

1 **Expectancy changes the self-monitoring of voice identity**

2

3 **Authors:**

4 Joseph F. Johnson¹, Michel Belyk², Michael Schwartze¹, Ana P. Pinheiro³, Sonja A. Kotz^{1,4}

5

6 **Affiliations:**

7 ¹ University of Maastricht, Department of Neuropsychology and Psychopharmacology, the
8 Netherlands

9 ² University College London, Division of Psychology and Language Sciences, London, the United
10 Kingdom

11 ³ Faculdade de Psicologia, Universidade de Lisboa, Lisboa, Portugal

12 ⁴ Max Planck Institute for Human and Cognitive Sciences, Leipzig, Germany

13

14 **Corresponding Author:** Sonja A. Kotz: Maastricht University, Universiteitssingel 40, 6229 ER
15 Maastricht, Netherlands. +31 (0)433881653. sonja.kotz@maastrichtuniversity.nl

16

17 **Running Title:** Expectancy in self-voice

18

19 **Keywords:** fMRI, Source attribution, Voice morphing, Motor-induced suppression, Auditory
20 feedback

21 ABSTRACT

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

34 INTRODUCTION

35 Self-monitoring of the voice relies on comparing what we expect to hear and what we actually
36 hear (Frith, 1992; Wolpert et al., 1998). However, in a dynamic environment sensory feedback is
37 often ambiguous, e.g., when listening to multiple speakers. Any judgment of the voice source
38 further depends on how much sensory feedback deviates from expectations (Feinberg, 1978).
39 Minor deviations regarding one's own voice are typically self-attributed and used to compensate
40 motor control. Major deviations may lead to source-attributing the voice to another person. People
41 who experience auditory verbal hallucinations (AVH) show dysfunctional self-monitoring
42 (Kumari et al., 2010b; Sapara et al., 2015). For example, schizophrenia patients who experience
43 AVH are more likely to incorrectly attribute their voice to an external source in ambiguous
44 conditions that result in uncertainty among healthy individuals (Johns et al., 2001; Allen et al.,
45 2004; Pinheiro et al., 2016a). However, AVH are not limited to persons with psychosis but are
46 also experienced along a spectrum of hallucination proneness in healthy individuals (Baumeister
47 et al., 2017). An externalization bias observed within the general population may relate to higher
48 proneness to experience AVH in otherwise healthy individuals (Asai & Tanno, 2013; Pinheiro et
49 al., 2019). Functional neuroimaging studies of self-voice monitoring in the healthy brain have
50 examined the neural substrates of self-other voice attribution but have so far not examined
51 responses to uncertainty in ambiguous conditions (e.g. Allen et al., 2006; Fu et al., 2006). It is
52 critical to not only know how the brain establishes correct self and other voice attribution but also
53 where and how the voice is processed in conditions of uncertainty to gain a better understanding
54 of the mechanisms underlying dysfunctional self-monitoring.

55 Previous research has reported that unaltered self-voice production leads to reduced functional
56 brain activity in the auditory cortex (Christoffels, Formisano, & Schiller, 2007). This motor-
57 induced suppression (MIS) is compatible with findings of numerous studies employing diverse
58 methodology. It is similar to the N1 suppression effect, a modulation of the event-related potential
59 of the electroencephalogram (EEG) (e.g. Heinks-Maldonado, Mathalon, Gray, & Ford, 2005;
60 Behroozmand & Larson, 2011; Sitek et al., 2013; Wang et al., 2014; Pinheiro, Schwartz, & Kotz,
61 2018), or M1 suppression in magnetoencephalography (Numminen, Salmelin, & Hari, 1999;
62 Houde et al., 2002; Ventura, Nagarajan, & Houde, 2009), weakened activity in
63 electrocorticography and at intracranial electrodes (Greenlee et al., 2011; Chang et al., 2013), or
64 direct- and inter-cell recordings in non-human primates (Müller-Preuss & Ploog, 1981; Eliades &
65 Wang, 2008). Contrasting with the suppressed activity in auditory cortex, self-voice monitoring
66 activates a widespread system of functionally connected brain regions, including the inferior
67 frontal gyrus (IFG), supplementary motor area, insula, pre- and postcentral gyrus, inferior parietal
68 lobule (IPL), motor cortex, thalamus, and cerebellum (Christoffels, Formisano, & Schiller, 2007;
69 Behroozmand et al., 2015). The right anterior superior temporal gyrus (aSTG) and the adjacent
70 upper bank of the superior temporal sulcus (STS) likely play a critical role in voice identity
71 perception (Belin, Fecteau, & Bedard, 2004; von Kriegstein et al., 2003; von Kriegstein & Giraud,
72 2004; Belin and Zatorre, 2003). Patient studies support this assumption as lesions or damage to
73 the aSTG can lead to deficits in voice identity recognition (Gainotti, Ferraccioli, & Marra, 2010;
74 Gainotti & Marra, 2011; Hailstone et al., 2011; van Lancker & Kreiman, 1987; van Lancker &
75 Canter, 1982).

76 MIS in voice monitoring is not only effective in voice production but also in response to voice
77 recordings activated via a button press (Ford et al., 2007; Whitford et al., 2011; Pinheiro,

78 Schwartze, & Kotz, 2018; Knolle, Schwartze, Schröger, & Kotz, 2019) as well as for non-verbal
79 sounds including tones (e.g. Aliu, Houde, & Nagarajan, 2009; Baess, Widmann, Roye, Schröger,
80 & Jacobsen, 2009; Knolle, Schröger, & Kotz, 2013). Moreover, MIS seems to operate across
81 modalities of sensory feedback and arises from various motor effectors (e.g. Miall & Wolpert,
82 1996; Wolpert et al., 1998; Leube et al., 2003; Blakemore, Wolpert, & Frith, 1998). One
83 explanation for MIS is that internal models of expected action outcomes are fed-forward to the
84 relevant cortical regions to cancel out impending activity to the anticipated stimulus (Jordan &
85 Rumelhart, 1992; Miall & Wolpert, 1996; Wolpert, 1997). Studies that experimentally manipulate
86 sensory feedback create a mismatch between expected and actual outcome and indicate
87 concomitant modulation or absence of MIS under such circumstances. EEG studies typically show
88 this as decreased N1 suppression (e.g. Heinks-Maldonado et al., 2005; Behroozmand & Larson,
89 2011), while fMRI studies report a relative increase of STG activity when expected feedback is
90 altered (McGuire, Silbersweig, & Frith, 1996; Fu et al., 2006; Christoffels et al., 2007; Zheng et
91 al., 2010; Christoffels et al., 2011). With this approach, it is not only possible to make listeners
92 uncertain about self- or other-voice attribution (Allen et al., 2004, 2005, 2006; Fu et al., 2006;
93 Vermissen et al., 2007), but to also lead listeners to incorrectly attribute self-voice to another
94 speaker (Johns et al., 2001, 2003, 2006; Fu et al., 2006; Allen et al., 2004, 2005, 2006, Kumari et
95 al., 2010a, 2010b; Sapara et al., 2015). STG suppression only persists when the voice is correctly
96 judged as self-voice in distorted feedback conditions (Fu et al., 2006). Critically, data reflecting
97 uncertain voice attribution are often removed from fMRI analyses (Allen et al., 2005; Fu et al.,
98 2006). However, in order to gain a better understanding of voice attribution to internal or external
99 sources, it is mandatory to specifically focus on such data and to define how the known voice
100 attribution region of the STG reacts in conditions of uncertainty.

101 In addition to auditory cortex, activation in the right inferior frontal gyrus increases in response to
102 distorted auditory feedback (Johnson et al., 2019). However, while attenuation of the right aSTG
103 activation reflects expected stimulus quality, the right IFG is selectively responsive to unexpected
104 sensory events (Aron, Robbins, & Poldrack, 2004). Increased right IFG activity has been reported
105 when feedback is acoustically altered (Behroozmand et al., 2015, Fu et al., 2006; Toyomura et al.,
106 2007; Tourville et al., 2008; Guo et al., 2016), delayed (Sakai et al., 2009; Watkins et al., 2005),
107 replaced with the voice of another speaker (Fu et al., 2006), or physically perturbed during vocal
108 production (Golfinopoulos et al., 2010). In response to unexpected sensory feedback in voice
109 production, the right IFG produces a “salient signal”, indicating the potential need to stop and
110 respond to stimuli that may be affected by or the result of some external influence.
111 Correspondingly, It has been hypothesized that the processing of salient stimuli with minimal
112 divergence from expectations leads to an externalization bias that may manifest in the experience
113 of AVH (Sommer et al., 2008).

114 In the current fMRI experiment, we investigated how cortical voice identity and auditory feedback
115 monitoring regions respond in (un)certain self-other voice attribution. Participants elicited voice
116 stimuli that varied along a morphing continuum from self to other voice, including intermediate
117 voice samples of ambiguous identity. Region of interest (ROI) analyses motivated by our research
118 question and a priori hypotheses focussed on the right aSTG and the right IFG. The right aSTG
119 ROI stems from a well-replicated temporal voice area (TVA) localizer task (Belin et al., 2000).
120 The right IFG ROI conforms to a region responsive to experimental manipulation of auditory
121 feedback previously identified via activation-likelihood estimation (ALE) analyses (Johnson et al.,
122 2019). Due to possible individual variability in thresholds for self-other voice attribution (Asai &
123 Tanno, 2013), each participant underwent psychometric testing to determine individualized points

124 of maximum uncertainty on a continuum from self to other voice. The primary goal was to test
125 hypotheses that i) MIS of self-voice in the right aSTG is present, and the degree of suppression is
126 greater when attribution of the self is certain compared to when uncertain, and ii) right IFG
127 activation would increase in response to an increase in voice uncertainty. Confirmation of these
128 results would further substantiate previous EEG findings regarding MIS for self-voice elicited via
129 button-press as compared to passively heard (Ford et al., 2007; Whitford et al., 2011; Pinheiro,
130 Schwartz, & Kotz, 2018; Knolle, Schwartz, Schröger, & Kotz, 2019), indicating that suppressed
131 activity in auditory cortex aligns with predicted self-voice quality and not only as a function of
132 expected quality of voice feedback.

133 METHODS

134 PARTICIPANT RECRUITMENT

135 Twenty-seven participants took part in the study. The data of two participants were discarded due
136 to scanning artefacts. Of the remaining 25 (17 female), the average age was 21.88 years (SD =
137 4.37; range 18 to 33). Inclusion criteria assured that participants had no diagnosis of psychological
138 disorder, normal or corrected-to-normal hearing and vision, and no evidence of phonagnosia. This
139 was tested using an adapted version of a voice-name recognition test described below
140 (Roswandowitz et al., 2014). All participants gave informed consent and were compensated with
141 university study participant credit. This study was approved by the Ethical Review Committee of
142 the Faculty of Psychology and Neuroscience at Maastricht University (ERCPN-176_08_02_2017).

143 PROCEDURES

144 *PHONAGNOSIA SCREENING*

145 Phonagnosia is a disorder restricting individuals from perceiving speaker identity in voice (Van
146 Lancker et al., 1988). We screened for phonagnosia using an adapted version of a phonagnosia
147 screening task (see Roswandowitz et al., 2014). The task was composed of four rounds of
148 successive learning and testing phases, in which participants initially listened to the voices of three
149 speakers of the same gender. Identification of each speaker was subsequently tested 10 times with
150 response accuracy feedback provided during the first half of test trials. Finally, the task was
151 repeated with stimuli of the gender not used in the first run. Presentation order of these runs was
152 counterbalanced across participants.

