Neural competition between concurrent speech production and other speech perception

Journal Article

Author(s):
Dietziker, Joris; Staib, Matthias; Frühholz, Sascha

Publication date:
2021-03

Permanent link:
https://doi.org/10.3929/ethz-b-000460864

Rights / license:
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

Originally published in:
NeuroImage 228, https://doi.org/10.1016/j.neuroimage.2020.117710

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Neural competition between concurrent speech production and other speech perception

Joris Dietziker[a,*], Matthias Staib[a,b], Sascha Frühholz[a,b,c,d,*]

[a] Cognitive and Affective Neuroscience Unit, University of Zurich, Zurich, Switzerland
[b] Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland
[c] Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Switzerland
[d] Department of Psychology, University of Oslo, Norway

ARTICLE INFO

Keywords:
Speech
Voice
Language
fMRI
Auditory cortex

ABSTRACT

Understanding others' speech while individuals simultaneously produce speech utterances implies neural competition and requires specific mechanisms for a neural resolution given that previous studies proposed opposing signal dynamics for both processes in the auditory cortex (AC). We here used neuroimaging in humans to investigate this neural competition by lateralized stimulations with other speech samples and ipsilateral or contralateral lateralized feedback of actively produced self speech utterances in the form of various speech vowels. In experiment 1, we show, first, that others' speech classifications during active self speech lead to activity in the planum temporale (PTe) when both self and other speech samples were presented together to only the left or right ear. The contralateral PTe also seemed to indifferently respond to single self and other speech samples. Second, specific activity in the left anterior superior temporal cortex (STC) was found during dichotic stimulations (i.e. self and other speech presented to separate ears). Unlike previous studies, this left anterior STC activity supported self speech rather than other speech processing. Furthermore, right mid and anterior STC was more involved in other speech processing. These results signify specific mechanisms for self and other speech processing in the left and right STC beyond a more general speech processing in PTe. Third, other speech recognition in the context of listening to recorded self speech in experiment 2 led to largely symmetric activity in STC and additionally in inferior frontal subregions. The latter was previously reported to be generally relevant for other speech perception and classification, but we found frontal activity only when other speech classification was challenged by recorded but not by active self speech samples. Altogether, unlike formerly established brain networks for uncompetitive other speech perception, active self speech during other speech perception seemingly leads to a neural reordering, functional reassignment, and unusual lateralization of AC and frontal brain activations.

1. Introduction

Auditory speech is an important feature of many social interactions, and understanding the speech of others is critical to successful communicative interactions. Accurately understanding the speech of other individuals requires neural processing in a distributed network of brain regions. Given the acoustic nature of auditory speech, the first important brain system to recognize speech is the auditory cortex (AC) located in the superior temporal cortex (STC). At the neural level of the AC, several studies have shown increased left anterior STC (aSTC) activity for understanding others' speech (Evans et al., 2014; Scott et al., 2000), especially when noise levels decreased that masked or degraded the speech. The latter notion refers to a specific line of research that investigated the neural effects of understanding the speech of others during suboptimal listening conditions. Decreasing noise levels allows a better recognition of others' speech based on a lesser degree of noise masking, and the cognition rate correlates roughly with this noise decrease. Besides the aSTC a critical cortical node for speech recognition, recent reports also pointed to the bilateral mid STC (mSTC) and posterior STC (pSTC) for others' speech perception (Evans et al., 2014; Okada et al., 2010; Osnes et al., 2011) using a similar line of research setups.

Humans not only listen to speech samples of other individuals, but they also listen to their own speech while talking. These own speech samples should have the same acoustic properties as speech samples of other individuals, with the only difference being that these speech samples are self-produced by the listener and include some acoustic effects introduced by bone conduction. Given the involvement of the AC/STC in the analysis of others' speech, it should be also involved in the acoustic analysis of self speech generated by the listener. Accurate overt self-generated speech production and self-speech feedback regis-

[*] Corresponding authors.
E-mail addresses: joris.dietziker@uzh.ch (J. Dietziker), sascha.fruehholz@uzh.ch (S. Frühholz).

https://doi.org/10.1016/j.neuroimage.2020.117710
Received 3 October 2020; Received in revised form 28 November 2020; Accepted 19 December 2020
Available online 30 December 2020

tration lead to activity in STC, but with apparent opposite neural effects in terms of lower activations compared to a baseline condition. Specifically, while decreasing the noise level during other speech perception leads generally to higher AC activity, decreasing the noise level during self speech perception leads to lower AC activity. Accordingly, studies on active speech production found decreased activity in the AC that is mainly localized to the bilateral mSTC and pSTC (Behroozmand et al., 2015; Christoffels et al., 2011, 2007), with stronger effects in the right hemisphere (Franken et al., 2018). This suppression effect in the AC during active self speech seems to originate when the difference between intended speech (potentially stored as a vocal motor template in the frontal pre-motor cortex) and actual speech (analyzed by auditory feedback registration) is minimal, such that no vocalizing errors occurred. This suppression effect might thus serve to more easily detect and correct vocalizations errors, and to minimize the interference with following speech productions (Hickok, 2012). Thus suppression effect for AC activity during active self speech, is commonly found to be stronger with decreasing levels of noise during feedback registration of self speech (Hickok, 2012).

Given these previous studies on other speech perception and self speech production, neural effects in the AC accompanying speech perception in others and self speech production thus seem to neurally overlap in the mSTC and pSTC subregions of the AC (Cheung et al., 2016; Evans and Davis, 2015; Rampinini et al., 2017), but seem to elicit contrasting signal dynamics. Given these opposing signal dynamics in the AC, this could cause neural competition and/or neural interaction when others' speech is recognized while self-speech is produced simultaneously. This neural competition also requires mechanisms for a neural resolution of this conflict, which so far are relatively unexplored. As mentioned above, a common line of research into the neural dynamics of speech perception uses noise to degrade and mask speech samples of other individuals. A critical difference in our study is that we introduced a source of noise during others' speech perception that is internal to the listener. Because this "noise" is produced by the listener, it is thus partly more predictable for the listener. Compared to common mechanisms of neural resolution and neural interaction, the case of internally generated noise in the listener could thus potentially cause more non-linear neural effects both of attenuated nature (given the predictability of the noise) and/or of increased nature (less neural suppression to self speech). The latter might be especially the case because introducing some kind of noise signal (other speech) during active self speech diminishes the AC suppression effect to certain degrees, which could interact with AC activity for other speech perception.

Besides neural activity in AC for other speech perception and self speech production, there is also commonly found activity in the inferior frontal cortex (IFC) for both functional domains, given also close structural (Friederici, 2011; Frühholz et al., 2015) and functional connections to the AC (Frühholz and Grandjean, 2012; Hickok, 2012). Specifically, the posterior IFC (inferior frontal gyrus (IFG), pars opercularis) shows increased activity during speech production (Brown et al., 2005; Eickhoff et al., 2009), whereas the more anterior IFC (IFG, pars triangularis) shows increased activity rather during the semantic decoding of speech (Elmer and Kühnis, 2016; Tyler et al., 2005). Additionally, the IFG displays hemispheric differences with the left IFG showing more activity for processing speech-related information, while right IFG activity seems to be predominantly found processing pitch contours in linguistic and paraverbal contexts (Frühholz and Grandjean, 2013a; Geiser et al., 2008; Merrill et al., 2012). Similarly, the ventral premotor cortex (vPMC) is active both during the perception and production of speech (Aziz-Zadeh et al., 2010; Hickok, 2010; Parkinson et al., 2012). While some models propose that vPMC to not be essential for speech perception (Scott et al., 2009), while other findings point to its relevance in the neutrally efficient analysis of speech sounds (Möttönen et al., 2013), especially through sensorimotor integration (Hickok et al., 2011).

