi.org/10.1177/1073858407299288

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right

inferior frontal gyrus in humans. *Nature Neuroscience*, 6(2), 115–116. https://doi.org/10.1038/nn1003

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: Role of the subthalamic nucleus. *Journal of Neuroscience*, 26(9), 2424–2433. https://doi.org/10.1523/JNEUROSCI.4682-05.2006

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170–177. https://doi.org/10.1016/j.tics.2004.02.010

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, 18(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003

Asai, T., & Tanno, Y. (2013). Why must we attribute our own action to ourselves? Auditory hallucination like-experiences as the results both from the explicit self-other attribution and implicit regulation in speech. *Psychiatry Research*, 207(3), 179–188. https://doi.org/10.1016/j.psychres.2012.09.055

Baess, P., Widmann, A., Roye, A., Schröger, E., & Jacobsen, T. (2009). Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. *European Journal of Neuroscience*, 29(7), 1514–1521. https://doi.org/10.1111/j.1460-9568.2009.06683.x

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, 12(1), 2. http://cran.r-project.org/package=lme4

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research Psychologische Forschung*, 74(1), 110. https://doi.org/10.1007/s00426-008-0185-z

Baumeister, D., Sedgwick, O., Howes, O., & Peters, E. (2017). Auditory verbal hallucinations and continuum models of psychosis: A systematic review of the healthy voice-hearer literature. *Clinical Psychology Review*, 51, 125–141. https://doi.org/10.1016/j.cpr.2016.10.010

Behroozmand, R., & Larson, C. (2011). Motor-induced suppression of auditory neural responses to pitch-shifted voice feedback. *The Journal of the Acoustical Society of America*, 129(4), 2454. https://doi.org/10.1121/1.3588063

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A. III, & Greenlee, J. D. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. *NeuroImage*, 109, 418–428. https://doi.org/10.1016/j.neuroimage.2015.01.040

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, 8(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport*, 14(16), 2105–2109. https://doi.org/10.1097/00001756-200311140-00019

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13(1), 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *NeuroImage*, 10(4), 417–429. https://doi.org/10.1006/nimg.1999.0480

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403(6767), 309–312. https://doi.org/10.1038/35002078

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, 1(7), 635–640. https://doi.org/10.1038/2870

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H., & Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology*, 20(2), 116–120. https://doi.org/10.1016/j.cub.2009.11.034

Cai, W., Ryali, S., Chen, T., Li, C. S. R., & Menon, V. (2014). Dissociable roles of right inferior frontal cortex and anterior insula in inhibitory control: Evidence from intrinsic and task-related functional parcellation, connectivity, and response profile analyses across multiple datasets. *Journal of Neuroscience*, 34(44), 14652–14667. https://doi.org/10.1523/JNEUROSCI.3048-14.2014

Candini, M., Zamagni, E., Nuzzo, A., Ruotolo, F., Iachini, T., & Frassinetti, F. (2014). Who is speaking? Implicit and explicit self and other voice recognition. *Brain and Cognition*, 92, 112–117. https://doi.org/10.1016/j.bandc.2014.10.001

Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., Morris, A. P., & Mattingley, J. B. (2006). Executive "brake failure" following deactivation of human frontal lobe. *Journal of Cognitive Neuroscience*, 18(3), 444–455. https://doi.org/10.1162/jocn.2006.18.3.444

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, 110(7), 2653–2658. https://doi.org/10.1073/pnas.1216827110

Charest, I., Pernet, C., Latinus, M., Crabbe, F., & Belin, P. (2013). Cerebral processing of voice gender studied using a continuous carryover fMRI design. *Cerebral Cortex*, 23(4), 958–966. https://doi.org/10.1093/cercor/bhs090

Chevrier, A. D., Noseworthy, M. D., & Schachar, R. (2007). Dissociation of response inhibition and performance monitoring in the stop signal task using event-related fMRI. *Human Brain Mapping*, 28(12), 1347–1358. https://doi.org/10.1002/hbm.20355

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Human Brain Mapping*, 28(9), 868–879. https://doi.org/10.1002/hbm.20315

Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E., & Schiller, N. O. (2011). The sensory consequences of speaking: Parametric neural cancellation during speech in auditory cortex. *PLoS One*, 6(5), e18307. https://doi.org/10.1371/journal.pone.0018307