153 *PSYCHOMETRIC TASK*

154 In a voice attribution task (VAT), participants heard samples of the vowels /a/ and /o/. These
155 samples varied in voice identity, which was morphed along a continuum from “self-voice” to
156 “other-voice” using the STRAIGHT voice morphing software package (Kawahara 2003, 2006)
157 running in MATLAB (R2019A, v9.6.0.1072779, The MathWorks, Inc., Natick, MA). For this
158 procedure, samples of the self-voice (SV) and other voice (OV) producing the two vowels were
159 obtained from each participant and normalized in duration (500ms) and amplitude (70db) using
160 the Praat software package (v6.0.28, <http://www.praat.org/>). The OV sample used matched the
161 gender of the participant. On this basis, 11 stimuli were created along a morphing spectrum in
162 steps of 10% morphing from SV to OV. In a two-alternative forced-choice (2AFC) task,
163 participants listened to each stimulus presented in random order and responded to the question: Is
164 the voice “more me” or “more other”? This procedure was repeated twice. In one run stimuli were
165 presented passively while in the other run participants were visually cued to press a button which
166 elicited the next stimulus (see Figure 1). This procedure was used to identify an individualized
167 point of maximum ambiguity (PMA) along the morphing spectrum for each participant. The PMA
168 was defined as the stimulus that was closest to chance level (50%) and used to inform subsequent
169 fMRI analyses.

170 *FMRI TASKS*

171 *Temporal Voice Area (TVA) Localizer:* To identify voice sensitive brain areas, participants were
172 scanned during a voice localizer task (Belin et al., 2000). This task is widely used to reliably probe
173 activity along the bilateral temporal cortices (e.g. Pernet et al., 2015) designated as anterior,
174 middle, and posterior TVA regions. Stimuli consisted of 8-second auditory clips with 20 vocal and

175 20 non-vocal sounds. In a single run, participants passively listened to these sounds and 20 silent
176 trials in pseudorandom order. Contrasting responses to vocal and non-vocal sounds identified brain
177 regions selectively sensitive to voice processing. The peak activation in the anterior STG of the
178 right hemisphere was then chosen as the voice-attribution ROIs in the subsequent fMRI analysis.

179 *Voice Perception Task (VPT):* Participants listened to passively presented or self-generated voice
180 stimuli. When shown a cue signifying the active button-press condition, participants pressed a
181 button to elicit voice stimuli, and conversely when shown a cue signifying the passive condition
182 were instructed to do nothing (Figure 2). In the active condition, half of the trials elicited a voice
183 following the button press, while in the other half no voice was presented. In the passive condition,
184 all trials involved the presentation of a voice. A subset of stimuli used in the VAT was selected for
185 the VPT, specifically the 100, 60, 50, 40, and 0% self-voice morphs. The intermediate steps of 60,
186 50, and 40% were selected as piloting revealed that individual PMA fell within a range of 35-65%
187 morphing, while morphs outside of this range produced high degrees of confidence in self vs. other
188 judgement. This ensured that every participant received the voice stimuli nearest to their subjective
189 PMA. Trial onsets were 9 seconds (+/- 500ms) apart to allow the BOLD response to return to
190 baseline before the presentation of the next stimulus started. To avoid the effects of adaptation
191 suppression (Andics et al., 2010, 2013; Belin et al., 2003; Latinus & Belin., 2011; Wong et al.,
192 2004), voice conditions were presented in a random order. Over two runs, a total of 100 trials were
193 presented in each condition of Source (Active and Passive). Within each condition of Source, each
194 voice stimulus (100, 60, 50, 40, and 0% morphs from self-to-other) was heard 20 times. 20 null
195 trials were included to provide a baseline comparison of activity in response to experimental trials.

196 FMRI DATA ACQUISITION AND ANALYSIS

197 Data acquisition was performed at a Siemens 3T Magnetom Prisma Fit Magnetic Resonance
198 Imaging (MRI) scanner at Scannexus facilities (Maastricht, NE), equipped with a 32-channel head
199 coil (Siemens Healthcare, Erlangen, Germany). A structural whole brain T1-weighted single-shot
200 echoplanar imaging (EPI) sequence was collected for each participant (field of view (FOV)
201 256mm; 192 axial slices; 1mm slice thickness; 1 x 1 x 1mm voxel size; repetition time (TR) of
202 2250ms seconds; echo-time (TE) 2.21ms). Two functional tasks were conducted with T2-weighted
203 EPI scans (FOV 208mm; 60 axial slices; 2mm slice thickness; 2 x 2 x 2mm voxel size; TE 30ms;
204 flip angle = 77°). Both tasks applied a long inter-acquisition interval where time between
205 consecutive image acquisition (2000ms) was delayed, resulting in a TR of 10 and 9 seconds for
206 the TVA localizer and VPT, respectively. This allowed auditory stimuli to be presented during a
207 period of relative silence to reduce noise artifacts and for volume acquisition to proceed during a
208 period of peak activation in the auditory cortex (Belin et al., 1999; Hall et al., 1999).

209 DICOM image data was converted to 4D NIFTI format using the Dcm2Nii converter provided in
210 the MRICron software package (<https://www.nitrc.org/projects/mricron/>). The topup tool (Smith,
211 et al., 2004) implemented in FSL (www.fmrib.ox.ac.uk/fsl) was used to estimate and correct for
212 susceptibility induced image distortions. Pre-processing was performed using SPM12 (Wellcome
213 Department of Cognitive Neurology, London, UK). A pre-processing pipeline applied slice timing
214 correction, realignment and unwarping, segmentation, normalization to standard Montreal
215 Neurological Institute (MNI) space (Fonov et al., 2009) as well as smoothing with a full width at
216 half maximum (FWHM) 8mm isotropic Gaussian kernel.

217 *General Linear Model (GLM) Analysis:* The TVA localizer and experimental VPT fMRI data were
218 analyzed with a standard two-level procedure in SPM12. For the TVA localizer, contrast images

219 for Vocal > Non-Vocal and Vocal > Silent were estimated for each participant. To test for the main
220 effect of interest, conjunction analysis $((V > NV) \cap (V > S))$ was performed. A second level
221 random-effects analysis tested for group-level significance. A first-level fixed-effects GLM of the
222 VPT data calculated contrast estimates for each participant. Contrast estimates were then used in
223 the subsequent hypothesis-driven ROI analysis to investigate TVA activity.

224 *Linear Mixed Model (LMM) ROI Analyses:* Two spherical (5mm) ROIs were selected for analysis:
225 the right aSTG/S in Brodmann Area (BA) 22 (MNI coordinates x 58, y 2, z -10) defined by the
226 TVA fMRI localizer task, and the right IFG opercular region in BA 44 (MNI coordinates x 46, y
227 10, z 4) (See Figure 3). A 2x3 factorial design was formulated using the factors of Source and
228 Voice. The two-leveled factor Source included self-generated (Active: A) and passively-heard
229 (Passive: P) playback of voice recordings. The three-leveled factor Voice included self-identified
230 voice (Self-voice: SV), externally-identified voice (Other-voice: OV), and voice of ambiguous
231 identity (Uncertain: UV) unattributed to self or external.

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

239 The main effects of Voice, Source and their interaction were tested with the afex package using
240 Kenward-Rogers degrees of freedom (Singmann et al., 2015). Estimated marginal means and

241 confidence intervals were computed with the emmeans package (Lenth, 2020) for visualization.

242 All p-values are corrected for multiple comparisons controlling at a false-discovery rate (FDR) of

243 0.05.

244 VAT RESULTS

245 Psychometric analysis of the VAT indicated little variability in the degree of morphing between
246 SV and OV required to elicit responses at chance level (50%), which we identified as the point of
247 maximum ambiguity. For the A condition, nine participants had PMAs at 40%, eight at 50% and
248 ten at 60% morphing. In the passive condition, eleven required 40%, seven 50%, and nine 60%
249 morphing. There was no significant difference between the average morphing required to elicit
250 PMA in A (μ 50%, SD 0.085) and P (μ 50%, SD 0.087) conditions. Although no participant
251 matched criteria for phonagnosia as specified by the screening task, VAT data from one participant
252 was excluded due to an inability to reliably differentiate between their own voice and other voices.

253 TVA LOCALIZER RESULTS

254 The TVA fMRI localizer produced four significant cluster-level activations (see Table 1 for
255 details). Within two large bilateral STG (BA 22) clusters, each included three peak-level
256 significant activations. These peaks correspond to the posterior (pSTG), middle (mSTG), and
257 aSTG. Two smaller clusters were found in the right precentral gyrus (BA 6), the left IFG (BA 44),
258 and the left IPL (BA 40). All significant cluster- and peak-level coordinates reported survived an
259 FDR correction of 0.05. The right aSTG peak was chosen for ROI analyses as the voice-attribution
260 ROI. These results replicate the pattern of TVA regions of peak activity (e.g. Belin et al., 2000;
261 Fecteau et al., 2004; Latinus et al., 2013; Pernet et al., 2015).

262 LMM ROI RESULTS:

263 Linear mixed model analysis of the right aSTG (Table 2A, Figure 4A) produced an FDR-corrected
264 significant main effect for the factor of voice, ($F(2, 118.94) = 4.90, p = 0.021$). No significant

265 effect was observed for source ($F(1, 118.92) = 0.53, p = 0.47$). A trend for the expected interaction
266 effect between voice and source was observed, although did not survive FDR correction for
267 multiple comparisons ($F(2, 118.94) = 3.40, p = 0.065$). However, based on our hypotheses and the
268 observed trend we conducted an exploratory post-hoc analysis to test the hypothesis that the
269 contrast $A > P$ differs for SV stimuli as compared to stimuli with OV or UV identities. This was
270 confirmed by the finding that motor-induced suppression is observed preferentially for SV stimuli
271 ($t(119) = -2.7, p = 0.021$).

272 The LMM analysis was repeated for the right IFG ROI (Table 2B, Figure 4B). A significant FDR-
273 corrected main effect of source was observed ($F(1, 116.04) = 9.93, p = 0.002$). No main effect
274 was found for the factor of voice ($F(2, 115.95) = 1.52, p = 0.26$), and no interaction between voice
275 and source were observed ($F(2, 115.81) = 1.60, p = 0.26$).

276 DISCUSSION

277 The current study investigated the interplay of auditory feedback regions involved in the
278 processing of (un)certainty in self-voice attribution, and unexpected quality of voice feedback. We
279 report the first fMRI evidence congruent with EEG reports that indicates self-voice MIS can be
280 observed even when voice stimuli are elicited by a button press rather than spoken. The predictable
281 qualities learned by long-term experience with self-voice feedback therefore are sufficient to
282 modulate MIS. Importantly, this effect was specific to vocal sounds matching the timbre of the
283 participant's own voice and was not observed when hearing the voice of another or being uncertain
284 about a speaker. The right IFG pars opercularis showed increased activation in response to self-
285 initiated voice relative to passive exposure. It is plausible that this differential response pattern is
286 driven by the higher proportion of voice trials not attributed to oneself. This region is known to be
287 more active when perceived stimuli are in conflict with expected sensory feedback. Together, these
288 findings suggest a differentiation between and an potential interplay of right IFG and aSTG in
289 voice processing, and more specifically feedback monitoring of self-generated voice and
290 differentiation of self- and other attribution.