Unlike the AC, neural suppression effects in the IFC have not been observed during active self speech production, such that both other speech perception and self speech production lead to increased activity in the IFC. For both domains, however, neural IFC activity seems spatially separated, with some potential overlap in vPMC. It is thus unclear how the IFC and its subregion deal with the simultaneous task of speech perception and speech production, but this specific condition could lead either to a potentiation of the combined neural processing in overlapping frontal regions or to an accentuation of each speech signal in certain IFC subregions. To thus investigate this neuronal interaction of simultaneous self-speech production and other-speech perception in the AC and the IFC, we conducted therefore two fMRI experiments including human participants. More specifically, we investigated the production and perception of simple speech sounds in the form of speech vowels as basic elements of human speech and that could easily be produced in functional neuroimaging environment. We focused on three major questions: (a) Are both processes relatively lateralized in the AC when performed simultaneously? (b) Do they overlap, compete, and interact (non-)linearly for neural resources at specific subregions of the STC (Assaneoet al., 2019)? and (c) What is the additional role of the ventral (pre)motor cortex (vPMC) (Behroozmand et al., 2015; Evans and Davis, 2015; Markiewicz and Bohland, 2016; Meister et al., 2007; Stokes et al., 2019; Wilson et al., 2004) and the IFC (Cogan et al., 2014) in simultaneous self speech production (i.e. speech motor planning) and other speech perception (i.e. motor mirroring, vowel identification), given the strong and partly differential involvement in both processes?

2. Materials and methods

2.1. Participants

Thirty-one healthy and right-handed adults with normal or corrected-to-normal vision, normal hearing, and no reported history of neurological or psychiatric disorder took part in the fMRI experiments. Three participants were excluded because they had excessive motion artifacts, resulting in a final sample of 28 participants (13 females; mean age 25.89y, SD=3.81). All participants gave written informed consent and were reimbursed for their participation. The study was approved by the Swiss governmental ethics committee of the Cantone Zürich.

2.2. Stimuli

The general purpose of the experiments was to investigate the ability of participants to identify the speech of others (i.e. other speech, "OTHER") while producing self-speech at the same time (i.e. speaker's self-speech, "SELF"). Both OTHER and SELF consisted of the vocally expressed vowels /a/, /e/, /i/, and /o/. For OTHER, we used prerecorded vocalizations of these vowels by one male and one female speaker, thus consisting of 8 different recordings. These recordings were 24-bit vowel vocalizations of 300 ms duration with a sampling rate of 44.1 kHz. OTHER were presented at an intensity level of 70 dB SPL.

2.3. Experimental procedure of experiment 1

The experiment consisted of two main experiments separated into different sessions. In experiment 1 ("active SELF speech task"), the participants were asked to produce SELF while listening to and classifying OTHER. For the identification and classification of OTHER (i.e. participants classified OTHER as /a/, /e/, /i/, or /o/) we used a 4-alternative forced-choice task using the index and middle finger of both hands, and with counterbalanced response option assignment across participants. For SELF, participants were asked to produce 1 of the 4 vowels with the restriction that the OTHER and the SELF were never the same on a single trial. For each trial, an uppercase letter on the screen ("A", "E", "I", or "O") cued the participants to produce the respective vowel for about 1 s duration on an MR-compatible microphone (OptoActive system, Optoacoustics). The cue remained on the screen for 2 s. If no SELF

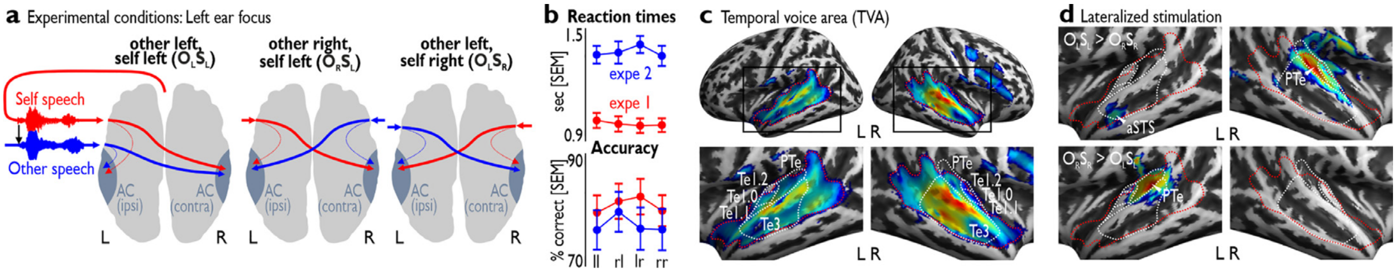
a Experimental conditions: Left ear focus
other left, self left (O_LS_L)
other right, self left (O_RS_L)
other left, self right (O_LS_R)
Self speech
Other speech
AC (ipsi)
AC (contra)
L
R
b Reaction times
1.5
0.9
sec [SEM]
expe 2
expe 1
Accuracy
90
70
% correct [SEM]
ll
rl
lr
rr
c Temporal voice area (TVA)
L R
PTe
Te1.2
Te1.0
Te1.1
Te3
d Lateralized stimulation
$O_LS_L > O_RS_R$
$O_RS_R > O_LS_L$
aSTS
PTe

Fig. 1. Behavioral and functional brain data for experiment 1. (a) The experimental conditions required participants to produce vowels by themselves (SELF, self-speech, red lines) while simultaneously listening to vowel recordings of other individuals (OTHER, other speech, blue lines). Own and other speech was presented to the same ear or to different ears. **(b)** Reaction time (upper panel) and accuracy data (lower panel) for the decision on OTHER for the 4 conditions in experiment 1 (expe 1, red) and experiment 2 (expe 2, blue). **(c)** Functional definition of the voice-sensitive area (TVA; red dashed line). **(d)** Presenting all speech to the left ear ([O_LS_L > O_RS_R], upper panel) or to the right ear ([O_RS_R > O_LS_L], lower panel) leads to increased activation in the AC in the contralateral brain centered on the PTe. Brain activity (n = 28) thresholded at a combined voxel (p<0.005) and cluster level (k = 42) threshold (p<0.05 corrected at the cluster level). White dotted lines in panel c-d mark the anatomical subregions of the AC: primary AC (Te1.0–1.2), secondary AC (PTe), and higher-level AC (Te3). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was produced by the participant on one trial, no OTHER was presented (see below), which happened in 3.71% of trials. These trials were excluded from further analyses. The SELFs were sampled and recorded at a sampling rate of 44.1 kHz with 24-bit encoding. We decided to include vowels instead of syllables or longer speech utterances as speech sounds in our experiments, given certain requirements for the task. We aimed to include speech sounds of short duration and a similar level of "complexity" that could be easily identified by participants in the 4-alternative-forced choice task in the fMRI environment. Also, we wanted that self and other speech could be rather easily separated since we did not want to challenge the acoustic confusion too much between both speech samples. Finally, vowels were chosen as speech sounds, such they could be accurately identified by the voice onset detector and to accurately trigger the presentation of OTHER speech and ensure a proper overlap between SELF and OTHER speech samples (see below).

A voice-onset detector indicated the onset of SELF by the participants and instantaneously prompted the presentation of OTHER. This procedure ensured that OTHER and SELF were always presented at the time of voice onset of SELF. SELF was played back to the participants at approximately 70 dB SPL, based on the (de-)amplification of the SELF speech sample, which was estimated from each participant's vocalization intensity during a pre-experimental training run (see below). SELF speech was passed through a filter, which increased frequencies below 1 kHz by +2 dB and frequencies above 1 kHz by −2 dB. This filtering was done to simulate the effects of bone conduction during self speech feedback processing, which enhances lower and attenuates higher voice frequency components (Dauman, 2013).