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2018). Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing. *International Journal of Psychophysiology*, 133, 66–78. https://doi.org/10.1016/j.ijpsycho.2018.08.007

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201–215. https://doi.org/10.1038/nrn755

Costafreda, S. G., Brébion, G., Allen, P., McGuire, P. K., & Fu, C. H. Y. (2008). Affective modulation of external misattribution bias in source monitoring in schizophrenia. *Psychological Medicine*, 38(6), 821–824. https://doi.org/10.1017/S0033291708003243

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., Fox, M. D., Snyder, A. Z., Vincent, J. L., Raichle, M. E., Schlaggar, B. L., & Petersen, S. E. (2007). Distinct brain networks for adaptive and stable task control in humans.

Proceedings of the National Academy of Sciences, *104*(26), 11073–11078. https://doi.org/10.1073/pnas.0704320104

Eckert, M. A., Menon, V., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., & Dubno, J. R. (2009). At the heart of the ventral attention system: The right anterior insula. *Human Brain Mapping*, *30*(8), 2530–2541. https://doi.org/10.1002/hbm.20688

Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*, *453*(7198), 1102–1106. https://doi.org/10.1038/nature06910

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., & Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *NeuroImage*, *30*(2), 580–587. https://doi.org/10.1016/j.neuroimage.2005.09.059

Ethofer, T., Wiethoff, S., Anders, S., Kreifelts, B., Grodd, W., & Wildgruber, D. (2007). The voices of seduction: Cross-gender effects in processing of erotic prosody. *Social Cognitive and Affective Neuroscience*, *2*(4), 334–337. https://doi.org/10.1093/scan/nsm028

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? *An fMRI Study. Neuroimage*, *23*(3), 840–848. https://doi.org/10.1016/j.neuroimage.2004.09.019

Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin*, *4*(4), 636. https://doi.org/10.1093/schbul/4.4.636

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5

Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., & Mathalon, D. H. (2007). Dissecting corollary discharge dysfunction in schizophrenia. *Psychophysiology*, *44*(4), 522–529. https://doi.org/10.1111/j.1469-8986.2007.00533.x

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, *322*(5903), 970–973. https://doi.org/10.1126/science.1164318

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences*, *103*(26), 10046–10051. https://doi.org/10.1073/pnas.0604187103

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Psychology Press. https://doi.org/10.4324/9781315785011

Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro, E., Andrew, C. M., Yáguez, L., van Haren, N. E. M., Matsumoto, K., & McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*(7), 969–977. https://doi.org/10.1093/cercor/bhj039

Gainotti, G., Ferraccioli, M., & Marra, C. (2010). The relation between person identity nodes, familiarity judgment and biographical information. Evidence from two patients with right and left anterior temporal atrophy. *Brain Research*, *1307*, 103–114. https://doi.org/10.1016/j.brainres.2009.10.009

Gainotti, G., & Marra, C. (2011). Differential contribution of right and left temporo-occipital and anterior temporal lesions to face recognition disorders. *Frontiers in Human Neuroscience*, *5*, 55. https://doi.org/10.3389/fnhum.2011.00055

Gaudrain, E., Li, S., Ban, V. S., & Patterson, R. D. (2009). The role of glottal pulse rate and vocal tract length in the perception of speaker identity. *Interspeech*, *1*(5), 152–155. https://doi.org/10.6084/m9.figshare.870509

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, *55*(3), 1324–1338. https://doi.org/10.1016/j.neuroimage.2010.12.065

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*(2), 145–146. https://doi.org/10.1038/nn1392

Greenlee, J. D. W., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., Chen, H., & Howard, M. A. (2011). Human auditory cortical activation during self-vocalization. *PLoS One*, *6*(3), e14744. https://doi.org/10.1371/journal.pone.0014744

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301. https://doi.org/10.1016/j.bandl.2005.06.001

Guo, Z., Huang, X., Wang, M., Jones, J. A., Dai, Z., Li, W., Liu, P., & Liu, H. (2016). Regional homogeneity of intrinsic brain activity correlates with auditory-motor processing of vocal pitch errors. *NeuroImage*, *142*, 565–575. https://doi.org/10.1016/j.neuroimage.2016.08.005

Hailstone, J. C., Ridgway, G. R., Bartlett, J. W., Goll, J. C., Buckley, A. H., Crutch, S. J., & Warren, J. D. (2011). Voice processing in dementia: A neuropsychological and neuroanatomical analysis. *Brain*, *134*(9), 2535–2547. https://doi.org/10.1093/brain/awr205