291 VOICE IDENTITY AND MOTOR-INDUCED SUPPRESSION IN THE STG

292 Our results confirm right aSTG/S involvement in processing voice identity and indicate that it may
293 play a particular role in segregating the speaker's voice from external voices when monitoring
294 auditory feedback. We replicate previous TVA findings that the STG and upper bank of the STS
295 contain three bilateral voice patches (Table 1) (Belin et al., 2000; Pernet et al., 2015). The
296 processing of speech-related linguistic ("what") features have been attributed predominantly to the
297 left hemisphere, while speaker-related paralinguistic ("who") feature processing has been

298 attributed predominantly to the right hemisphere (Belin et al., 2002; Formisano et al., 2008;
299 Moerel, De Martino, & Formisano, 2012; Grandjean et al., 2005; Ethofer et al., 2006; Schirmer &
300 Kotz, 2006; Ethofer et al., 2007; Wiethoff et al., 2008; Kotz et al., 2003). Moreover, right
301 hemisphere paralinguistic processing of speaker-identity has been localized to the anterior region
302 of the STG/S (Belin et al., 2003; von Kriegstein et al., 2003; von Kriegstein & Giraud, 2004;
303 Fecteau et al., 2004; Latinus et al., 2013; Schelenski, Borowiak, & von Kriegstein, 2016). Different
304 low-level acoustics used in voice identity perception are processed in the pSTG, the extracted cues
305 relevant for speaker identification are then processed in the mSTG, and finally differential
306 processing of voice identity occurs in aSTG (Maguiness et al., 2019). To provide sufficient
307 duration for the extraction of paralinguistic speaker-related features, steady 500ms vowel excerpts
308 were chosen as voice samples in our study (Pinheiro et al., 2018; Schweinberger et al., 2011;
309 Schweinberger, Herholz, & Sommer, 1997; Van Berkum, van den Brink, Tesink, Kos, & Hagoort,
310 2008). Although vowels provide fundamental cues that allow differentiating between speakers
311 (Belin, Fecteau, & Bedard, 2004; Kreiman & Sidtis, 2011; Latinus & Belin, 2011; Schweinberger
312 et al., 2014), to the best of our knowledge no study has yet confirmed whether such basic stimuli
313 carry enough identity cues to allow for explicit self-recognition (Conde, Goncalves, & Pinheiro,
314 2018). We conducted ROI analyses on voice identity processing only in the right aSTG due to its
315 responsiveness to variation in voice identity and did not include other TVA regions in our analysis.
316 In doing so, this allowed us to detect fine-grain differences in activation patterns influenced by
317 only identity processing in a region that is highly active in the perception of the voice. We
318 confirmed prior to fMRI testing via psychometric analysis on behavioural data the ability for
319 participants to correctly attribute voice to self and other. Furthermore, we provide the first evidence

320 that 500ms steady vowel recordings of SV and OV allow for accurate recognition of self- and
321 other-attribution.

322 We observed that motor induced suppression in the right aSTG occurred only for SV. One possible
323 interpretation for this selective finding is that participants are most familiar with the acoustic
324 characteristics of their self-voice, and that they can therefore predict the features of their self-voice
325 more efficiently. Activity in cortical sensory processing regions activate more strongly by stimuli
326 that are unexpected than stimuli that are easily predicted. In the right aSTG, voice identity
327 processing is determined by the extent that speaker-related cues deviate from prototypes of
328 expected voice qualities (Mullenix et al., 2011; Bruckert et al., 2010; Andics et al., 2010, 2013;
329 Latinus & Belin, 2011; Latinus et al., 2013; Petkov & Vuong, 2013; Schweinberger et al., 2014).
330 These prototypes are learned through mean-based coding (Hoffman & Logothesis, 2009). While
331 it is clear that low-level acoustic processing is involved in this comparison (Smith & Patterson,
332 2005; Smith, Waters, & Patterson, 2007; Gaudrain et al., 2009; Baumann & Belin, 2010; Nolan,
333 McDougall, & Hudson, 2011; Zheng et al., 2011; Kreitewolf, Gaudrain, & von Kriegstein, 2014),
334 the specific features which drive identification vary from voice to voice (Lavner, Gath, &
335 Rosenhouse, 2000; Lavner, Rosenhouse, & Gath, 2001; Kreiman et al., 1992; Latinus & Belin,
336 2012; Xu et al., 2013). Furthermore, variability in the acoustic features of the voice do not only
337 exist between speakers, but also occurs within individual speakers (Lavan et al., 2019). Therefore,
338 increased experience with the voice of a specific speaker facilitates more efficient recognition of
339 voice identity. Indeed, people have the most experience with the qualities of their self-voice,
340 allowing for easy identification of their own identity, as little divergence from mean-based coding
341 is detected.

342 Alternatively, MIS of self-voice in a dynamic multi-speaker environment is important for the
343 segregation of internally- and externally-controlled voice stimuli. As verbal communication is
344 typically performed with the synchronized perception of one's own voice, the sound of self-voice
345 may therefore gain a privileged status that is also reflected in auditory feedback processing. During
346 vocalization, an efference copy of the motor command is sent from motor planning areas to
347 auditory and sensorimotor cortical regions to notify of impending feedback (Rauschecker & Scott,
348 2009; Rauschecker, 2011; Tourville & Guenther, 2011; Hickok, Houde, & Rong, 2011; Hickok,
349 2012; Kearney & Guenther, 2019). In specific, error-cells in the pSTG (planum temporale) receive
350 these signals from Broca's to remain inactive in response to the expected sound of self-voice, and
351 to engage when perceiving voice feedback outside the control of the speaker (Guenther et al.,
352 2006). To date, fMRI research using vocal feedback paradigms has provided evidence for this form
353 of MIS dependent on vocal production. For example, MIS has been reported for unaltered vocal
354 production relative to when hearing a recording of self-voice or when in a noisy environmental
355 (Christoffels et al., 2007), when acoustically distorted (McGuire, Silbersweig, & Frith, 1996; Fu
356 et al., 2006; Zheng et al., 2010; Christoffels et al., 2011) or replaced with the voice of another
357 speaker (McGuire, Silbersweig, & Frith, 1996; Fu et al., 2006). However, as these paradigms all
358 rely on vocal production, they are unable to indicate how the identity processing region of the STG
359 responds specifically to self-identity in voice during action. EEG research has provided evidence
360 for MIS in the auditory cortex that does not depend on vocal speech production as it is observed
361 even when sounds are elicited by a button press. For example, MIS of the N1 response was reported
362 for both, vocal (Heinks-Maldonado, Mathalon, Gray, & For, 2005; Behroozmand & Larson, 2011;
363 Sitek et al., 2013; Wang et al., 2014; Pinheiro, Schwartz, & Kotz, 2018) and button-press elicited
364 self-voice (Ford et al., 2007; Whitford et al., 2011; Pinheiro, Schwartz, & Kotz, 2018; Knolle,

365 Schwartze, Schröger, & Kotz, 2019). In line with previous EEG evidence, the current findings of
366 our button-press fMRI experiment in the voice identity auditory cortex ROI (right aSTG) indicate
367 suppressed activity in response to self-attributed voice during action. The reported MIS is specific
368 to self-voice processing, providing further evidence of voice identity suppression separate from
369 previously described cortical suppression during unperturbed speech. Importantly, this pattern was
370 observed only for voice attributed to oneself with certainty, and not present when voice was
371 distorted to an extent where self-attribution was uncertain.

372 EXPECTED FEEDBACK AND THE IFG

373 The right IFG was more strongly activated when participants generated vocal stimuli with a button
374 press as compared to passive perception. This finding confirms that this region is more responsive
375 to sounds triggered by the participant, potentially as part of auditory feedback. Increased activity
376 in this region has been observed in response to acoustically altered (Behroozmand et al., 2015, Fu
377 et al., 2006; Toyomura et al., 2007; Tourville et al., 2008; Guo et al., 2016), physically perturbed
378 (Golfopoulos et al., 2010), and externalized voice feedback (Fu et al., 2006).

379 In response to unexpected sensory information, the right IFG plays a crucial role in relaying salient
380 signals to attention networks. Moreover, the right IFG pars opercularis is part of a prediction
381 network, which forms expectations and detects unexpected sensory outcomes (Siman-Tov et al.,
382 2019). When prediction errors are detected, an inferior frontal network produces a salience
383 response (Cai et al., et al., 2014; Seeley, 2010; Power et al., 2011; Chang et al., 2013). Salience
384 signals engage ventral and dorsal attention networks, overlapping the right inferior frontal cortex.
385 The ventral attention network responds with bottom-up inhibition of ongoing action (Aron,
386 Robbins, & Poldrack, 2004, 2014), such as halting manual or speech movement (Aron & Poldrack,

387 2006; Aron, 2007; Chevrier et al., 2007; Xue et al., 2008). Correspondingly, damage to prefrontal
388 regions affects the ability one has in halting action in response to a stop signal (Aron et al., 2003),
389 and is similarly diminished when the pars opercularis is deactivated with TMS (Chambers et al.,
390 2006). The salience response may also engage the dorsal attention network to facilitate a top-down
391 response (Dosenbach et al., 2007; Eckert et al., 2009; Corbetta & Shulman, 2002; Fox et al., 2006),
392 for example, in goal-directed vocal compensation to pitch-shift (Riecker et al., 2000; Zarate and
393 Zatorre, 2005; Toyomura et al., 2007) or somatosensory perturbation (Golfinopoulos et al. 2011).
394 The localization of the right IFG ROI in the current study was determined by an ALE meta-analysis
395 on neuroimaging studies that experimentally manipulated auditory feedback from both vocal and
396 manual production (Johnson et al., 2019). As the current experimental design required no explicit
397 response to a change in stimulus quality, we hypothesized that increased activity in the right IFG
398 pars opercularis may represent the initial salience response to unexpected voice quality. However,
399 the effect of voice identity in the right IFG did not reach significance, and there was no significant
400 interaction between stimulus source and voice identity in this region. We note that the main effect
401 of source appears most strongly driven by unfamiliar or ambiguous voices, with an intermediate
402 level increase in the uncertain condition (see Figure 4B). It is possible that substantial variability
403 in the data limiting these results was due to the passive nature of the task with no overt attention
404 to the stimulus quality. As activity in this region is associated with attention and subsequent
405 inhibition/adaptation responses, the degree to which each participant attended to the change in
406 stimulus quality is unclear. Furthermore, although psychometric testing confirmed the subjective
407 ability of participants to correctly recognize voice as their own or another speaker's at a behaviour
408 level, it is possible that the brief vowel stimuli did not provide sufficient information to signal a
409 strong response to unexpected changes in self-voice. Further research is therefore needed to clarify

410 whether the right IFG is responsive to voice identity, and to which extent this may be driven by
411 the degree of salience elicited in divergence from expected qualities of self-voice.

412 VARIABILITY IN SELF-MONITORING THRESHOLDS

413 Although recordings of self-voice can produce a feeling of eeriness for listeners as compared to
414 when spoken (Kimura et al., 2018), people nevertheless recognize recorded voice samples as their
415 own (Nakamura et al., 2001; Kaplan et al., 2008; Rosa et al., 2008; Hughes & Nicholson, 2010;
416 Xu et al., 2013; Candini et al., 2014; Pinheiro et al., 2016a, 2016b, 2019). However, in ambiguous
417 conditions (i.e. acoustic distortion), the ability to accurately attribute a voice to oneself becomes
418 diminished (Allen et al., 2004, 2005, 2006, 2007; Fu et al., 2006; Kumari et al., 2010a, 2010b). As
419 ambiguity increases, an attribution threshold is passed, initiating a transition from uncertainty to
420 externalization (Johns et al., 2001, 2003, 2006; Vermissen et al., 2007). This threshold however
421 varies from person to person (Asai and Tanno, 2013). Here, it was therefore necessary to determine
422 the degree of morphing required to elicit uncertainty in the attribution of voice identity via separate
423 2AFC psychometric analysis for each participant. In doing so, we could confirm that fMRI
424 responses to the PMA condition were specific to the experience of maximum uncertainty,
425 regardless of any variability in the individual thresholds present in our healthy sample. The results
426 confirmed that participants were able to discriminate their self-voice from an unfamiliar voice,
427 with relatively little variation regarding the point of maximum ambiguity.