SELF and the OTHER were presented by using MR-compatible headphones (OptoActive II^{TM} ANC Headphones). These headphones included active noise cancelation of scanner noise of ~20 dB. They were randomly presented to the left and/or right ear with online mixing of OTHER and SELF with a frame size of 64 samples based on a 44.1 kHz sampling rate, resulting in 4 different stimulation conditions: condition O_LS_L with both sounds on the left ear or O_RS_R on the right ear; condition O_LS_R with OTHER on the left and SELF on the right ear, or condition O_RS_L with SELF on the left and OTHER on the right ear (Fig 1a). This lateralized presentation allowed us to investigate laterality effects, assuming that auditory signals elicit stronger contra- than ipsilateral AC activity, and to disentangle the neural overlap of speech production and perception mechanisms.

Experiment 1 consisted of 2 experimental runs (see below) preceded by 3 training runs to familiarize the participants with the experiment. In the first training run, a pseudorandom sequence of 32 OTHERs was first presented. Each OTHER had to be identified by the participants by pressing one of the corresponding keys on a keyboard with the index and middle fingers of both hands. This training was repeated until a 95% accuracy rate was reached. This response button assignment remained the same throughout the experiment for each participant but was counterbalanced across participants. During the second training run, participants were asked to vocalize each of the 4 vowels 4 times. A pseudorandom sequence of 16 vowels was visually presented to the participants, who were given an uppercase cue as described earlier. When a vowel instruction was displayed, the participants had to vocalize it for 1 s. The 16 SELFs were recorded and the mean vocalization intensity (i.e. mean root mean square) was calculated and used for amplifying or de-amplifying the level of the SELF played back to participants to match the intensity of the OTHER at about 70 dB SPL. This calculation of the (de-)amplification level was used throughout all runs of the main experiment and for the third training run. The latter consisted of a pseudorandom sequence of 16 cued vowels for SELF while participants had to both vocalize the SELF (as in training run 2) and to listen and identify the OTHER by using the 4 response buttons (as in training run 1). The identification of OTHER was also the task during the main experiment, using a 4-alternative forced-choice task using the index and middle finger of both hands, and with counterbalanced response option assignment across participants.

After the 3 training runs, participants were asked to take part in the 2 runs of the main experiment. Participants accomplished 96 trials in each run, including active SELF during the 4 experimental conditions as described earlier, thus including 24 trials for each condition with conditions and vowels equally distributed across the trials. The trial order was random, with the exceptions of more than 3 times the same SELF, OTHER, or laterality condition in a row.

2.4. Experimental procedure of experiment 2

In addition to the active speaking task in experiment 1, we ran a second experiment 2 in a separate scanning session by using only a passive listening to SELF and OTHER ("passive SELF speech task"), but including the same classification task on the OTHER as in experiment 1. These 2 passive runs were included to be used as a baseline comparison condition. Since the major hypothesis of the study concerned the influence of active SELF on the perception and identification of OTHER, we aimed to include another experimental condition in which participants were presented with the same setup of listening to SELF and to OTHER, but without an active vocalization condition. For this purpose, SELF recordings from the active conditions in experiment 1 were played back to the participants while they listened both to their SELF recordings and to the OTHER at the same time. The SELF was randomly combined with an OTHER with the restriction that the vowel of the SELF and the OTHER were not the same. The participants' task was to identify the OTHER by using the same 4 buttons as during experiment 1. The participants accomplished 2 runs with 96 trials each and with the same trial order restrictions as during experiment 1.

2.5. Functional voice localizer scan

To identify human voice-sensitive regions in the bilateral superior cortex, we used sound clips of 8 s length from an existing database (Belin et al., 2000). These sound files consisted of 20 vocal sounds and 20 non-vocal sounds. Participants were instructed to listen passively to the stimuli. The functional voice localizer scan was used to determine voice-sensitive regions along the STC in both hemispheres that are commonly referred to as temporal voice area (TVA; Fig. 1c). Using the functional definition of the TVA in both left and right AC, we restricted the discussion of activation pattern found in experiment 1 and 2 to activity that was located inside the TVA to ensure that we only discuss AC activity that is related to voice processing as the carrier of our speech samples.

2.6. Behavioral data analysis

Each experiment included 4 different conditions represented by the within-subject factors laterality of SELF (2 levels: left, right) and laterality of OTHER (2 levels: left, right). Across these 4 different conditions, we quantified the reaction times and the accuracy level of OTHER identifications for each participant. RTs and accuracy were considered only for trials in which the participants produced the correct vocalizations according to the cued instruction. Trials during the active runs with incorrect target SELF speech were included in the passive runs, but these trials were overall discarded from further analyses for both experiments. The RTs and accuracy level of each participant was subjected to a random-effects group analysis of a 2 × 2 repeated measures ANOVA, including the within-subject factors mentioned earlier. The significance threshold was set at α=0.05.

2.7. Functional image acquisition

All structural and functional MRI images were acquired on a 3T Philips Ingenia System. The structural images were obtained by using a high-resolution magnetization prepared rapid acquisition gradient-echo T1-weighted sequence (301 contiguous 1.2 mm slices, TR/TE=1.96 s/3.71 ms, FOV=250 mm, in-plane resolution 1 × 1 mm) obtained in sagittal orientation. Functional brain data were recorded by using 31 axial slices covering the whole brain aligned to the anterior (AC) or to the posterior (PC) commissure plane (thickness/gap=3.5/0.4 mm, FOV=219 mm, in-plane resolution 1.71 × 1.71 mm). A sparse temporal acquisition protocol was used with TR=3.29 s, which consisted of TA=1.71 s for volume acquisition and 1.58 s of a silent gap. For the functional voice localizer scan (see below) we used a continuous whole-head acquisition with 31 slices (thickness/gap=3.5/0.4 mm, FOV=219 mm, in-plane 1.71 × 1.71 mm) aligned to the AC-to-PC plane with a TR/TE=1700/30 ms.

2.8. Functional image analysis

Statistical parametric mapping software (SPM12, fil.ion.ucl.ac.uk/spm/software/spm12) was used for the preprocessing and analysis of the functional brain data. Functional images were realigned and coregistered to the anatomical image. The realigned functional images were spatially normalized to the Montreal Neurological Institute (MNI) stereotactic template brain by using the segmentation procedure implemented in the Computational Anatomy Toolbox (CAT12; neuro.uni-jena.de/cat/). Normalized images were spatially smoothed by using an isotropic Gaussian kernel with a full-width half-maximum of 8 mm.

A general linear model was used for the first-level statistical analyses, including boxcar functions defined by the onset and duration of the auditory stimuli. These boxcar functions were convolved with a canonical hemodynamic response function. We created separate regressors for each of the 4 experimental conditions. The model included 6 additional regressors of no interest that were based on motion estimates to account for motion artifacts.

For the main experiment, contrast images for each experimental condition were entered into a second-level group analysis. All contrasts between conditions were thresholded at a voxel threshold of p<0.005 and a minimum cluster size of k = 42. This combined voxel and cluster threshold corresponds to p = 0.05 corrected at the cluster level and was determined by the 3DClustSim algorithm implemented in AFNI software (afni.nimh.nih.gov/afni; version AFNI_18.3.01), including the recent extension to estimate the (spatial) autocorrelation function according to the median estimated smoothness of the residual images. For the voice localizer, we contrasted vocal against non-vocal trials on the group level by using a threshold of p<0.005 (uncorrected) and a cluster extent of k = 42 voxels (see earlier). We determined voice-sensitive regions along the STC in both hemispheres.