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., & Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain Mapping*, *7*(3), 213–223. https://doi.org/10.1002/(SICI)1097-0193(1999)7:3<213:AID-HBM5>3.0.CO;2-N

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, *42*(2), 180–190. https://doi.org/10.1111/j.1469-8986.2005.00272.x

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of General Psychiatry*, *64*(3), 286–296. https://doi.org/10.1001/archpsyc.64.3.286

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145. https://doi.org/10.1038/nrn3158

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–422. https://doi.org/10.1016/j.neuron.2011.01.019

Hoffman, K. L., & Logothetis, N. K. (2009). Cortical mechanisms of sensory learning and object recognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1515), 321–329. https://doi.org/10.1098/rstb.2008.0271

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, *14*(8), 1125–1138. https://doi.org/10.1162/089892902760807140

Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition of self-stimuli. *Consciousness*

and Cognition, *19*(4), 1124–1134. https://doi.org/10.1016/j.concog.2010.03.001

Johns, L. C., Allen, P., Valli, I., Winton-Brown, T., Broome, M., Woolley, J., Tabraham, P., Day, F., Howes, O., Wykes, T., & McGuire, P. (2010). Impaired verbal self-monitoring in individuals at high risk of psychosis. *Psychological Medicine*, *40*(9), 1433. https://doi.org/10.1017/s0033291709991991

Johns, L. C., Gregg, L., Allen, P., & McGuire, P. K. (2006). Impaired verbal self-monitoring in psychosis: Effects of state, trait and diagnosis. *Psychological Medicine*, *36*(4), 465–474. https://doi.org/10.1017/S0033291705006628

Johns, L. C., Gregg, L., Vythelingum, N., & McGuire, P. K. (2003). Establishing the reliability of a verbal self-monitoring paradigm. *Psychopathology*, *36*(6), 299–303. https://doi.org/10.1159/000075188

Johns, L. C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., & McGuire, P. K. (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine*, *31*(4), 705–715. https://doi.org/10.1017/S0033291701003774

Johnson, J. F., Belyk, M., Schwartze, M., Pinheiro, A. P., & Kotz, S. A. (2019). The role of the cerebellum in adaptation: ALE meta-analyses on sensory feedback error. *Human Brain Mapping*, *40*(13), 3966–3981. https://doi.org/10.1002/hbm.24681

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science*, *16*(3), 307–354. https://doi.org/10.1207/s15516709cog1603_1

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: An fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience*, *3*(3), 218–223. https://doi.org/10.1093/scan/nsn014

Kawahara, H. (2003). Exemplar-based voice quality analysis and control using a high quality auditory morphing procedure based on STRAIGHT. *Parameters*, *4*(5), 2. Retrieved from https://www.isca-speech.org/archive_open/archive_papers/voqual03/voq3_109.pdf

Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoustical Science and Technology*, *27*(6), 349–353. https://doi.org/10.1250/ast.27.349

Kearney, E., & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Language, Cognition and Neuroscience*, *34*(9), 1214–1229. https://doi.org/10.1080/23273798.2019.1589541

Kimura, M., & Yotsumoto, Y. (2018). Auditory traits of "own voice". *PLoS One*, *13*(6), e0199443. https://doi.org/10.1371/journal.pone.0199443

Knolle, F., Schröger, E., & Kotz, S. A. (2013). Prediction errors in self- and externally-generated deviants. *Biological Psychology*, *92*(2), 410–416. https://doi.org/10.1016/j.biopsycho.2012.11.017

Knolle, F., Schwartze, M., Schröger, E., & Kotz, S. A. (2019). Auditory predictions and prediction errors in response to self-initiated vowels. *Frontiers in Neuroscience*, *13*, 1146. https://doi.org/10.3389/fnins.2019.01146

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, *86*(3), 366–376. https://doi.org/10.1016/S0093-934X(02)00532-1

Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in voice quality perception. *Journal of Speech, Language, and Hearing Research*, *35*(3), 512–520. https://doi.org/10.1044/jshr.3503.512

Kreiman, J., & Sidtis, D. (2011). Foundations of voice studies: An interdisciplinary approach to voice production and perception. John Wiley & Sons, https://doi.org/10.1002/9781444395068