428 In contrast, it is known that persons with schizophrenia display a bias to misattribute self-voice to
429 an external source, both when they listen to recordings of their voice (Ilankovic et al., 2011;
430 Kambeitz-Ilankovic et al., 2013) and when they are speaking (Kumari et al., 2008, 2010b; Sapara
431 et al., 2015). This externalization bias is particularly prominent in schizophrenia patients who

432 experience AVH (Johns et al., 2001, 2006; Allen et al., 2004, 2007; Heinks-Maldonado et al.,
433 2007; Costafreda et al., 2008). Moreover, these individuals are highly confident in their
434 misattributions, as they are more likely to perceive a voice in ambiguous conditions as external
435 rather than remaining uncertain (Johns et al., 2001; Allen et al., 2004; Pinheiro et al., 2016a). It
436 was hypothesised that voice misattribution may underlie AVH as self-voice, either spoken aloud
437 or subvocalized, is mistaken for the voice of an external agent (Frith & Done, 1988; Bentall, 1990;
438 Brookwell, Bentall, & Varese, 2013). Correspondingly, as the severity of AVH symptoms
439 increase, accuracy in self-attribution voice diminishes (Allen et al., 2004, 2006; Pinheiro et al.,
440 2016a). Furthermore, the propensity to externalize self-voice has been linked to hypersalient
441 processing of auditory signals seen in persons with schizophrenia and other populations
442 experiencing AVH (Waters et al., 2012). Notably, this symptomology does not only exist within
443 patient groups. Individuals who are sub-clinical but at a high risk to develop psychosis, display
444 levels of self-monitoring performance similar to patients who meet a clinical diagnosis of
445 schizophrenia (Vermissem et al., 2007; Johns et al., 2010). Indeed, proneness to hallucinate is a
446 continuum and AVH are experienced in the general populations as well, although at lower rates
447 (Baumeister et al., 2017). Even in non-clinical populations, AVH are associated with a bias
448 towards external voice attributions (Asai & Tanno, 2013; Pinheiro et al., 2019). The current
449 findings may be of value in the understanding of the neural substrates underlying dysfunctional
450 self-other voice attribution. In light of our observation that the aSTG displays a qualitatively
451 different activation tendencies for self-voice relative to an unfamiliar voice and the hypothesized
452 influence of right IFG overactivity in salience detection in AVH, we suggest future research in
453 high risk groups to assess a possible abnormal interaction between these two regions. Structural
454 and functional connectivity MRI analysis may help explain if it is abnormalities in the

455 communication between these two regions, or individual disturbances in either or both regions that
456 leads to this symptomatology.

457 5. CONCLUSION

458 The goal of this experiment was to investigate how levels of self-voice certainty alter brain activity
459 in voice identity and feedback quality monitoring regions of the brain. By replicating earlier
460 findings using a voice area localizer task, we isolated a putative voice identity processing region
461 in the right aSTG. Our results indicate activity in this TVA is suppressed only when self-generating
462 a voice that is definitively attributed to oneself. Furthermore, in the right IFG pars opercularis
463 region responsive to unexpected feedback quality, we demonstrate increased activity while
464 monitoring voice during action relative to when passively heard. It is possible that this activity is
465 driven by salience responses to self-produced stimuli that do not match the expected quality of
466 self-voice. Using a novel self-monitoring paradigm, we provide the first fMRI evidence for the
467 effectiveness of button-press voice-elicitation in modulating an identity-related MIS in the
468 auditory cortex. Furthermore, we present novel findings on the effectiveness of brief vowel
469 excerpts to provide sufficient paralinguistic information to explicitly identify one's own voice
470 identity. Finally, we suggest a dynamic interaction between the roles of the right aSTG and IFG in
471 the voice self-monitoring network. One may speculate that the feedback monitoring frontal region
472 informs the temporal identity region whenever a salience threshold has been passed and voice
473 feedback is influenced by or under control of an external actor. The implications of variability in
474 the function of these mechanisms are particularly relevant to AVH and may provide specific
475 substrates for the symptomatology seen across the population, independent from broader neural
476 dysfunction associated with clinical pathology.

477 ACKNOWLEDGMENTS

478 This work has been supported by the Fundação Bial, Grant/Award Number: BIAL 238/16;
479 Fundação para a Ciência e a Tecnologia, Grant/Award Number: PTDC/MHC-PCN/0101/2014.
480 Further funding was provided by the Maastricht Brain Imaging Center, MBIC Funding Number:
481 F8000E14, F8000F14, F8042, F8051. We thank Lisa Goller for support in coordination and data
482 collection.

483 COMPETING INTERESTS

484 All authors disclose no potential sources of conflict of interest.

485 AUTHOR CONTRIBUTIONS

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

489 REFERENCES

- 490 1. Aliu, S. O., Houde, J. F., & Nagarajan, S. S. (2009). Motor-induced suppression of the
491 auditory cortex. *Journal of Cognitive Neuroscience*, 21(4), 791-802.
492 <https://doi.org/10.1162/jocn.2009.21055>
- 493 2. Allen, P. P., Johns, L. C., Fu, C. H., Broome, M. R., Vythelingum, G. N., & McGuire, P. K.
494 (2004). Misattribution of external speech in patients with hallucinations and delusions.
495 *Schizophrenia Research*, 69(2-3), 277-287. <https://doi.org/10.1016/j.schres.2003.09.008>
- 496 3. Allen, P. P., Amaro, E., Fu, C. H., Williams, S. C., Brammer, M., Johns, L. C., & McGuire,
497 P. K. (2005). Neural correlates of the misattribution of self-generated speech. *Human Brain*
498 *Mapping*, 26(1), 44-53. <https://doi.org/10.1002/hbm.20120>
- 499 4. Allen, P., Freeman, D., Johns, L., & McGuire, P. (2006). Misattribution of self-generated
500 speech in relation to hallucinatory proneness and delusional ideation in healthy volunteers.
501 *Schizophrenia Research*, 84(2-3), 281-288. <https://doi.org/10.1016/j.schres.2006.01.021>
- 502 5. Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., & Vidnyánszky, Z.
503 (2010). Neural mechanisms for voice recognition. *Neuroimage*, 52(4), 1528-1540.
504 <https://doi.org/10.1016/j.neuroimage.2010.05.048>
- 505 6. Andics, A., McQueen, J. M., & Petersson, K. M. (2013). Mean-based neural coding of
506 voices. *Neuroimage*, 79, 351-360. <https://doi.org/10.1016/j.neuroimage.2013.05.002>
- 507 7. Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-
508 signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature*
509 *Neuroscience*, 6(2), 115-116. <https://doi.org/10.1038/mn1003>

- 510 8. Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior
511 frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170-177.
512 <https://doi.org/10.1016/j.tics.2004.02.010>
- 513 9. Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal
514 response inhibition: role of the subthalamic nucleus. *Journal of Neuroscience*, 26(9), 2424-
515 2433. <https://doi.org/10.1523/JNEUROSCI.4682-05.2006>
- 516 10. Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The Neuroscientist*,
517 13(3), 214-228. <https://doi.org/10.1177/1073858407299288>
- 518 11. Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior
519 frontal cortex: one decade on. *Trends in Cognitive Sciences*, 18(4), 177-185.
520 <https://doi.org/10.1016/j.tics.2013.12.003>
- 521 12. Asai, T., & Tanno, Y. (2013). Why must we attribute our own action to ourselves? Auditory
522 hallucination like-experiences as the results both from the explicit self-other attribution and
523 implicit regulation in speech. *Psychiatry Research*, 207(3), 179-188.
524 <https://doi.org/10.1016/j.psychres.2012.09.055>
- 525 13. Baess, P., Widmann, A., Roye, A., Schröger, E., & Jacobsen, T. (2009). Attenuated human
526 auditory middle latency response and evoked 40-Hz response to self-initiated sounds.
527 *European Journal of Neuroscience*, 29(7), 1514-1521. [https://doi.org/10.1111/j.1460-
528 9568.2009.06683.x](https://doi.org/10.1111/j.1460-9568.2009.06683.x)
- 529 14. Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... &
530 Bolker, M. B. (2015). Package 'lme4'. *Convergence*, 12(1), 2. [http://cran.r-
531 project.org/package=lme4](http://cran.r-project.org/package=lme4)

- 532 15. Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: common dimensions
533 for different vowels and speakers. *Psychological Research*, 74(1), 110.
534 <https://doi.org/10.1007/s00426-008-0185-z>
- 535 16. Baumeister, D., Sedgwick, O., Howes, O., & Peters, E. (2017). Auditory verbal
536 hallucinations and continuum models of psychosis: a systematic review of the healthy voice-
537 hearer literature. *Clinical Psychology Review*, 51, 125-141.
538 <https://doi.org/10.1016/j.cpr.2016.10.010>
- 539 17. Behroozmand, R., & Larson, C. (2011). Motor-induced suppression of auditory neural
540 responses to pitch-shifted voice feedback. *The Journal of the Acoustical Society of America*,
541 129(4), 2454-2454. <https://doi.org/10.1121/1.3588063>
- 542 18. Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard III, M. A., &
543 Greenlee, J. D. (2015). Sensory–motor networks involved in speech production and motor
544 control: An fMRI study. *Neuroimage*, 109, 418-428.
545 <https://doi.org/10.1016/j.neuroimage.2015.01.040>
- 546 19. Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the
547 auditory cortex. *Neuroimage*, 10(4), 417-429. <https://doi.org/10.1006/nimg.1999.0480>
- 548 20. Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in
549 human auditory cortex. *Nature*, 403(6767), 309-312. <https://doi.org/10.1038/35002078>
- 550 21. Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds.
551 *Cognitive Brain Research*, 13(1), 17-26. [https://doi.org/10.1016/S0926-6410\(01\)00084-2](https://doi.org/10.1016/S0926-6410(01)00084-2)
- 552 22. Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal
553 lobe. *Neuroreport*, 14(16), 2105-2109. <https://doi.org/10.1097/00001756-200311140-00019>

- 554 23. Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice
555 perception. *Trends in Cognitive Sciences*, 8(3), 129-135.
556 <https://doi.org/10.1016/j.tics.2004.01.008>
- 557 24. Bentall, R. P. (1990). The illusion of reality: a review and integration of psychological
558 research on hallucinations. *Psychological Bulletin*, 107(1), 82-95.
559 <https://doi.org/10.1037/0033-2909.107.1.82>
- 560 25. Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-
561 produced tickle sensation. *Nature Neuroscience*, 1(7), 635-640. <https://doi.org/10.1038/2870>
- 562 26. Brookwell, M. L., Bentall, R. P., & Varese, F. (2013). Externalizing biases and hallucinations
563 in source-monitoring, self-monitoring and signal detection studies: a meta-analytic review.
564 *Psychological Medicine*, 43(12), 2465-2475. <https://doi.org/10.1017/S0033291712002760>
- 565 27. Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., ... &
566 Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology*, 20(2), 116-
567 120. <https://doi.org/10.1016/j.cub.2009.11.034>
- 568 28. Cai, W., Ryali, S., Chen, T., Li, C. S. R., & Menon, V. (2014). Dissociable roles of right
569 inferior frontal cortex and anterior insula in inhibitory control: evidence from intrinsic and
570 task-related functional parcellation, connectivity, and response profile analyses across
571 multiple datasets. *Journal of Neuroscience*, 34(44), 14652-14667.
572 <https://doi.org/10.1523/JNEUROSCI.3048-14.2014>
- 573 29. Candini, M., Zamagni, E., Nuzzo, A., Ruotolo, F., Iachini, T., & Frassinetti, F. (2014). Who
574 is speaking? Implicit and explicit self and other voice recognition. *Brain and Cognition*, 92,
575 112-117. <https://doi.org/10.1016/j.bandc.2014.10.001>