3. Results

3.1. Experiment 1: active production of self speech during concurrent other speech perception

In experiment 1, participants produced self-speech utterances (SELF) that instantaneously triggered the presentation of non-identical recordings of others' speech signals from unfamiliar speakers (OTHER). Participants had to identify the vowel spoken by OTHER, which revealed no difference ($F_{1,27}$<2.321, p>0.139; all rmANOVA unless stated otherwise, n = 28) and no interaction effect ($F_{1,27}$=0.956, p = 0.337, n = 28) in reaction times across the 2 × 2 conditions (SELF left or right, OTHER left or right) and no difference in the accuracy level ($F_{1,27}$<0.460, p>0.503, n = 28), but did reveal an interaction effect ($F_{1,27}$=6.354, p = 0.018, n = 28). Trials of O_LS_R had significantly better performance compared with those of O_LS_L (p = 0.019), such that displaying SELF to the right ear led to better performance in classifying OTHER speech from the left ear compared with presenting SELF with OTHER together to the left ear (Fig 1b). We performed the same analysis for reaction times quantified from SELF speech *offset* instead of SELF speech *onset* (Fig. S1), which resulted in the same patterns of behavioral data.

To investigate the neural effects of concurrent SELF and OTHER processing, we first determined the voice-sensitive areas (VA) in the AC by using a separate functional localizer scan (Belin et al., 2000; Pernet et al., 2015). This functional voice localizer scan revealed spatially extended activity in the bilateral AC covering areas of primary (Te1.0–1.2), secondary (planum temporale [PTe]), and higher-order AC (peaks located in Te3) (Fig 1c). The following analyses of AC activity for the main experiments were accordingly limited to activations within the VA. For the main experiment 1, we first determined neural effects from unilateral stimulations of both SELF and OTHER according to the experimental conditions, either to the left [$O_LS_L > O_RS_R$] or to the right ear [$O_RS_R > O_LS_L$]. Given that we presented both speech samples ipsilateral to one ear, we accordingly found contralateral activity mainly in the secondary AC (PTe) for each contrast. One activation peak stood out from this overall contralateral and symmetric activation pattern, as for [$O_LS_L > O_RS_R$] we found additional peak activations in the ipsilateral left aSTS when presenting both speech samples to the left ear (Fig. 1d).

In the next step, we compared the conditions while presenting SELF and OTHER to separate ears and brain hemispheres to assess potential lateralized neural indications for SELF suppression and OTHER positive cortical effects. This revealed no activations for comparing [$O_LS_R > O_RS_L$] or [$O_RS_L > O_LS_R$]; these contrasts directly compared if presenting SELF and OTHER to separate ears would elicit similar or different cortical effects of neural activations or neural suppression of both types of speech samples. It seemed like both SELF and OTHER elicited similar levels of neural activations with no obvious suppression effects for SELF.

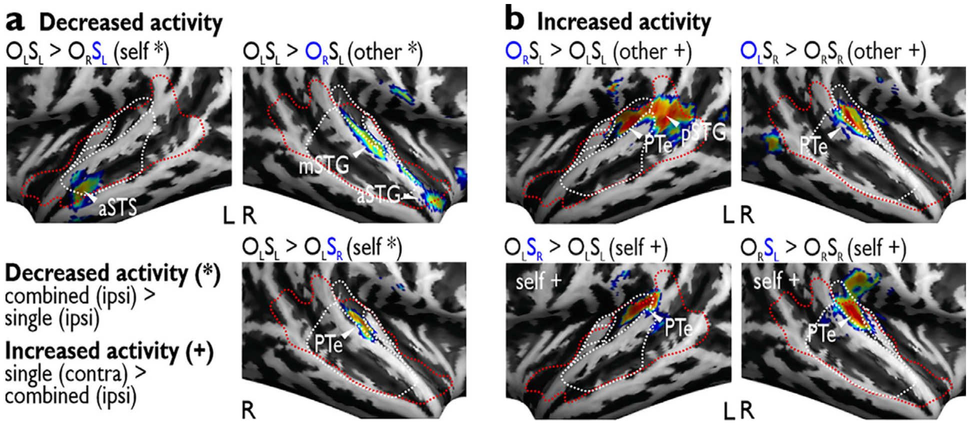
a Decreased activity
$O_LS_L > O_RS_L$ (self *)
$O_LS_L > O_RS_L$ (other *)
aSTS
mSTG
aSTG
L R
$O_LS_L > O_LS_R$ (self *)
PTe
R
Decreased activity (*)
combined (ipsi) > single (ipsi)
Increased activity (+)
single (contra) > combined (ipsi)
b Increased activity
$O_RS_L > O_LS_L$ (other +)
$O_LS_R > O_RS_R$ (other +)
PTe
pSTG
PTe
L R
$O_LS_R > O_LS_L$ (self +)
$O_RS_L > O_RS_R$ (self +)
self +
self +
PTe
PTe
L R

Fig. 2. Functional brain data for experiment 1. (a) Decreased activation found in the bilateral AC for SELF* and in the right AC for OTHER*. **(b)** Increased activation found in contralateral AC both for SELF+ and OTHER+. Voxel p<0.005, cluster k = 42 (p<0.05 corrected at the cluster level), n = 28. White dotted lines mark the anatomical subregions of the AC: primary AC (Te1.0–1.2), secondary AC (PTe), and higher-level AC (Te3).

In the next step, we accordingly looked at functional activity resulting from a combined presentation of SELF and OTHER to either the left (O_LS_L) or the right ear (O_RS_R) compared with the bilateral conditions of presenting both speech samples (O_RS_L or O_LS_R). We quantified two types of activation profiles. First, we quantified activity that we refer to as "decreased activity". For example, we performed the contrast [$O_LS_L > \mathbf{O_R}S_L$] that compared the presentation of both samples to the left ear with the presentation of only SELF to the left ear (SELF left). Given that acoustic signals are predominantly processed in the contralateral AC, we assumed that if this contrast leads to significant contralateral activity in the same ear condition (O_LS_L), it should be the missing condition on the left ear (OTHER left) that is responsible for the lower activity for the bilateral presentation (O_RS_L). This is why we termed this activity as "decreased activity" and marked the responsible condition for the reduced activity with a star (OTHER*) (Fig. 2a). Second, we quantified "increased activity" by performing a different set of contrasts. For example, we performed the contrast [$O_L\mathbf{S_R} > O_LS_L$] and found left brain activity, and marked it with a plus (SELF+), since SELF was the single right ear condition contralateral to the combined left ear condition (Fig. 2b). Regarding the latter increased activity, we found consistent increased and symmetrical activity in the secondary AC (PTe), both for SELF+ (based on [$O_LS_R > O_LS_L$] and [$O_RS_L > O_RS_R$]) and OTHER+ (based on [$O_RS_L > O_LS_L$] and [$O_LS_R > O_RS_R$]), with additional activity in the pSTG for OTHER+ for the contrast [$O_RS_L > O_LS_L$].

Besides these large symmetrical effects of SELF and OTHER in the ipsilateral secondary AC, we found decreased and rather asymmetric activations that were very specific for comparisons with the O_LS_L condition, thus when all speech samples were presented to the left ear and predominantly processed in the right AC (Fig. 2a). In the right AC, we found peak locations for OTHER* ([$O_LS_L > O_RS_L$]) in the right mSTC and aSTC and for SELF* ([$O_LS_L > O_LS_R$]) in right PTe. Furthermore, we found activity in the left aSTC for SELF* ([$O_LS_L > O_RS_L$]) resembling the activity found for the unilateral stimulation with O_LS_L (Fig. 1d). This aSTC activity for SELF* did not follow our general definition of decreased activity, because it appeared ipsilateral to the combined left ear condition (O_LS_L) instead of contralaterally to it. But given that O_LS_L can induce strong ipsilateral activity in left aSTC (see above; Fig. 1d), and for O_RS_L the missing right ear condition was SELF, we labeled the left aSTC as negative activity originating from SELF*.

3.2. Experiment 2: listening to recorded self speech during concurrent other speech perception

To test whether these neural mechanisms reflected by increased and decreased activations in experiment 1 are specific to the active self-speech condition, we conducted a similar experiment (experiment 2), the only difference being that participants were presented passively with recordings of their own SELF speech samples extracted from experiment 1 instead of producing them actively. Participants again had to identify OTHER speech, leading to no differences in reaction times (all $F_{1,27}$<2.817, p>0.105, n = 28) and in the accuracy level (all $F_{1,27}$<2.832, p>0.104, n = 28) according to the main factors and for interaction effects (Fig. 1b).