Kreitewolf, J., Gaudrain, E., & von Kriegstein, K. (2014). A neural mechanism for recognizing speech spoken by different speakers. *NeuroImage*, *91*, 375–385. https://doi.org/10.1016/j.neuroimage.2014.01.005

Kumari, V., Antonova, E., Fannon, D., Peters, E. R., Ffytche, D. H., Premkumar, P., & Williams, S. R. C. (2010). Beyond dopamine: Functional MRI predictors of responsiveness to cognitive behaviour therapy for psychosis. *Frontiers in Behavioural Neuroscience*, *4*(4), 1–10. https://doi.org/10.3389/neuro.08.004.2010

Kumari, V., Fannon, D., Ffytche, D. H., Raveendran, V., Antonova, E., Premkumar, P., & Johns, L. C. (2010). Functional MRI of verbal self-monitoring in schizophrenia: Performance and illness-specific effects. *Schizophrenia Bulletin*, *36*(4), 740–755. https://doi.org/10.1093/schbul/sbn148

Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, *21*(4), R143–R145. https://doi.org/10.1016/j.cub.2010.12.033

Latinus, M., & Belin, P. (2012). Perceptual auditory aftereffects on voice identity using brief vowel stimuli. *PLoS One*, *7*(7), https://doi.org/10.1371/journal.pone.0041384

Latinus, M., McAleer, P., Bestelmeyer, P. E., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, *23*(12), 1075–1080. https://doi.org/10.1016/j.cub.2013.04.055

Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin & Review*, *26*(1), 90–102. https://doi.org/10.3758/s13423-018-1497-7

Lavner, Y., Gath, I., & Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication*, *30*(1), 9–26. https://doi.org/10.1016/S0167-6393(99)00028-X

Lavner, Y., Rosenhouse, J., & Gath, I. (2001). The prototype model in speaker identification by human listeners. *International Journal of Speech Technology*, *4*(1), 63–74. https://doi.org/10.1023/A:1009656816383

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). emmeans: Estimated marginal means. *R Package Version*, *1*(4), 4. Retrieved from https://CRAN.R-project.org/package=emmeans

Leube, D. T., Knoblich, G., Erb, M., Grodd, W., Bartels, M., & Kircher, T. T. (2003). The neural correlates of perceiving one's own movements. *NeuroImage*, *20*(4), 2084–2090. https://doi.org/10.1016/j.neuroimage.2003.07.033

Maguinness, C., Roswandowitz, C., & von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia*, *116*, 179–193. https://doi.org/10.1016/j.neuropsychologia.2018.03.039

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, *119*(3), 907–917. https://doi.org/10.1093/brain/119.3.907

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279. https://doi.org/10.1016/S0893-6080(96)00035-4

Moerel, M., De Martino, F., & Formisano, E. (2012). Processing of natural sounds in human auditory cortex: Tonotopy, spectral tuning, and relation to voice sensitivity. *Journal of Neuroscience*, *32*(41), 14205–14216. https://doi.org/10.1523/JNEUROSCI.1388-12.2012

Mullennix, J. W., Ross, A., Smith, C., Kuykendall, K., Conard, J., & Barb, S. (2011). Typicality effects on memory for voice: Implications for earwitness testimony. *Applied Cognitive Psychology*, *25*(1), 29–34. https://doi.org/10.1002/acp.1635

Müller-Preuss, P., & Ploog, D. (1981). Inhibition of auditory cortical neurons during phonation. *Brain Research*, *215*(1–2), 61–76. https://doi.org/10.1016/0006-8993(81)90491-1

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., & Kojima, S. (2001). Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia*, *39*(10), 1047–1054. https://doi.org/10.1016/S0028-3932(01)00037-9

Nolan, F., McDougall, K., & Hudson, T. (2011). Some acoustic correlates of perceived (dis)similarity between same-accent voices. *ICPhS*, *17*, 1506–1509.Retrieved from http://icphs2011.hk.lt.cityu.edu.hk/resources/OnlineProceedings/RegularSession/Nolan/Nolan.pdf

Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, *265*(2), 119–122. https://doi.org/10.1016/S0304-3940(99)00218-9

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., & Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *NeuroImage*, *119*, 164–174. https://doi.org/10.1016/j.neuroimage.2015.06.050