- 576 30. Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H.,
577 Robertson, I. H., ... & Mattingley, J. B. (2006). Executive “brake failure” following
578 deactivation of human frontal lobe. *Journal of Cognitive Neuroscience*, 18(3), 444-455.
579 <https://doi.org/10.1162/jocn.2006.18.3.444>
- 580 31. Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human
581 cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the*
582 *National Academy of Sciences*, 110(7), 2653-2658. <https://doi.org/10.1073/pnas.1216827110>
- 583 32. Chevrier, A. D., Noseworthy, M. D., & Schachar, R. (2007). Dissociation of response
584 inhibition and performance monitoring in the stop signal task using event-related fMRI.
585 *Human brain mapping*, 28(12), 1347-1358. <https://doi.org/10.1002/hbm.20355>
- 586 33. Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal
587 feedback processing: an fMRI study employing overt speech. *Human Brain Mapping*, 28(9),
588 868-879. <https://doi.org/10.1002/hbm.20315>
- 589 34. Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E., & Schiller, N. O. (2011).
590 The sensory consequences of speaking: parametric neural cancellation during speech in
591 auditory cortex. *PLoS One*, 6(5). <https://doi.org/10.1371/journal.pone.0018307>
- 592 35. Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2018). Stimulus complexity matters when
593 you hear your own voice: Attention effects on self-generated voice processing. *International*
594 *Journal of Psychophysiology*, 133, 66-78. <https://doi.org/10.1016/j.ijpsycho.2018.08.007>
- 595 36. Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven
596 attention in the brain. *Nature Reviews Neuroscience*, 3(3), 201-215.
597 <https://doi.org/10.1038/nrn755>

- 598 37. Costafreda, S. G., Brébion, G., Allen, P., McGuire, P. K., & Fu, C. H. Y. (2008). Affective
599 modulation of external misattribution bias in source monitoring in schizophrenia.
600 *Psychological Medicine*, 38(6), 821-824. <https://doi.org/10.1017/S0033291708003243>
- 601 38. Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R.
602 A., ... & Schlaggar, B. L. (2007). Distinct brain networks for adaptive and stable task control
603 in humans. *Proceedings of the National Academy of Sciences*, 104(26), 11073-11078.
604 <https://doi.org/10.1073/pnas.0704320104>
- 605 39. Eckert, M. A., Menon, V., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., & Dubno, J.
606 R. (2009). At the heart of the ventral attention system: the right anterior insula. *Human brain*
607 *mapping*, 30(8), 2530-2541. <https://doi.org/10.1002/hbm.20688>
- 608 40. Eliades, S. J., & Wang, X. (2008). Neural substrates of vocalization feedback monitoring in
609 primate auditory cortex. *Nature*, 453(7198), 1102-1106. <https://doi.org/10.1038/nature06910>
- 610 41. Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., ... & Wildgruber, D.
611 (2006). Cerebral pathways in processing of affective prosody: a dynamic causal modeling
612 study. *Neuroimage*, 30(2), 580-587. <https://doi.org/10.1016/j.neuroimage.2005.09.059>
- 613 42. Ethofer, T., Wiethoff, S., Anders, S., Kreifelts, B., Grodd, W., & Wildgruber, D. (2007). The
614 voices of seduction: cross-gender effects in processing of erotic prosody. *Social Cognitive*
615 *and Affective Neuroscience*, 2(4), 334-337. <https://doi.org/10.1093/scan/nsm028>
- 616 43. Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-
617 specific in human auditory cortex? An fMRI study. *Neuroimage*, 23(3), 840-848.
618 <https://doi.org/10.1016/j.neuroimage.2004.09.019>
- 619 44. Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its
620 disorders. *Schizophrenia bulletin*, 4(4), 636. <https://doi.org/10.1093/schbul/4.4.636>

- 621 45. Fonov, V. S., Evans, A. C., McKinstry, R. C., Almlil, C. R., & Collins, D. L. (2009).
622 Unbiased nonlinear average age-appropriate brain templates from birth to adulthood.
623 *NeuroImage*, 47, S102. [https://doi.org/10.1016/S1053-8119\(09\)70884-5](https://doi.org/10.1016/S1053-8119(09)70884-5)
- 624 46. Ford, J. M., Gray, M., Faustman, W. O., Roach, B. J., & Mathalon, D. H. (2007). Dissecting
625 corollary discharge dysfunction in schizophrenia. *Psychophysiology*, 44(4), 522-529.
626 <https://doi.org/10.1111/j.1469-8986.2007.00533.x>
- 627 47. Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). " Who" is saying" what"?
628 Brain-based decoding of human voice and speech. *Science*, 322(5903), 970-973.
629 <https://doi.org/10.1126/science.1164318>
- 630 48. Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006).
631 Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems.
632 *Proceedings of the National Academy of Sciences*, 103(26), 10046-10051.
633 <https://doi.org/10.1073/pnas.0604187103>
- 634 49. Frith, C. D., & Done, D. J. (1988). Towards a neuropsychology of schizophrenia. *The British*
635 *Journal of Psychiatry*, 153(4), 437-443. <https://doi.org/10.1192/bjp.153.4.437>
- 636 50. Frith, C. D. (1992). The cognitive neuropsychology of schizophrenia. *Psychology press*.
637 <https://doi.org/10.4324/9781315785011>
- 638 51. Fu, C. H., Vythelingum, G. N., Brammer, M. J., Williams, S. C., Amaro Jr, E., Andrew, C.
639 M., ... & McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: neural correlates
640 of auditory verbal feedback. *Cerebral Cortex*, 16(7), 969-977.
641 <https://doi.org/10.1093/cercor/bhj039>
- 642 52. Gainotti, G., Ferraccioli, M., & Marra, C. (2010). The relation between person identity
643 nodes, familiarity judgment and biographical information. Evidence from two patients with

- 644 right and left anterior temporal atrophy. *Brain Research*, 1307, 103-114.
- 645 <https://doi.org/10.1016/j.brainres.2009.10.009>
- 646 53. Gainotti, G., & Marra, C. (2011). Differential contribution of right and left temporo-occipital
647 and anterior temporal lesions to face recognition disorders. *Frontiers in Human*
648 *Neuroscience*, 5, 55. <https://doi.org/10.3389/fnhum.2011.00055>
- 649 54. Gaudrain, E., Li, S., Ban, V. S., & Patterson, R. D. (2009). The Role of Glottal Pulse Rate
650 and Vocal Tract Length in the Perception of Speaker Identity. *Interspeech 2009* 1(5), 152-
651 155. <https://doi.org/10.6084/m9.figshare.870509>
- 652 55. Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., &
653 Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback
654 perturbation during speech. *Neuroimage*, 55(3), 1324-1338.
655 <https://doi.org/10.1016/j.neuroimage.2010.12.065>
- 656 56. Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., &
657 Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless
658 speech. *Nature Neuroscience*, 8(2), 145-146. <https://doi.org/10.1038/nn1392>
- 659 57. Greenlee, J. D., Jackson, A. W., Chen, F., Larson, C. R., Oya, H., Kawasaki, H., ... &
660 Howard III, M. A. (2011). Human auditory cortical activation during self-vocalization. *PloS*
661 *One*, 6(3). <https://doi.org/10.1371/journal.pone.0014744>
- 662 58. Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the
663 cortical interactions underlying syllable production. *Brain and language*, 96(3), 280-301.
664 <https://doi.org/10.1016/j.bandl.2005.06.001>

- 665 59. Guo, Z., Huang, X., Wang, M., Jones, J. A., Dai, Z., Li, W., ... & Liu, H. (2016). Regional
666 homogeneity of intrinsic brain activity correlates with auditory-motor processing of vocal
667 pitch errors. *Neuroimage*, *142*, 565-575. <https://doi.org/10.1016/j.neuroimage.2016.08.005>
- 668 60. Hailstone, J. C., Ridgway, G. R., Bartlett, J. W., Goll, J. C., Buckley, A. H., Crutch, S. J., &
669 Warren, J. D. (2011). Voice processing in dementia: a neuropsychological and
670 neuroanatomical analysis. *Brain*, *134*(9), 2535-2547. <https://doi.org/10.1093/brain/awr205>
- 671 61. Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M.
672 R., ... & Bowtell, R. W. (1999). "Sparse" temporal sampling in auditory fMRI. *Human Brain*
673 *Mapping*, *7*(3), 213-223. [https://doi.org/10.1002/\(SICI\)1097-0193\(1999\)7:3<213::AID-
674 HBM5>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N)
- 675 62. Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of
676 auditory cortex during speech production. *Psychophysiology*, *42*(2), 180-190.
677 <https://doi.org/10.1111/j.1469-8986.2005.00272.x>
- 678 63. Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., &
679 Ford, J. M. (2007). Relationship of imprecise corollary discharge in schizophrenia to auditory
680 hallucinations. *Archives of General Psychiatry*, *64*(3), 286-296.
681 <https://doi.org/10.1001/archpsyc.64.3.286>
- 682 64. Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing:
683 computational basis and neural organization. *Neuron*, *69*(3), 407-422.
684 <https://doi.org/10.1016/j.neuron.2011.01.019>
- 685 65. Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature reviews*
686 *neuroscience*, *13*(2), 135-145. <https://doi.org/10.1038/nrn3158>

- 687 66. Hoffman, K. L., & Logothetis, N. K. (2009). Cortical mechanisms of sensory learning and
688 object recognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
689 364(1515), 321-329. <https://doi.org/10.1098/rstb.2008.0271>
- 690 67. Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the
691 auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, 14(8),
692 1125-1138. <https://doi.org/10.1162/089892902760807140>
- 693 68. Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition
694 of self-stimuli. *Consciousness and Cognition*, 19(4), 1124-1134.
695 <https://doi.org/10.1016/j.concog.2010.03.001>
- 696 69. Ilankovic, L. M., Allen, P. P., Engel, R., Kambeitz, J., Riedel, M., Müller, N., & Hennig-
697 Fast, K. (2011). Attentional modulation of external speech attribution in patients with
698 hallucinations and delusions. *Neuropsychologia*, 49(5), 805-812.
699 <https://doi.org/10.1016/j.neuropsychologia.2011.01.016>
- 700 70. Johns, L. C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., & McGuire, P. K.
701 (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with
702 schizophrenia. *Psychological Medicine*, 31(4), 705-715.
703 <https://doi.org/10.1017/S0033291701003774>
- 704 71. Johns, L. C., Gregg, L., Vythelingum, N., & McGuire, P. K. (2003). Establishing the
705 reliability of a verbal self-monitoring paradigm. *Psychopathology*, 36(6), 299-303.
706 <https://doi.org/10.1159/000075188>
- 707 72. Johns, L. C., Gregg, L., Allen, P., & McGuire, P. K. (2006). Impaired verbal self-monitoring
708 in psychosis: effects of state, trait and diagnosis. *Psychological Medicine*, 36(4), 465-474.
709 <https://doi.org/10.1017/S0033291705006628>