In terms of neural activity, the comparison of unilateral stimulations ([$O_LS_L > O_RS_R$] and [$O_RS_R > O_LS_L$]) revealed results comparable to those for experiment 1 in PTe, but missing the left aSTS activation for SELF* (Fig. 2a). Furthermore, unlike in experiment 1, we found only increased but no decreased activity for comparing unilateral with bilateral conditions (Fig. 3). Unlike the PTe activations in experiment 1, increased activity was located predominantly in ipsilateral AC (ipsilateral to the unilateral combined ear stimulation) but partly also in contralateral high-level AC ranging from mSTC to pSTC (Fig. 3b-c). While the ipsilateral increased activations followed the definition of "increased activity" defined above, the contralateral activity reflected the increased activity of a single speech sample (e.g. S_R) compared to the combined presentation of both speech samples (e.g. O_RS_R).

In addition, in experiment 2 only we found increased activity in the bilateral frontal cortices (vPMC, IFC). Since we did not expect lateralization effects on the frontal cortex according to lateralized experimental conditions, all frontal activity has been labeled according to the general definition of increased activity. This was also confirmed by a lateralization analysis, which showed that frontal activity was not lateralized to the left or right brain (Fig. S2).

These neural findings are summarized in Fig. 2a, including a laterality analysis of the AC activations from experiments 1 and 2 (Fig. 2b; see Fig. S2 for the laterality analysis on frontal activations). Unlike the frontal activations (t_{27}<1.141, p>0.419, t-tests; n = 28; all FDR corrected), all AC activations were significantly lateralized to their reported brain hemisphere (n = 28, t_{27}=2.662–9.287, p = 1.830 × 10^{-8}–0.023, FDR), except for left mSTC activity (t_{27}=1.681, p = 0.176) in experiment 2.

4. Discussion

The task of understanding other speech in noisy contextual conditions is a critical challenge in daily life conversations and is the topic of a broad research field in human neuroscience. While many previous studies investigated speech-in-noise perception while introducing external sources of noise, we here introduced a source of noise that is internal to the person listening to and decoding other speech samples. This internal noise of self speech is different from external noise in two ways: first, unlike external noise that is more of a random acoustic nature, self speech as a source of noise is acoustically more similar to the target other speech samples; second, this internal noise however is more predictable since it is self-initiated and self-shaped by the person that simultaneously produces self speech while listening to other speech.

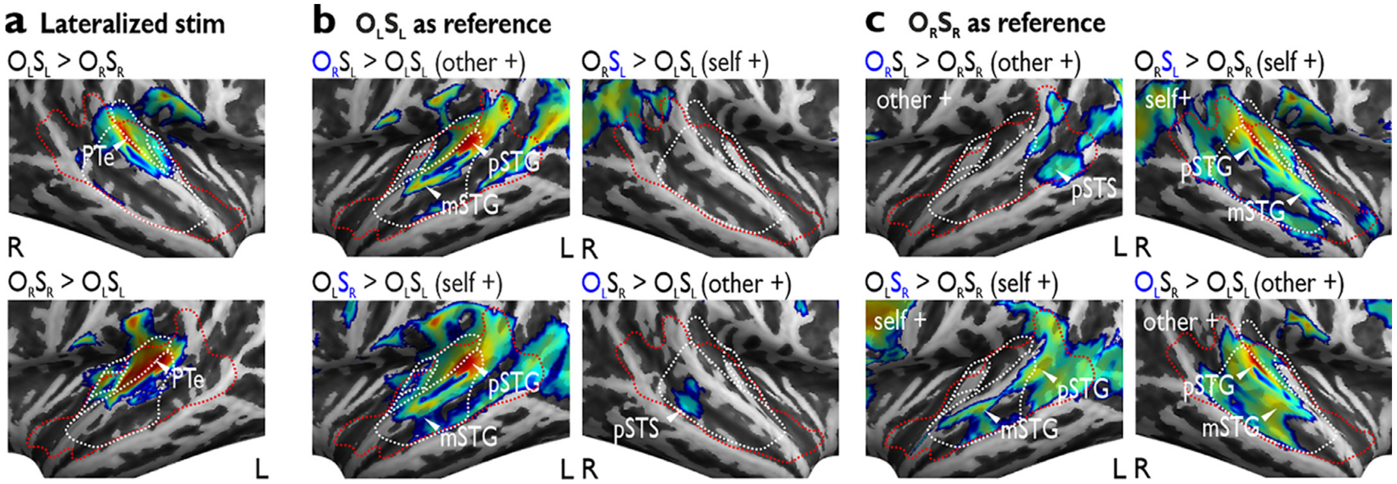
a Lateralized stim
$O_LS_L > O_RS_R$
PTe
R
$O_RS_R > O_LS_L$
PTe
L
b O_LS_L as reference
$O_RS_L > O_LS_L$ (other +)
pSTG
mSTG
L
$O_RS_L > O_LS_L$ (self +)
R
$O_LS_R > O_LS_L$ (self +)
pSTG
mSTG
L
$O_LS_R > O_LS_L$ (other +)
pSTS
R
c O_RS_R as reference
$O_RS_L > O_RS_R$ (other +)
other +
pSTS
L
$O_RS_L > O_RS_R$ (self +)
self+
pSTG
mSTG
R
$O_LS_R > O_RS_R$ (self +)
self +
pSTG
mSTG
L
$O_LS_R > O_LS_L$ (other +)
other +
pSTG
mSTG
R

Fig. 3. Functional AC activity in experiment 2. **(a)** Brain activation for unilateral stimulation. **(b)** Increased activity for SELF+ and OTHER+ largely in the left AC when compared to the O_LS_L condition as reference, and **(c)** for the same contrasts with O_RS_R as reference condition. Voxel $p<0.005$, cluster $k = 42$ ($p<0.05$ corrected at the cluster level), $n = 28$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

In our study, participants were asked to actively produce self speech (internal noise) while listening to and classifying other speech samples in experiment 1. In terms of behavioral classification performance of other speech, we found a similar performance in terms of reaction times for all four conditions. While we found no delay in the performance for a certain condition, we found that the condition O_LS_L (when compared to condition O_LS_R) seems to be a suboptimal case in terms of lower performance accuracy. This lower performance accuracy might be caused by the fact that during O_LS_L both speech signals are predominantly presented to the non-dominant right hemisphere (Price, 2012), but this challenge becomes attenuated when self speech is swapped to the right ear as in O_LS_R. The finding that O_LS_R revealed a better performance (compared to O_LS_L) however contradicts the general observation of a right ear advantage for other speech recognition (Lazard et al., 2012), which would imply that the O_RS_L should have resulted in the best performance. It might be that during concurrent self speech production while decoding other speech the O_LS_R condition leads to the least interference between both tasks, such that the left AC can accurately monitor self speech production (Christoffels et al., 2007; Hickok, 2012), and other speech recognition is performed by associated right brain mechanisms for speech decoding (Abrams et al., 2008; Xing et al., 2016).

We next looked at the brain activity that is elicited when comparing the four experimental conditions. First, when comparing unilateral stimulations with both speech samples, we found largely symmetric activity in the contralateral hemisphere that was largely focused on PTe. This area is supposed to represent secondary AC that integrates acoustic features for further processing (Griffiths and Warren, 2002). This PTe might serve to represent and separate both speech samples at a rather early stage of auditory processing. This is likely given that this separation should be rather easy to accomplish because the acoustic effects of self speech seem predictable based on correlative effects with vocal motor execution (Hickok, 2012). One exception from the symmetric PTe effects was activity in left aSTC for the O_LS_L condition. This latter asymmetric finding in the aSTC points to a specific interaction effect of self and other speech processing when presented to the left but not the right ear, accompanied by neural activity in a region that was thought to be at the center for other speech recognition (Evans et al., 2014; Scott et al., 2000). Given that we unilaterally stimulated with both speech samples, this left aSTC activity might more strongly respond to the other speech signal in the combined acoustic sample for accurate recognition of other speech. However, an alternative explanation for the left aSTC might be possible as we discuss below. Interestingly, this aSTC activity was only found in experiment 1 when active self speech had to be produced during the recognition of other speech.

