



Converging Evidence From Electrocorticography and BOLD fMRI for a Sharp Functional Boundary in Superior Temporal Gyrus Related to Multisensory Speech Processing

Muge Ozker¹, Daniel Yoshor^{1,2} and Michael S. Beauchamp^{1*}

¹Department of Neurosurgery, Baylor College of Medicine, Houston, TX, United States, ²Michael E. DeBakey Veterans Affairs Medical Center, Houston, TX, United States

OPEN ACCESS

Edited by:

Cheryl Olman,
University of Minnesota,
United States

Reviewed by:

Arun Bokde,
Trinity College, Dublin, Ireland
Julian Keil,
Christian-Albrechts-Universität zu
Kiel, Germany

*Correspondence:

Michael S. Beauchamp
michael.beauchamp@bcm.edu

Received: 05 January 2018

Accepted: 28 March 2018

Published: 24 April 2018

Citation:

Ozker M, Yoshor D and Beauchamp MS (2018) Converging Evidence From Electrocorticography and BOLD fMRI for a Sharp Functional Boundary in Superior Temporal Gyrus Related to Multisensory Speech Processing. *Front. Hum. Neurosci.* 12:141. doi: 10.3389/fnhum.2018.00141

Although humans can understand speech using the auditory modality alone, in noisy environments visual speech information from the talker's mouth can rescue otherwise unintelligible auditory speech. To investigate the neural substrates of multisensory speech perception, we compared neural activity from the human superior temporal gyrus (STG) in two datasets. One dataset consisted of direct neural recordings (electrocorticography, ECoG) from surface electrodes implanted in epilepsy patients (this dataset has been previously published). The second dataset consisted of indirect measures of neural activity using blood oxygen level dependent functional magnetic resonance imaging (BOLD fMRI). Both ECoG and fMRI participants viewed the same clear and noisy audiovisual speech stimuli and performed the same speech recognition task. Both techniques demonstrated a sharp functional boundary in the STG, spatially coincident with an anatomical boundary defined by the posterior edge of Heschl's gyrus. Cortex on the anterior side of the boundary responded more strongly to clear audiovisual speech than to noisy audiovisual speech while cortex on the posterior side of the boundary did not. For both ECoG and fMRI measurements, the transition between the functionally distinct regions happened within 10 mm of anterior-to-posterior distance along the STG. We relate this boundary to the multisensory neural code underlying speech perception and propose that it represents an important functional division within the human speech perception network.

Keywords: multisensory, speech perception, temporal lobe, electrocorticography (ECoG), BOLD fMRI, audiovisual speech perception, multisensory integration, speech in noise

INTRODUCTION

The human ability to understand speech is one of our most important cognitive abilities. While speech can be understood using the auditory modality alone, vision provides important additional cues about speech. In particular, the mouth movements made by the talker can compensate for degraded or noisy auditory speech (Sumby and Pollack, 1954; Bernstein et al., 2004; Ross et al., 2007). While it has been known since Wernicke that posterior lateral temporal cortex is important for language comprehension, the advent of blood-oxygen level dependent functional magnetic resonance imaging (BOLD fMRI) led to important advances, such as the discovery that multiple regions in temporal cortex are selective for human voices (Belin et al., 2000). However, BOLD fMRI

suffers from a major limitation, in that it is a slow and indirect measure of neural function. Spoken speech contains five or more syllables per second, requiring the neural processes that decode each syllable to be completed in less than 200 ms. In contrast, the sluggish hemodynamic response that underlies BOLD fMRI does not peak until several seconds after the neural activity that prompted it.

This drawback underscores the importance of complementing fMRI with other techniques that directly measure neural activity. The non-invasive techniques of EEG and MEG have led to a better understanding of the temporal dynamics of speech perception (Salmelin, 2007; Shahin et al., 2012; Crosse et al., 2016; Sohoglu and Davis, 2016). Recently, there has also been tremendous interest in electrocorticography (ECoG), a technique in which electrodes are implanted in the brains of patients with medically intractable epilepsy. Compared with EEG and MEG, ECoG allows localization of activity to the small population of neurons nearest each electrode, leading to the discovery of selective responses in the superior temporal gyrus (STG) for various speech features, including categorical representations of speech (Chang et al., 2010) phonetic features (Mesgarani et al., 2014) and prosody (Tang et al., 2017).