- 710 73. Johnson, J. F., Belyk, M., Schwartz, M., Pinheiro, A. P., & Kotz, S. A. (2019). The role of
711 the cerebellum in adaptation: ALE meta-analyses on sensory feedback error. *Human Brain*
712 *Mapping*, 40(13), 3966-3981. <https://doi.org/10.1002/hbm.24681>
- 713 74. Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal
714 teacher. *Cognitive Science*, 16(3), 307-354. https://doi.org/10.1207/s15516709cog1603_1
- 715 75. Kambeitz-Ilankovic, L., Hennig-Fast, K., Benetti, S., Kambeitz, J., Pettersson-Yeo, W.,
716 O'Daly, O., ... & Allen, P. (2013). Attentional modulation of source attribution in first-
717 episode psychosis: A functional magnetic resonance imaging study. *Schizophrenia Bulletin*,
718 39(5), 1027-1036. <https://doi.org/10.1093/schbul/sbs101>
- 719 76. Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the
720 senses: an fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective*
721 *Neuroscience*, 3(3), 218-223. <https://doi.org/10.1093/scan/nsn014>
- 722 77. Kawahara, H. (2003). Exemplar-based Voice Quality Analysis and Control using a High
723 Quality Auditory Morphing Procedure based on STRAIGHT. *Parameters*, 4(5), 2. Retrieved
724 from https://www.isca-speech.org/archive_open/archive_papers/voqual03/voq3_109.pdf
- 725 78. Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER:
726 Perceptually isomorphic decomposition of speech sounds. *Acoustical Science and*
727 *Technology*, 27(6), 349-353. <https://doi.org/10.1250/ast.27.349>
- 728 79. Kearney, E., & Guenther, F. H. (2019). Articulating: the neural mechanisms of speech
729 production. *Language, Cognition and Neuroscience*, 34(9), 1214-1229.
730 <https://doi.org/10.1080/23273798.2019.1589541>
- 731 80. Kimura M., Yotsumoto Y. (2018). Auditory traits of "own voice". *PLoS ONE* 13(6):
732 e0199443. <https://doi.org/10.1371/journal.pone.0199443>

- 733 81. Knolle, F., Schröger, E., & Kotz, S. A. (2013). Prediction errors in self-and externally-
734 generated deviants. *Biological Psychology*, *92*(2), 410-416.
735 <https://doi.org/10.1016/j.biopsycho.2012.11.017>
- 736 82. Knolle, F., Schwartz, M., Schröger, E., & Kotz, S. A. (2019). Auditory predictions and
737 prediction errors in response to self-initiated vowels. *Frontiers in Neuroscience*, *13*, 1146.
738 <https://doi.org/10.3389/fnins.2019.01146>
- 739 83. Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D.
740 (2003). On the lateralization of emotional prosody: an event-related functional MR
741 investigation. *Brain and Language*, *86*(3), 366-376. [https://doi.org/10.1016/S0093-](https://doi.org/10.1016/S0093-934X(02)00532-1)
742 [934X\(02\)00532-1](https://doi.org/10.1016/S0093-934X(02)00532-1)
- 743 84. Kreiman, J., Gerratt, B. R., Precoda, K., & Berke, G. S. (1992). Individual differences in
744 voice quality perception. *Journal of Speech, Language, and Hearing Research*, *35*(3), 512-
745 520. <https://doi.org/10.1044/jshr.3503.512>
- 746 85. Kreiman, J., & Sidtis, D. (2011). Foundations of voice studies: An interdisciplinary approach
747 to voice production and perception. John Wiley & Sons.
748 <https://doi.org/10.1002/9781444395068>
- 749 86. Kreitewolf, J., Gaudrain, E., & von Kriegstein, K. (2014). A neural mechanism for
750 recognizing speech spoken by different speakers. *Neuroimage*, *91*, 375-385.
751 <https://doi.org/10.1016/j.neuroimage.2014.01.005>
- 752 87. Kumari, V., Antonova, E., Fannon, D., Peters, E. R., Ffytche, D. H., Premkumar, P., ... &
753 Williams, S. R. C. (2010a). Beyond dopamine: functional MRI predictors of responsiveness
754 to cognitive behaviour therapy for psychosis. *Frontiers in Behavioural Neuroscience*, *4*(4), 1-
755 10. <https://doi.org/10.3389/neuro.08.004.2010>

- 756 88. Kumari, V., Fannon, D., Ffytche, D. H., Raveendran, V., Antonova, E., Premkumar, P., ... &
757 Johns, L. C. (2010b). Functional MRI of verbal self-monitoring in schizophrenia:
758 performance and illness-specific effects. *Schizophrenia bulletin*, 36(4), 740-755.
759 <https://doi.org/10.1093/schbul/sbn148>
- 760 89. Latinus, M., & Belin, P. (2011). Human voice perception. *Current Biology*, 21(4), R143-
761 R145. <https://doi.org/10.1016/j.cub.2010.12.033>
- 762 90. Latinus, M., & Belin, P. (2012). Perceptual auditory aftereffects on voice identity using brief
763 vowel stimuli. *PLoS One*, 7(7). <https://doi.org/10.1371/journal.pone.0041384>
- 764 91. Latinus, M., McAleer, P., Bestelmeyer, P. E., & Belin, P. (2013). Norm-based coding of
765 voice identity in human auditory cortex. *Current Biology*, 23(12), 1075-1080.
766 <https://doi.org/10.1016/j.cub.2013.04.055>
- 767 92. Lavner, Y., Gath, I., & Rosenhouse, J. (2000). The effects of acoustic modifications on the
768 identification of familiar voices speaking isolated vowels. *Speech Communication*, 30(1), 9-
769 26. [https://doi.org/10.1016/S0167-6393\(99\)00028-X](https://doi.org/10.1016/S0167-6393(99)00028-X)
- 770 93. Lavner, Y., Rosenhouse, J., & Gath, I. (2001). The prototype model in speaker identification
771 by human listeners. *International Journal of Speech Technology*, 4(1), 63-74.
772 <https://doi.org/10.1023/A:1009656816383>
- 773 94. Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). emmeans: estimated
774 marginal means. R package version 1.4. 4. Retrieved from [https://CRAN.R-](https://CRAN.R-project.org/package=emmeans)
775 [project.org/package=emmeans](https://CRAN.R-project.org/package=emmeans)
- 776 95. Leube, D. T., Knoblich, G., Erb, M., Grodd, W., Bartels, M., & Kircher, T. T. (2003). The
777 neural correlates of perceiving one's own movements. *Neuroimage*, 20(4), 2084-2090.
778 <https://doi.org/10.1016/j.neuroimage.2003.07.033>

- 779 96. Maguinness, C., Roswadowitz, C., & von Kriegstein, K. (2018). Understanding the
780 mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia*,
781 *116*, 179-193. <https://doi.org/10.1016/j.neuropsychologia.2018.03.039>
- 782 97. McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of
783 verbal self-monitoring. *Brain*, *119*(3), 907-917. <https://doi.org/10.1093/brain/119.3.907>
- 784 98. Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control.
785 *Neural networks*, *9*(8), 1265-1279. [https://doi.org/10.1016/S0893-6080\(96\)00035-4](https://doi.org/10.1016/S0893-6080(96)00035-4)
- 786 99. Moerel, M., De Martino, F., & Formisano, E. (2012). Processing of natural sounds in human
787 auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *Journal of*
788 *Neuroscience*, *32*(41), 14205-14216. <https://doi.org/10.1523/JNEUROSCI.1388-12.2012>
- 789 100. Mullennix, J. W., Ross, A., Smith, C., Kuykendall, K., Conard, J., & Barb, S. (2011).
790 Typicality effects on memory for voice: Implications for eyewitness testimony. *Applied*
791 *Cognitive Psychology*, *25*(1), 29-34. <https://doi.org/10.1002/acp.1635>
- 792 101. Müller-Preuss, P., & Ploog, D. (1981). Inhibition of auditory cortical neurons during
793 phonation. *Brain research*, *215*(1-2), 61-76. [https://doi.org/10.1016/0006-8993\(81\)90491-1](https://doi.org/10.1016/0006-8993(81)90491-1)
- 794 102. Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., ... &
795 Kojima, S. (2001). Neural substrates for recognition of familiar voices: a PET study.
796 *Neuropsychologia*, *39*(10), 1047-1054. [https://doi.org/10.1016/S0028-3932\(01\)00037-9](https://doi.org/10.1016/S0028-3932(01)00037-9)
- 797 103. Nolan, F., McDougall, K., & Hudson, T. (2011). Some acoustic correlates of perceived
798 (dis)similarity between same-accent voices. *ICPhS (17)*, 1506-1509. Retrieved from
799 <http://icphs2011.hk.lt.cityu.edu.hk/resources/OnlineProceedings/RegularSession/Nolan/Nolan.pdf>
800 [n.pdf](#)

- 801 104. Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity
802 of the human auditory cortex. *Neuroscience Letters*, 265(2), 119-122.
803 [https://doi.org/10.1016/S0304-3940\(99\)00218-9](https://doi.org/10.1016/S0304-3940(99)00218-9)
- 804 105. Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P.
805 E., ... & Belin, P. (2015). The human voice areas: Spatial organization and inter-individual
806 variability in temporal and extra-temporal cortices. *Neuroimage*, 119, 164-174.
807 <https://doi.org/10.1016/j.neuroimage.2015.06.050>
- 808 106. Petkov, C. I., & Vuong, Q. C. (2013). Neuronal Coding: The Value in Having an
809 Average Voice. *Current Biology*, 23(12), R521-R523.
810 <https://doi.org/10.1016/j.cub.2013.04.077>
- 811 107. Pinheiro, A. P., Rezaii, N., Rauber, A., & Niznikiewicz, M. (2016a). Is this my voice or
812 yours? The role of emotion and acoustic quality in self-other voice discrimination in
813 schizophrenia. *Cognitive Neuropsychiatry*, 21(4), 335-353.
814 <https://doi.org/10.1080/13546805.2016.1208611>
- 815 108. Pinheiro, A. P., Rezaii, N., Nestor, P. G., Rauber, A., Spencer, K. M., & Niznikiewicz,
816 M. (2016b). Did you or I say pretty, rude or brief? An ERP study of the effects of speaker's
817 identity on emotional word processing. *Brain and Language*, 153, 38-49.
818 <https://doi.org/10.1016/j.bandl.2015.12.003>
- 819 109. Pinheiro, A. P., Schwartze, M., & Kotz, S. A. (2018). Voice-selective prediction
820 alterations in nonclinical voice hearers. *Scientific Reports*, 8(1), 1-10.
821 <https://doi.org/10.1038/s41598-018-32614-9>