Petkov, C. I., & Vuong, Q. C. (2013). Neuronal coding: The value in having an average voice. *Current Biology*, *23*(12), R521–R523. https://doi.org/10.1016/j.cub.2013.04.077

Pinheiro, A. P., Farinha-Fernandes, A., Roberto, M. S., & Kotz, S. A. (2019). Self-voice perception and its relationship with hallucination predisposition. *Cognitive Neuropsychiatry*, *24*(4), 237–255. https://doi.org/10.1080/13546805.2019.1621159

Pinheiro, A. P., Rezaii, N., Nestor, P. G., Rauber, A., Spencer, K. M., & Niznikiewicz, M. (2016). Did you or I say pretty, rude or brief? An ERP study of the effects of speaker's identity on emotional word processing. *Brain and Language*, *153*, 38–49. https://doi.org/10.1016/j.bandl.2015.12.003

Pinheiro, A. P., Rezaii, N., Rauber, A., & Niznikiewicz, M. (2016). Is this my voice or yours? The role of emotion and acoustic quality in self-other voice discrimination in schizophrenia. *Cognitive Neuropsychiatry*, *21*(4), 335–353. https://doi.org/10.1080/13546805.2016.1208611

Pinheiro, A. P., Schwartze, M., & Kotz, S. A. (2018). Voice-selective prediction alterations in nonclinical voice hearers. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-32614-9

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665–678. https://doi.org/10.1016/j.neuron.2011.09.006

R Development Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in sensorimotor control and integration. *Hearing Research*, *271*(1–2), 16–25. https://doi.org/10.1016/j.heares.2010.09.001

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *NeuroReport*, *11*(9), 1997–2000. https://doi.org/10.1097/00001756-200006260-00038

Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition*, *68*(2), 204–214. https://doi.org/10.1016/j.bandc.2008.04.007

Roswandowitz, C., Mathias, S. R., Hintz, F., Kreitewolf, J., Schelinski, S., & von Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments. *Current Biology*, *24*(19), 2348–2353. https://doi.org/10.1016/j.cub.2014.08.048

Sakai, N., Masuda, S., Shimotomai, T., & Mori, K. (2009). Brain activation in adults who stutter under delayed auditory feedback: An fMRI study. *International Journal of Speech-Language Pathology*, *11*(1), 2–11. https://doi.org/10.1080/17549500802588161

Sapara, A., ffytche, D. H., Cooke, M. A., Williams, S. C. R., & Kumari, V. (2015). Is it me? Verbal self-monitoring neural network and clinical insight in schizophrenia. *Psychiatry Research: Neuroimaging*, *234*(3), 328–335. https://doi.org/10.1016/j.pscychresns.2015.10.007

Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience*, *11*(11), 1812–1822. https://doi.org/10.1093/scan/nsw089

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language, and Hearing Research*, *40*(2), 453–463. https://doi.org/10.1044/jslhr.4002.453

Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014). Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(1), 15–25. https://doi.org/10.1002/wcs.1261

Schweinberger, S. R., Walther, C., Zäske, R., & Kovács, G. (2011). Neural correlates of adaptation to voice identity. *British Journal of Psychology*, *102*(4), 748–764. https://doi.org/10.1111/j.2044-8295.2011.02048.x

Seeley, W. W. (2010). Anterior insula degeneration in frontotemporal dementia. *Brain Structure and Function*, *214*(5–6), 465–475. https://doi.org/10.1007/s00429-010-0263-z

Siman-Tov, T., Granot, R. Y., Shany, O., Singer, N., Hendler, T., & Gordon, C. R. (2019). Is there a prediction network? Meta-analytic evidence for a cortical-subcortical network likely subserving prediction. *Neuroscience & Biobehavioral Reviews*, *105*, 262–275. https://doi.org/10.1016/j.neubiorev.2019.08.012

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2015). afex: Analysis of factorial experiments. *R Package Version*, 13–145. Retrieved from https://cran.r-project.org/web/packages/afex/index.html

Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., & Ford, J. M. (2013). Auditory cortex processes variation in our own speech. *PLoS One*, *8*(12), e82925. https://doi.org/10.1371/journal.pone.0082925

Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society of America*, *118*(5), 3177–3186. https://doi.org/10.1121/1.2047107

Smith, D. R., Walters, T. C., & Patterson, R. D. (2007). Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. *The Journal of the Acoustical Society of America*, *122*(6), 3628–3639. https://doi.org/10.1121/1.2799507