In a second step we directly compared the bilateral stimulation conditions against each other (O_LS_R with O_RS_L). This was done to compare neural suppression effects for self speech with neural enhancement effects for other speech. Using these unilateral conditions ensured that the respective conditions (self speech, other speech) are directly compared in either left or right AC. If the AC would respond with opposing signal dynamics to own and other speech samples, this effect should be maximized in the comparison of these conditions. Surprisingly, none of these comparisons led to significant neural activations, especially in the case when we compared other speech processing (i.e. enhanced neural activity) with self speech processing (i.e. suppressed neural activity), which according to previous studies should lead to large signal differences. Our data thus seem to suggest that the contralateral suppressed cortical effect for self speech is either of a similar neural activity level as that for common other speech processing, or that ipsilateral influences of other speech lead to minimization or modulation of the contralateral self speech suppression effect. The latter could resemble the general observation that the self speech suppression effect disappears when additional acoustic noise is introduced during active self speech production and feedback registration (Christoffels et al., 2007). In any case, it seems like the commonly observed neural suppression effect for self speech in the AC is only observed in conditions with pure self speech (Hickok, 2012), but this effect seems to be modulated when additional listening and task conditions are introduced. Presenting other speech during concurrent self speech might figure as a challenging noise condition, and the self speech suppression effect is known to diminish with external noise (Christoffels et al., 2007).

In a third step, we then compared bilateral with unilateral stimulation conditions to either quantify decreased (i.e. when unilateral stimulation elicited higher activity than bilateral stimulation) or increased neural effects (i.e. when bilateral stimulation elicited higher activity than unilateral stimulation). In terms of increased activity, we found relatively symmetric effects in PTe, both for self speech and for other speech processing in experiment 1. These effects resemble the effects that we found when directly comparing the unilateral conditions against each other (see above). As all increased activation comparisons and the unilateral comparisons show similar activation profiles, it seems likely the PTe activity in our study resembles the notion of PTe being a computational hub for acoustic feature integration (Griffiths and Warren, 2002). This might support auditory classification and low-level speech detection indiscriminately for self- or other speech in a cortical region that is midway between low- and higher-order AC (Kumar et al., 2007).

Unlike these symmetric increased neural activations, the neural patterns for decreased activations were rather asymmetric and they were very specific for experiment 1. We found separate peaks for other speech processing in right mSTC and aSTC, and in right PTe and left aSTC for self speech processing. This points to a potential interaction effect between speech signals (i.e. decreased activations were found only when one but not the other speech signal was missing) of self and other speech processing, especially in the right hemisphere during the active speech task. The interaction effect between both speech signals is also highlighted by the fact that we did not find these peaks when comparing the bilateral conditions against each other (i.e. when both speech samples were presented to separate ears; see above). The activity that we found in left aSTC resembled the activity in left aSTC that we found when comparing the unilateral conditions, which also revealed an asymmetric effect in left aSTC. While the previous notion on the aSTC activity was relatively ambiguous about which speech sample could have caused the left aSTC activity, the more specific comparison between uni- and bilateral conditions pointed to effects of the missing right ear self speech sample that caused this decreased activity (i.e. self*). We therefore labeled this region as being predominantly responsible for decoding self speech signals. For other speech processing (Evans et al., 2014; Scott et al., 2000), the left aSTC is usually found to be part of a ventral processing stream for sound meaning analysis (Rauschecker and Scott, 2009). Surprisingly, our data seem to indicate that left aSTC activity might be specific to self rather than other speech processing when both are presented at the same time. Although task-relevant other speech processing might still be accomplished by the aSTC in our task (Evans et al., 2014; Scott et al., 2000), self speech analysis super-additively outweighs the effects of other speech processing. This effect is most likely based on self speech monitoring demands during the active production of the listeners' own-speech signals, but could potentially also be driven by the absence of strong effects of other speech signals in this area.

Concerning the functional data from experiment 1, we finally have to note that we did not find any activation in IFC when comparing the active speech conditions. This finding does not mean that there was no activity in the IFC during the active speech task while listening to other speech, but there was no significant difference in terms of neural activity between the four conditions. We also did not find frontal activity when comparing conditions for which we found differences in behavioral performance. This points to similar frontal activations between the conditions, and processing concurrent active self and other speech signals seem to be resolved by neural mechanisms of the AC. This highlights the notion that competition in experiment 1 largely seems to happen at the level of acoustic processing rather than higher-order speech identification.

A different pattern of neural effects, especially in the IFC, emerged however for neural activity comparisons in experiment 2. Experiment 2 was identical to experiment 1, with the exception that participants passively listened to self speech samples while simultaneous classifying other speech signals. We first looked at the behavioral performance in experiment 2, and found no difference in reaction times and the accuracy level across the four conditions. We however observed that reaction times were generally longer in experiment 2, pointing to a higher task difficulty of other speech classification. Other speech classification might have been easier in experiment 1, because self speech as a source of noise was more predictable in experiment 1 than in experiment 2. Experiment 2 required first to assign self and other labels to the two speech samples, which obviously requires more processing time. This increase in reaction time might have diminished differences between experimental conditions in terms of performance parameters.

To quantify neural activations in experiment 2, we applied the same comparisons as in experiment 1. When directly comparing the unilateral stimulation condition, we found the same neural effects in PTe as in experiment 1 except for the activity in left aSTC. Thus, the combined unilateral presentation of self and other speech samples leads to the same effects in PTe, independent of self speech being actively produced or being heard from recorded self speech samples. Only left aSTC activations seem specific to actively produced self speech. When comparing uni- with bilateral stimulations in experiment 2, we only found increased activations, but no decreased activations as in experiment 1. Unlike experiment 1, the increased activations were not predominantly located in PTe, but were shifted to more high-order AC along different locations in pSTC and mSTG. These increased activations were also relatively symmetric across the left and right STC. This shift of activation from secondary to high-level AC might represent a shift in processing requirements. Self and other speech can be more easily separated during active self speech, and thus both require only an acoustic analysis and integration (Griffiths and Warren, 2002). Separating self and other speech when listening to recorded self speech is a complex matter of speech signal attribution, which more likely involves high-level AC regions for auditory object recognition (Kumar et al., 2007) and social signals analysis (Belin et al., 2000).

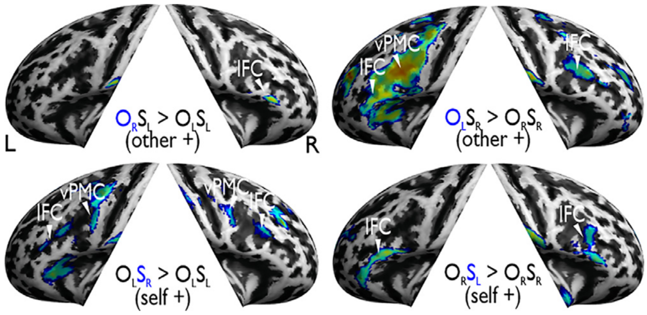

Fig. 4. Functional frontal activity in experiment 2. Increased activity in the bilateral vPMC and IFC only during experiment 2 for SELF+ and OTHER+. Voxel $p<0.005$, cluster $k = 42$ ($p<0.05$ corrected at the cluster level), $n = 28$.