- 822 110. Pinheiro, A. P., Farinha-Fernandes, A., Roberto, M. S., & Kotz, S. A. (2019). Self-voice
823 perception and its relationship with hallucination predisposition. *Cognitive Neuropsychiatry*,
824 24(4), 237-255. <https://doi.org/10.1080/13546805.2019.1621159>
- 825 111. Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., ... &
826 Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, 72(4),
827 665-678. <https://doi.org/10.1016/j.neuron.2011.09.006>
- 828 112. Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex:
829 nonhuman primates illuminate human speech processing. *Nature Neuroscience*, 12(6), 718-
830 724. <https://doi.org/10.1038/nn.2331>
- 831 113. Rauschecker, J. P. (2011). An expanded role for the dorsal auditory pathway in
832 sensorimotor control and integration. *Hearing research*, 271(1-2), 16-25.
833 <https://doi.org/10.1016/j.heares.2010.09.001>
- 834 114. Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite
835 hemispheric lateralization effects during speaking and singing at motor cortex, insula and
836 cerebellum. *Neuroreport*, 11(9), 1997-2000. [https://doi.org/10.1097/00001756-200006260-](https://doi.org/10.1097/00001756-200006260-00038)
837 [00038](https://doi.org/10.1097/00001756-200006260-00038)
- 838 115. Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of
839 Hemispheric Specialization of Self-Voice Recognition. *Brain and Cognition*, 68(2), 204-214.
840 <https://doi.org/10.1016/j.bandc.2008.04.007>
- 841 116. Roswadowitz, C., Mathias, S. R., Hintz, F., Kreitewolf, J., Schelinski, S., & von
842 Kriegstein, K. (2014). Two cases of selective developmental voice-recognition impairments.
843 *Current Biology*, 24(19), 2348-2353. <https://doi.org/10.1016/j.cub.2014.08.048>

- 844 117. Sakai, N., Masuda, S., Shimotomai, T., & Mori, K. (2009). Brain activation in adults who
845 stutter under delayed auditory feedback: An fMRI study. *International Journal of Speech-*
846 *Language Pathology*, 11(1), 2-11. <https://doi.org/10.1080/17549500802588161>
- 847 118. Sapara, A., Cooke, M. A., Williams, S. C., & Kumari, V. (2015). Is it me? Verbal self-
848 monitoring neural network and clinical insight in schizophrenia. *Psychiatry Research:*
849 *Neuroimaging*, 234(3), 328-335. <https://doi.org/10.1016/j.psychresns.2015.10.007>
- 850 119. Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms
851 mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10(1), 24-30.
852 <https://doi.org/10.1016/j.tics.2005.11.009>
- 853 120. Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in
854 autism spectrum disorder but are dysfunctional for voice identity recognition. *Social*
855 *Cognitive and Affective Neuroscience*, 11(11), 1812-1822.
856 <https://doi.org/10.1093/scan/nsw089>
- 857 121. Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices:
858 Influence of stimulus duration and different types of retrieval cues. *Journal of Speech,*
859 *Language, and Hearing Research*, 40(2), 453-463. <https://doi.org/10.1044/jslhr.4002.453>
- 860 122. Schweinberger, S. R., Walther, C., Zäske, R., & Kovács, G. (2011). Neural correlates of
861 adaptation to voice identity. *British Journal of Psychology*, 102(4), 748-764.
862 <https://doi.org/10.1111/j.2044-8295.2011.02048.x>
- 863 123. Schweinberger, S. R., Kawahara, H., Simpson, A. P., Skuk, V. G., & Zäske, R. (2014).
864 Speaker perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(1), 15-25.
865 <https://doi.org/10.1002/wcs.1261>

- 866 124. Seeley, W. W. (2010). Anterior insula degeneration in frontotemporal dementia. *Brain*
867 *Structure and Function*, 214(5-6), 465-475. <https://doi.org/10.1007/s00429-010-0263-z>
- 868 125. Siman-Tov, T., Granot, R. Y., Shany, O., Singer, N., Hendler, T., & Gordon, C. R.
869 (2019). Is there a prediction network? Meta-analytic evidence for a cortical-subcortical
870 network likely subserving prediction. *Neuroscience & Biobehavioral Reviews*, 105, 262-275.
871 <https://doi.org/10.1016/j.neubiorev.2019.08.012>
- 872 126. Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2015). afex: Analysis of factorial
873 experiments. R package version 0.13–145. Retrieved from [https://cran.r-](https://cran.r-project.org/web/packages/afex/index.html)
874 [project.org/web/packages/afex/index.html](https://cran.r-project.org/web/packages/afex/index.html)
- 875 127. Sitek, K. R., Mathalon, D. H., Roach, B. J., Houde, J. F., Niziolek, C. A., & Ford, J. M.
876 (2013). Auditory cortex processes variation in our own speech. *PloS one*, 8(12), e82925.
877 <https://doi.org/10.1371/journal.pone.0082925>
- 878 128. Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E.,
879 Johansen-Berg, H., ... & Niazy, R. K. (2004). Advances in functional and structural MR
880 image analysis and implementation as FSL. *Neuroimage*, 23, S208-S219.
881 <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- 882 129. Smith, D. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-
883 tract length in judgements of speaker size, sex, and age. *The Journal of the Acoustical Society*
884 *of America*, 118(5), 3177-3186. <https://doi.org/10.1121/1.2047107>
- 885 130. Smith, D. R., Walters, T. C., & Patterson, R. D. (2007). Discrimination of speaker sex
886 and size when glottal-pulse rate and vocal-tract length are controlled. *The Journal of the*
887 *Acoustical Society of America*, 122(6), 3628-3639. <https://doi.org/10.1121/1.2799507>

- 888 131. Sommer, I. E., Diederer, K. M., Blom, J. D., Willems, A., Kushan, L., Slotema, K., ... &
889 Kahn, R. S. (2008). Auditory verbal hallucinations predominantly activate the right inferior
890 frontal area. *Brain*, *131*(12), 3169-3177. <https://doi.org/10.1093/brain/awn251>
- 891 132. Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying
892 auditory feedback control of speech. *Neuroimage*, *39*(3), 1429-1443.
893 <https://doi.org/10.1016/j.neuroimage.2007.09.054>
- 894 133. Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech
895 acquisition and production. *Language and Cognitive Processes*, *26*(7), 952-981.
896 <https://doi.org/10.1080/01690960903498424>
- 897 134. Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., &
898 Kuriki, S. (2007). Neural correlates of auditory feedback control in human. *Neuroscience*,
899 *146*(2), 499-503. <https://doi.org/10.1016/j.neuroscience.2007.02.023>
- 900 135. Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The
901 neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580-
902 591. <https://doi.org/10.1162/jocn.2008.20054>
- 903 136. Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in
904 patients with hemispheric damage. *Brain and Cognition*, *1*(2), 185-195.
905 [https://doi.org/10.1016/0278-2626\(82\)90016-1](https://doi.org/10.1016/0278-2626(82)90016-1)
- 906 137. Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are
907 separate abilities. *Neuropsychologia*, *25*(5), 829-834. [https://doi.org/10.1016/0028-](https://doi.org/10.1016/0028-3932(87)90120-5)
908 [3932\(87\)90120-5](https://doi.org/10.1016/0028-3932(87)90120-5)

- 909 138. Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988).
910 Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex*, 24(2), 195-209.
911 [https://doi.org/10.1016/S0010-9452\(88\)80029-7](https://doi.org/10.1016/S0010-9452(88)80029-7)
- 912 139. Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates
913 speaking induced suppression in auditory cortex. *BMC Neuroscience*, 10(1), 58.
914 <https://doi.org/10.1186/1471-2202-10-58>
- 915 140. Versmissen, D., Janssen, I., Johns, L., McGUIRE, P. H. I. L. I. P., Drukker, M., Campo,
916 J. A., ... & Krabbendam, L. (2007). Verbal self-monitoring in psychosis: a non-replication.
917 *Psychological Medicine*, 37(4), 569-576. <https://doi.org/10.1017/S0033291706009780>
- 918 141. von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of
919 neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain*
920 *Research*, 17(1), 48-55. [https://doi.org/10.1016/S0926-6410\(03\)00079-X](https://doi.org/10.1016/S0926-6410(03)00079-X)
- 921 142. Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right
922 superior temporal sulcus for the processing of voices. *Neuroimage*, 22(2), 948-955.
923 <https://doi.org/10.1016/j.neuroimage.2004.02.020>
- 924 143. Wang, J., Mathalon, D. H., Roach, B. J., Reilly, J., Keedy, S. K., Sweeney, J. A., & Ford,
925 J. M. (2014). Action planning and predictive coding when speaking. *Neuroimage*, 91, 91-98.
926 <https://doi.org/10.1016/j.neuroimage.2014.01.003>
- 927 144. Waters, F., Allen, P., Aleman, A., Fernyhough, C., Woodward, T. S., Badcock, J. C., ... &
928 Vercammen, A. (2012). Auditory hallucinations in schizophrenia and nonschizophrenia
929 populations: a review and integrated model of cognitive mechanisms. *Schizophrenia Bulletin*,
930 38(4), 683-693. <https://doi.org/10.1093/schbul/sbs045>

- 931 145. Watkins, K., Patel, N., Davis, S., & Howell, P. (2005). Brain activity during altered
932 auditory feedback: an fMRI study in healthy adolescents. *Neuroimage*, 26(Supp 1), 304.
933 Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2801072/>
- 934 146. Whitford, T. J., Mathalon, D. H., Shenton, M. E., Roach, B. J., Bammer, R., Adcock, R.
935 A., ... & Schneiderman, J. S. (2011). Electrophysiological and diffusion tensor imaging
936 evidence of delayed corollary discharges in patients with schizophrenia. *Psychological*
937 *Medicine*, 41(5), 959-969. <https://doi.org/10.1017/S0033291710001376>
- 938 147. Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., &
939 Ethofer, T. (2008). Cerebral processing of emotional prosody—influence of acoustic
940 parameters and arousal. *Neuroimage*, 39(2), 885-893.
941 <https://doi.org/10.1016/j.neuroimage.2007.09.028>
- 942 148. Wickham, H. (2017). tidyverse: Easily Install and Load the ‘Tidyverse’. R package
943 version 1.2. 1. Retrieved from <https://cran.r-project.org/web/packages/tidyverse/index.html>
- 944 149. Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive*
945 *Sciences*, 1(6), 209-216. [https://doi.org/10.1016/S1364-6613\(97\)01070-X](https://doi.org/10.1016/S1364-6613(97)01070-X)
- 946 150. Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for
947 motor control. *Neural Networks*, 11(7-8), 1317-1329. [https://doi.org/10.1016/S0893-](https://doi.org/10.1016/S0893-6080(98)00066-5)
948 [6080\(98\)00066-5](https://doi.org/10.1016/S0893-6080(98)00066-5)
- 949 151. Xu, M., Homae, F., Hashimoto, R. I., & Hagiwara, H. (2013). Acoustic cues for the
950 recognition of self-voice and other-voice. *Frontiers in Psychology*, 4, 735.
951 <https://doi.org/10.3389/fpsyg.2013.00735>

- 952 152. Xue, G., Aron, A. R., & Poldrack, R. A. (2008). Common neural substrates for inhibition
953 of spoken and manual responses. *Cerebral Cortex*, 18(8), 1923-1932.
954 <https://doi.org/10.1093/cercor/bhm220>
- 955 153. Zarate, J. M., & Zatorre, R. J. (2005). Neural substrates governing audiovocal integration
956 for vocal pitch regulation in singing. *Annals of the New York Academy of Sciences*, 1060,
957 404-408. <https://doi.org/10.1196/annals.1360.058>
- 958 154. Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between
959 regions involved in speech perception and in monitoring one's own voice during speech
960 production. *Journal of Cognitive Neuroscience*, 22(8), 1770-1781.
961 <https://doi.org/10.1162/jocn.2009.21324>
- 962 155. Zheng, Z. Z., MacDonald, E. N., Munhall, K. G., & Johnsrude, I. S. (2011). Perceiving a
963 stranger's voice as being one's own: A 'rubber voice' illusion?. *PloS One*, 6(4).
964 <https://doi.org/10.1371/journal.pone.0018655>
- 965

966 TABLES

967 **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}	4551
	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}	4565
	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

968

969 Table Caption:

970 Results from TVA localizer task: Coordinates listed in MNI space; (p/a/m)STG:

971 posterior/anterior/middle superior temporal gyrus, preCG: precentral gyrus, IFG: inferior frontal

972 gyrus; 7 peak-level activations in 4 clusters: 1. left STG, 2. right STG, 3. right preCG, 4. left

973 IFG; All listed significant regions survived FDR-corrected threshold 0.05.

974 FIGURE CAPTIONS:

975 **1. Psychometric Voice Attribution Task (VAT):** Active = button-press condition; Passive =
976 hearing conditions, * = affected by individual motor response-time variability; Response = two-
977 alternate forced-choice (“The voice sounded more like me.” or “The voice sounded more like
978 someone else.”).