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., & Niazy, R. K. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Sommer, I. E., Diederen, K. M., Blom, J. D., Willems, A., Kushan, L., Slotema, K., & Kahn, R. S. (2008). Auditory verbal hallucinations predominantly activate the right inferior frontal area. *Brain*, *131*(12), 3169–3177. https://doi.org/10.1093/brain/awn251

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

Toyomura, A., Koyama, S., Miyamoto, T., Terao, A., Omori, T., Murohashi, H., & Kuriki, S. (2007). Neural correlates of auditory feedback control in human. *Neuroscience*, *146*(2), 499–503. https://doi.org/10.1016/j.neuroscience.2007.02.023

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain and Cognition*, *1*(2), 185–195. https://doi.org/10.1016/0278-2626(82)90016-1

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex*, *24*(2), 195–209. https://doi.org/10.1016/S0010-9452(88)80029-7

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia*, *25*(5), 829–834. https://doi.org/10.1016/0028-3932(87)90120-5

Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced suppression in auditory cortex. *BMC Neuroscience*, *10*(1), 58. https://doi.org/10.1186/1471-2202-10-58

Vermissen, D., Janssen, I., Johns, L., McGuire, P., Drukker, M., Campo, J., Myin-Germeys, I., Van Os, J., & Krabbendam, L. (2007). Verbal self-monitoring in psychosis: A non-replication. *Psychological Medicine*, *37*(4), 569–576. https://doi.org/10.1017/s0033291706009780

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research*, *17*(1), 48–55. https://doi.org/10.1016/S0926-6410(03)00079-X

von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, *22*(2), 948–955. https://doi.org/10.1016/j.neuroimage.2004.02.020

Wang, J., Mathalon, D. H., Roach, B. J., Reilly, J., Keedy, S. K., Sweeney, J. A., & Ford, J. M. (2014). Action planning and predictive coding when speaking. *NeuroImage*, *91*, 91–98. https://doi.org/10.1016/j.neuroimage.2014.01.003

Watkins, K., Patel, N., Davis, S., & Howell, P. (2005). Brain activity during altered auditory feedback: An FMRI study in healthy adolescents. *NeuroImage*, *26*(Supp 1), 304. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2801072/

Whitford, T. J., Mathalon, D. H., Shenton, M. E., Roach, B. J., Bammer, R., Adcock, R. A., Bouix, S., Kubicki, M., De Siebenthal, J., Rausch, A. C., Schneiderman, J. S., & Ford, J. M. (2011). Electrophysiological and diffusion tensor imaging evidence of delayed corollary discharges in patients with schizophrenia. *Psychological Medicine*, *41*(5), 959–969. https://doi.org/10.1017/S0033291710001376

Wickham, H. (2017). tidyverse: Easily Install and Load the 'Tidyverse'. *R Package Version*, *1*(2), 1. Retrieved from https://cran.r-project.org/web/packages/tidyverse/index.html

Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. *NeuroImage*, *39*(2), 885–893. https://doi.org/10.1016/j.neuroimage.2007.09.028

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 209–216. https://doi.org/10.1016/S1364-6613(97)01070-X

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7–8), 1317–1329. https://doi.org/10.1016/S0893-6080(98)00066-5

Wong, P. C., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *Journal of Cognitive Neuroscience*, *16*(7), 1173–1184. https://doi.org/10.1162/0898929041920522

Xu, M., Homae, F., Hashimoto, R. I., & Hagiwara, H. (2013). Acoustic cues for the recognition of self-voice and other-voice. *Frontiers in Psychology*, *4*, 735. https://doi.org/10.3389/fpsyg.2013.00735

Xue, G., Aron, A. R., & Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex*, *18*(8), 1923–1932. https://doi.org/10.1093/cercor/bhm220

Zarate, J. M., & Zatorre, R. J. (2005). Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *Annals of the New York Academy of Sciences*, *1060*, 404–408. https://doi.org/10.1196/annals.1360.058

Zheng, Z. Z., MacDonald, E. N., Munhall, K. G., & Johnsrude, I. S. (2011). Perceiving a stranger's voice as being one's own: A 'rubber voice' illusion? *PLoS One*, *6*(4), e18655. https://doi.org/10.1371/journal.pone.0018655

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, *22*(8), 1770–1781. https://doi.org/10.1162/jocn.2009.21324

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.