Unlike in experiment 1, we found various frontal activations in experiment 2, located both in the inferior frontal gyrus (IFG) and in the ventral premotor cortex (vPMC). These frontal activations might help to resolve competition, especially in some higher-order cognitive processes to facilitate other speech classification in experiment 2 including various functions to support this classification. Specifically, the frontal activity might be related to attributing the two simultaneous presented speech signals to self and other signal sources, to orient attention to the other speech sample by articulatory representations of other and /or self speech, and to support an accurate decision on other speech samples. Specifically, activity in vPMC might support accurate other speech processing, presumably from motor mirroring (Meister et al., 2007; Wilson et al., 2004) and by accessing articulatory representations (Evans and Davis, 2015). These processes seem increased in experiment 2, because other speech classification happens somehow under suboptimal listening conditions caused by the concurrently presented self speech samples. Complementarily, the IFG activity might be more directly related to sound meaning classification as part of the auditory ventral stream (Frühholz and Grandjean, 2013b; Rauschecker and Scott, 2009). This is in correspondence with the increased reaction time found in experiment 2. The increased frontal activity was specifically found during the conditions with separate speech signals to both ears, specifically for the O_LS_R condition. The self/other attribution might require increased neural support, especially when these signals are presented to their non-dominant hemispheres as in O_LS_R during passive listening to both speech samples.

Taken together, the neural activations that we found in experiments 1 and 2 are summarized in Fig. 5. Our data suggest several important findings. First, our data seem to suggest that the previously described neural brain network for other speech processing (Evans et al., 2014;

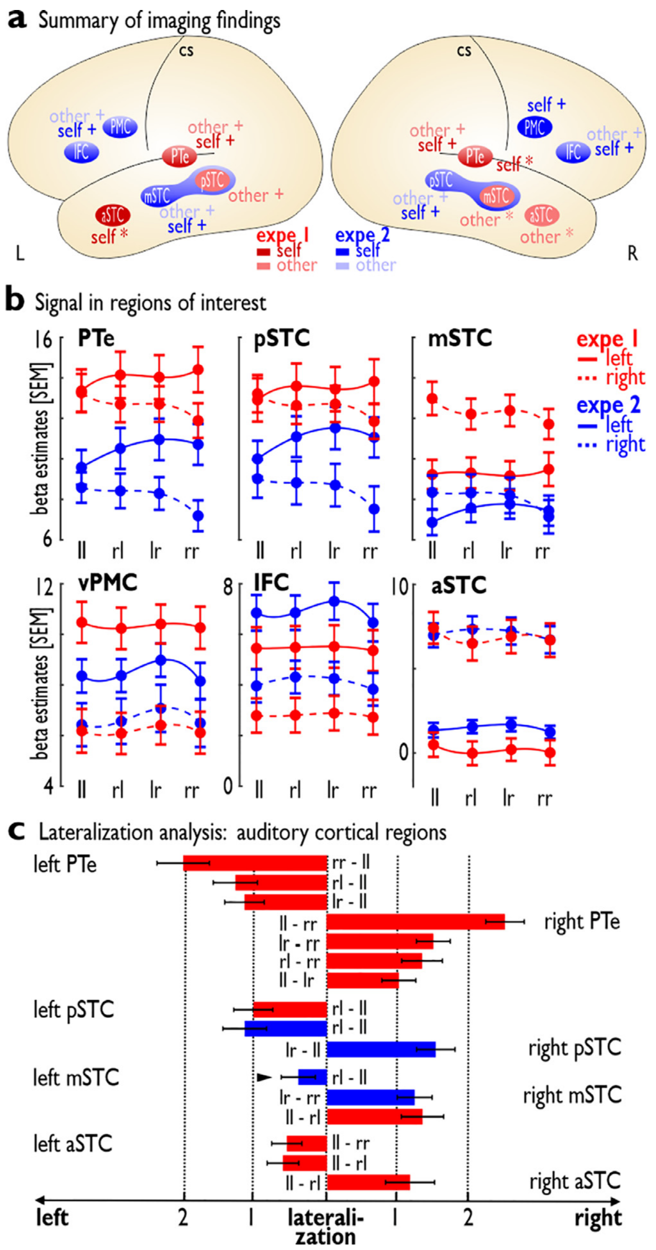

Fig. 5. Summary of findings, regions of interest, and laterality analysis. **(a)** Summary of the imaging findings for the left and right AC subregions (PTe, STC) and the frontal brain areas (PMC, IFC); other/self marks the preference for other- or self-speech; +/* indicates whether activity resulted from increased or decreased activity (see text). **(b)** Signal in regions of interest ($n = 28$) for all 4 conditions in experiment 1 (expe 1, red) and experiment 2 (expe 2, blue) for the left and right hemispheres. The signal profile was fitted with a 3rd degree polynomial function. **(c)** Lateralization analysis ($n = 28$) for AC subregions; all areas showed significant lateralization effects ($p<0.05$, FDR corrected) except for the left mSTC for the contrast $[O_R S_L > O_L S_L]$ (marked with ►). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Okada et al., 2010; Osnes et al., 2011; Scott et al., 2000) might show some functional flexibility and seem to functionally reorganize when there is a perceptual and cognitive rivalry with simultaneous self speech. More specifically, this functional flexibility and reorganization are was based on competitive acoustic processing of self and other speech depending on the lateralized presentation of these speech samples. Second, self speech processing during other speech recognition seems to elicit activity in left AC regions that were assumed to be relatively specialized to other speech perception. This points to a broader functional role of left AC regions in registering speech signals produced both by others and by the speakers' own speech. Third, other speech processing seems to recruit additional neural resources in the right AC compared to a rather dominant left lateralization in normal speech processing. Fourth, bilateral frontal speech processing regions, such as the IFC and vPM, were active only when individuals need to distinguish other from self speech in simultaneously presented playback speech samples. Together, these results indicate a more flexible brain organization for other speech perception, with neural resources being more adaptively distributed, depending on the challenging requirements of contextual constraints such as self speech production during other speech processing.

Data availability

The data and code used in thus study are available from the corresponding author upon reasonable request.

CRediT statements

J.D. and S.F. designed the experiment, acquired and analyzed the data, and wrote the manuscript; M.S. was involved in the data analysis. Fig. 4

Funding

The study was supported by the Swiss National Science Foundation (SNSF PP00P1_157409/1 and PP00P1_183711/1 to SF).

Declaration of Competing Interests

The authors declare no competing interests.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2020.117710.

References

Abrams, D.A., Nicol, T., Zecker, S., Kraus, N., 2008. Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. J. Neurosci. 28, 3958–3965. doi:10.1523/JNEUROSCI.0187-08.2008.

Assaneo, M.F., Ripollés, P., Orpella, J., Lin, W.M., de Diego-Balaguer, R., Poeppel, D., 2019. Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nat. Neurosci. 22, 627–632. doi:10.1038/s41593-019-0353-z.

Aziz-Zadeh, L., Sheng, T., Gheytanchi, A., 2010. Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. PLoS ONE 5. doi:10.1371/journal.pone.0008759.

Behroozmand, R., Shebek, R., Hansen, D.R., Oya, H., Robin, D.A., Howard, M.A., Greenlee, J.D.W., 2015. Sensory-motor networks involved in speech production and motor control: an fMRI study. Neuroimage 109, 418–428. doi:10.1016/j.neuroimage.2015.01.040.

Belin, P., Zatorre, R.J., Lafaillie, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312. doi:10.1038/35002078.

Brown, S., Ingham, R.J., Ingham, J.C., Laird, A.R., Fox, P.T., 2005. Stuttered and fluent speech production: an ALE meta-analysis of functional neuroimaging studies. In: Human Brain Mapping, pp. 105–117. doi:10.1002/hbm.20140.

Cheung, C., Hamiton, L.S., Johnson, K., Chang, E.F., 2016. The auditory representation of speech sounds in human motor cortex. Elife 5. doi:10.7554/eLife.12577.