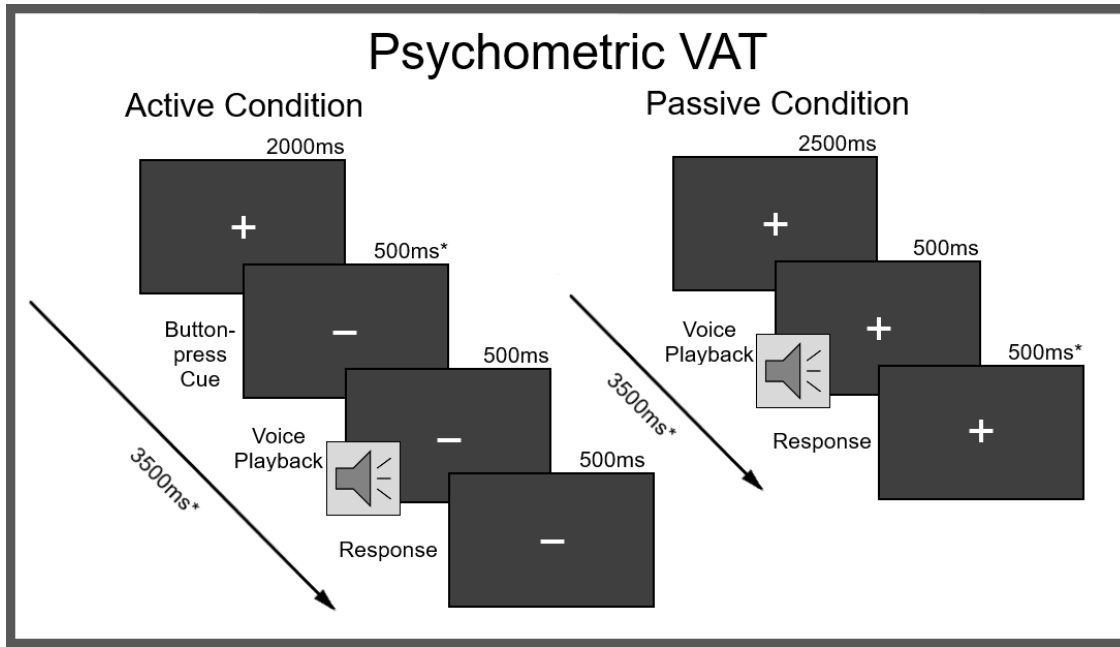
979 **2. fMRI Voice Perception Task (VPT):** Active = button-press condition; Passive = hearing
980 conditions, * = affected by individual motor response-time variability.

981 **3. fMRI Regions of Interest:** Blue: inferior frontal gyrus; MNI coordinates x 58, y 2, z -10;
982 determined from ALE neuroimaging meta-analysis (Johnson et al., 2019). Red: right anterior
983 superior temporal gyrus; MNI coordinates x 46, y 10, z 4; determined from fMRI temporal voice
984 area localizer task.

985 **4. fMRI Voice Perception Task (VPT) LMM Results:** Linear mixed model analysis on ROIs
986 in A) right anterior superior temporal gyrus (aSTG) and B) right inferior frontal gyrus (IFG).
987 Active: button-press condition, Passive: passive hearing condition, SV: self-voice, UV:
988 uncertain-voice, OV: other-voice. Post-hoc analysis in right aSTG revealed motor induced
989 suppression (for contrast Active > Passive) for only SV as compared to UV or OV ($t(119) = -2.7$,
990 $p = 0.021$).

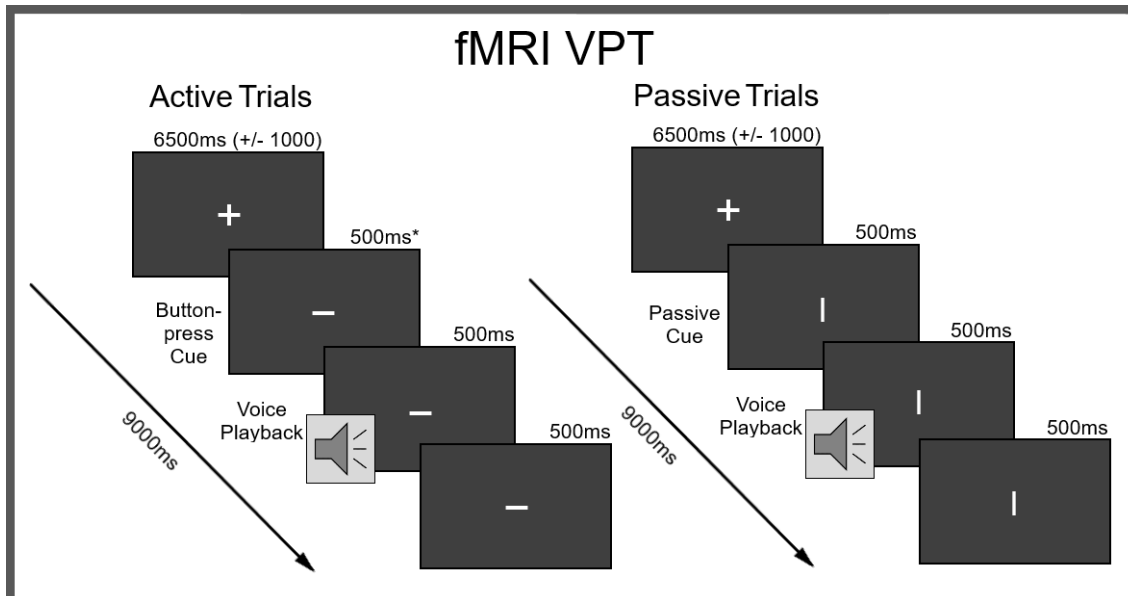
991 FIGURES

992 **1. Psychometric Voice Attribution Task (VAT)**



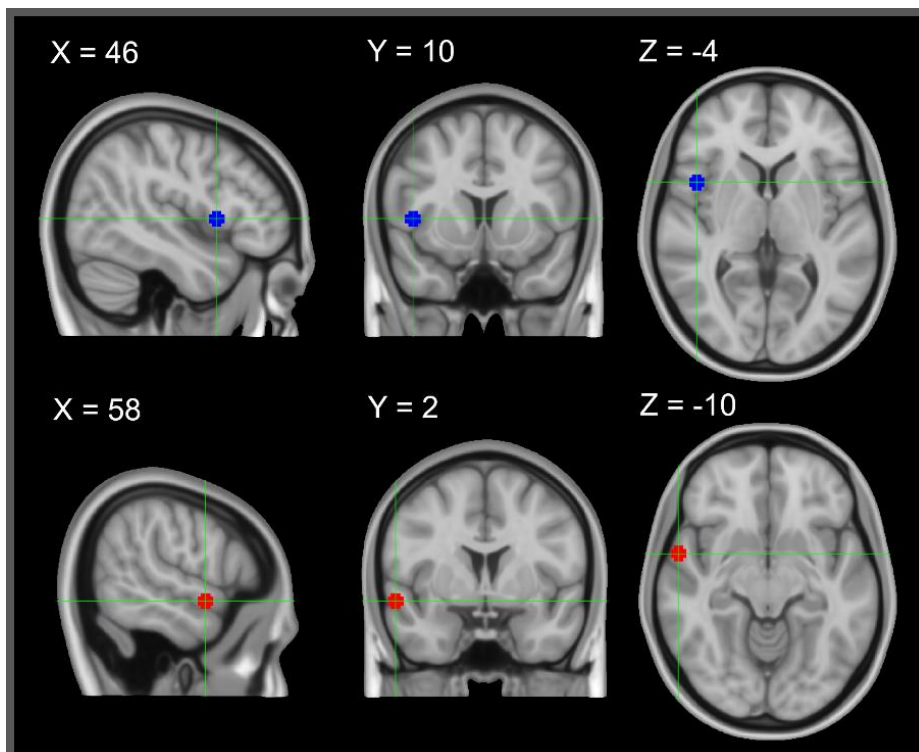
993

994 **2. fMRI Voice Perception Task (VPT):**



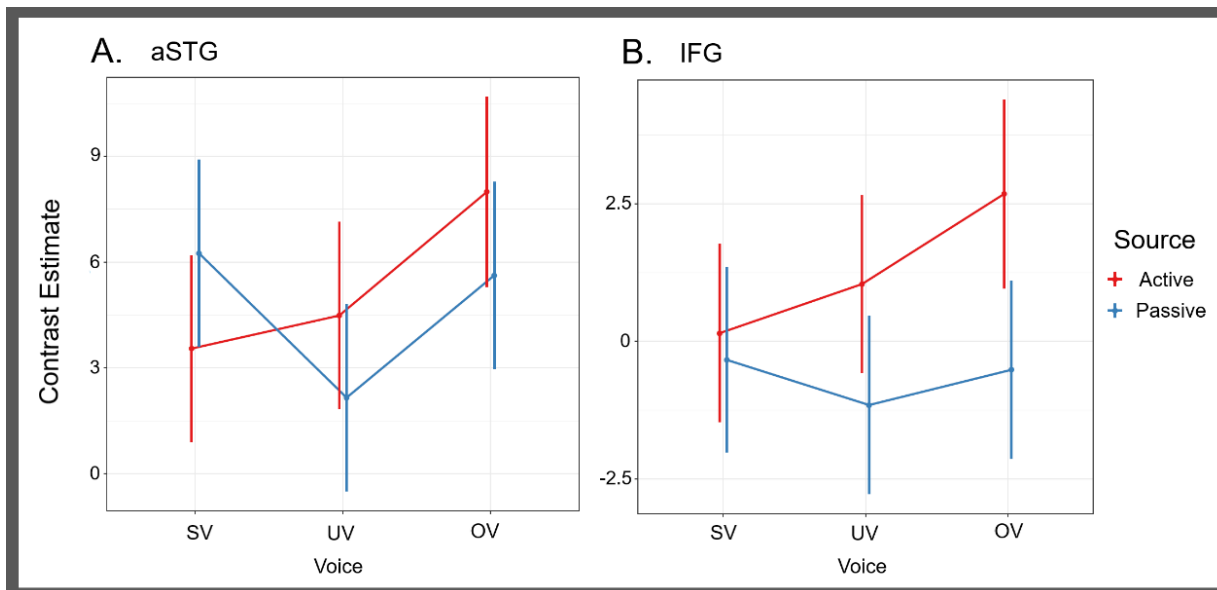
995

996 **3. fMRI Regions of Interest**



997

998 **4. fMRI Voice Perception Task (VPT) LMM Results**



999