Christoffels, I.K., de Van 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, e18307. doi:10.1371/journal.pone.0018307.

Christoffels, I.K., Formisano, E., Schiller, N.O., 2007. Neural correlates of verbal feedback processing: an fMRI study employing overt speech. Hum. Brain Mapp. 28, 868–879. doi:10.1002/hbm.20315.

Cogan, G.B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., Pesaran, B., 2014. Sensory-motor transformations for speech occur bilaterally. Nature 507, 94–98. doi:10.1038/nature12935.

Dauman, R., 2013. Bone conduction: an explanation for this phenomenon comprising complex mechanisms. Eur. Ann. Otorhinolaryngol. Head Neck Dis. doi:10.1016/j.anorl.2012.11.002.

Eickhoff, S.B., Heim, S., Zilles, K., Amunts, K., 2009. A systems perspective on the effective connectivity of overt speech production. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 367, 2399–2421. doi:10.1098/rsta.2008.0287.

Elmer, S., Kühnis, J., 2016. Functional connectivity in the left dorsal stream facilitates simultaneous language translation: an EEG study. Front. Hum. Neurosci. 10. doi:10.3389/fnhum.2016.00060.

Evans, S., Davis, M.H., 2015. Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. Cereb. Cortex 25, 4772–4788. doi:10.1093/cercor/bhv136.

Evans, S., Kyong, J.S., Rosen, S., Golestani, N., Warren, J.E., McGettigan, C., Mourão-Miranda, J., Wise, R.J.S., Scott, S.K., 2014. The pathways for intelligible speech: multivariate and univariate perspectives. Cereb. Cortex 24, 2350–2361. doi:10.1093/cercor/bht083.

Franken, M.K., Eisner, F., Acheson, D.J., McQueen, J.M., Hagoort, P., Schoffelen, J.M., 2018. Self-monitoring in the cerebral cortex: neural responses to small pitch shifts in auditory feedback during speech production. Neuroimage 179, 326–336. doi:10.1016/j.neuroimage.2018.06.061.

Friederici, A.D., 2011. The brain basis of language processing: from structure to function. Physiol. Rev. 91, 1357–1392. doi:10.1152/physrev.00006.2011.

Frühholz, S., Grandjean, D., 2013a. Processing of emotional vocalizations in bilateral inferior frontal cortex. Neurosci. Biobehav. Rev. 37, 2847–2855. doi:10.1016/j.neubiorev.2013.10.007.

Frühholz, S., Grandjean, D., 2013b. Processing of emotional vocalizations in bilateral inferior frontal cortex. Neurosci. Biobehav. Rev. 37, 2847–2855. doi:10.1016/j.neubiorev.2013.10.007.

Frühholz, S., Grandjean, D., 2012. Towards a fronto-temporal neural network for the decoding of angry vocal expressions. Neuroimage 62, 1658–1666. doi:10.1016/j.neuroimage.2012.06.015.

Frühholz, S., Gschwind, M., Grandjean, D., 2015. Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. Neuroimage 109, 27–34. doi:10.1016/j.neuroimage.2015.01.016.

Geiser, E., Zaehle, T., Jancke, L., Meyer, M., 2008. The neural correlate of speech rhythm as evidenced by metrical speech processing. J. Cogn. Neurosci. 20, 541–552. doi:10.1162/jocn.2008.20029.

Griffiths, T.D., Warren, J.D., 2002. The planum temporale as a computational hub. Trends Neurosci 25, 348–353. doi:10.1016/S0166-2236(02)02191-4.

Hickok, G., 2012. Computational neuroanatomy of speech production. Nat. Rev. Neurosci. 13, 135–145. doi:10.1038/nrn3158.

Hickok, G., 2010. The role of mirror neurons in speech and language processing. Brain Lang 112, 1–2. doi:10.1016/j.bandl.2009.10.006.

Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69, 407–422. doi:10.1016/j.neuron.2011.01.019.

Kumar, S., Stephan, K.E., Warren, J.D., Friston, K.J., Griffiths, T.D., 2007. Hierarchical processing of auditory objects in humans. PLoS Comput. Biol. 3, 0977–0985. doi:10.1371/journal.pcbi.0030100.

Lazard, D.S., Collette, J.L., Perrot, X., 2012. Speech processing: from peripheral to hemispheric asymmetry of the auditory system. Laryngoscope doi:10.1002/lary.22370.

Markiewicz, C.J., Bohland, J.W., 2016. Mapping the cortical representation of speech sounds in a syllable repetition task. Neuroimage 141, 174–190. doi:10.1016/j.neuroimage.2016.07.023.

Meister, I.G., Wilson, S.M., Deblieck, C., Wu, A.D., Iacoboni, M., 2007. The essential role of premotor cortex in speech perception. Curr. Biol. 17, 1692–1696. doi:10.1016/j.cub.2007.08.064.

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., Friederici, A.D., 2012. Perception of words and pitch patterns in song and speech. Front. Psychol. 3, 76. doi:10.3389/fpsyg.2012.00076.

Möttönen, R., Dutton, R., Watkins, K.E., 2013. Auditory-motor processing of speech sounds. Cereb. Cortex 23, 1190–1197. doi:10.1093/cercor/bhs110.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb. Cortex 20, 2486–2495. doi:10.1093/cercor/bhp318.

Osnes, B., Hugdahl, K., Specht, K., 2011. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. Neuroimage 54, 2437–2445. doi:10.1016/j.neuroimage.2010.09.078.

Parkinson, A.L., Flagmeier, S.G., Manes, J.L., Larson, C.R., Rogers, B., Robin, D.A., 2012. Understanding the neural mechanisms involved in sensory control of voice production. Neuroimage 61, 314–322. doi:10.1016/j.neuroimage.2012.02.068.

Pernet, C.R., McAleer, P., Latinus, M., Gorgolewski, K.J., Charest, I., Bestelmeyer, P.E.G., Watson, R.H., Fleming, D., Crabbe, F., Valdes-Sosa, M., Belin, P., 2015. The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. Neuroimage 119, 164–174. doi:10.1016/j.neuroimage.2015.06.050.

Price, C.J., 2012. A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. Neuroimage doi:10.1016/j.neuroimage.2012.04.062.

Rampinini, A.C., Handjaras, G., Leo, A., Cecchetti, L., Ricciardi, E., Marotta, G., Pietrini, P., 2017. Functional and spatial segregation within the inferior frontal and superior temporal cortices during listening, articulation imagery, and production of vowels. Sci. Rep. 7. doi:10.1038/s41598-017-17314-0.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724. doi:10.1038/nn.2331.

Scott, S.K., Catrin Blank, C., Rosen, S., Wise, R.J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406. doi:10.1093/brain/123.12.2400.

Scott, S.K., McGettigan, C., Eisner, F., 2009. A little more conversation, a little less action - candidate roles for the motor cortex in speech perception. Nat. Rev. Neurosci. 10, 295–302. doi:10.1038/nrn2603.

Stokes, R.C., Venezia, J.H., Hickok, G., 2019. The motor system's [modest] contribution to speech perception. Psychon. Bull. Rev. 26, 1354–1366. doi:10.3758/s13423-019-01580-2.

Tyler, L.K., Stamatakis, E.A., Post, B., Randall, B., Marslen-Wilson, W., 2005. Temporal and frontal systems in speech comprehension: an fMRI study of past tense processing. Neuropsychologia 43, 1963–1974. doi:10.1016/j.neuropsychologia.2005.03.008.

Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nat. Neurosci. 7, 701–702. doi:10.1038/nn1263.

Xing, S., Lacey, E.H., Skipper-Kallal, L.M., Jiang, X., Harris-Love, M.L., Zeng, J., Turkeltaub, P.E., 2016. Right hemisphere grey matter structure and language outcomes in chronic left hemisphere stroke. Brain 139, 227–241. doi:10.1093/brain/awv323.