While the broad outlines of the organization of visual cortex are well-established (Grill-Spector and Malach, 2004), the layout of auditory cortex is less well known. Early areas of auditory cortex centered on Heschl's gyrus contain maps of auditory frequency and spectral temporal modulation (Moerel et al., 2014). In contrast, within auditory association cortex in the STG, organization by auditory features is weaker, and the location and number of different functional areas is a matter of controversy (Leaver and Rauschecker, 2016). Recently, we used ECoG to document a double dissociation between anterior and posterior regions of the STG (Ozker et al., 2017). Both regions showed strong responses to audiovisual speech, but the anterior area responded more strongly to speech in which the auditory component was clear while the posterior area showed a different response pattern, responding similarly to clear and noisy speech or even responding more strongly to noisy audiovisual speech. There was a sharp anatomical boundary, defined by the posterior edge of Heschl's gyrus, between the two areas. All electrodes anterior to the boundary responded more to clear speech and no electrodes posterior to the boundary did. These results were interpreted in the conceptual framework of multisensory integration. Auditory association areas in anterior STG respond strongly to clear auditory speech but show a reduced response because of the reduced information available in noisy auditory speech, paralleling the reduction in speech intelligibility. Multisensory areas in posterior STG are able to use the visual speech information to compensate for the noisy auditory speech, restoring intelligibility. However, this demands recruitment of additional neuronal resources,

leading to an increased response during noisy audiovisual speech perception.

While there have been numerous previous fMRI studies of noisy and clear audiovisual speech (e.g., Callan et al., 2003; Sekiyama et al., 2003; Bishop and Miller, 2009; Stevenson and James, 2009; Lee and Noppeney, 2011; McGgettigan et al., 2012), none described a sharp boundary in the response patterns to clear and noisy speech within the STG. BOLD fMRI has the spatial resolution necessary to detect fine-scale cortical boundaries, such as between neighboring ocular dominance columns (Cheng et al., 2001), ruling out sensitivity of the technique itself as an explanation. Instead, we considered two other possibilities. One possible explanation is that the analysis or reporting strategies used in previous fMRI studies (such as group averaging or reporting only activation peaks) could have obscured a sharp functional boundary present in the fMRI data. A second, more worrisome, explanation is that the sharp boundary observed with ECoG reflects anomalous brain organization in the ECoG participants. Brain reorganization due to repeated seizures could have resulted in different STG functional properties in epileptic patients compared with healthy controls (Janszky et al., 2003; Kramer and Cash, 2012).

To distinguish these possibilities, in the present manuscript, we compare the ECoG dataset previously published in Ozker et al. (2017) with a new BOLD fMRI dataset not previously published. The healthy controls in the fMRI experiment viewed the same clear and noisy audiovisual speech stimuli viewed by the ECoG patients, and both set of participants performed the identical speech identification task. The BOLD fMRI data was analyzed without any spatial blurring or group averaging to ensure that these would not obscure areal boundaries within the STG. fMRI samples the entire brain volume, instead of the limited coverage obtained with ECoG electrodes, allowing for an examination of the responses to clear and noisy speech across the entire length of the STG.

MATERIALS AND METHODS

All participants (see **Table 1** for demographic information) provided written informed consent and underwent experimental procedures approved by the Baylor College of Medicine (BCM) Institutional Review Board.

For the main experiment, identical stimuli were used for the ECoG and fMRI participants. The stimuli consisted of audiovisual recordings of a female talker from the Hoosier Audiovisual Multi-Talker Database speaking single words ("rain" or "rock") in which the auditory component was either unaltered (auditory-clear) or replaced with speech-specific noise that matched the spectrotemporal power distribution

TABLE 1 | Subject demographics.

Participant group	N	Age (mean)	Age (range)	Gender
ECoG	5	31	21–51	3F, 2M
fMRI	6	25	19–31	3F, 3M

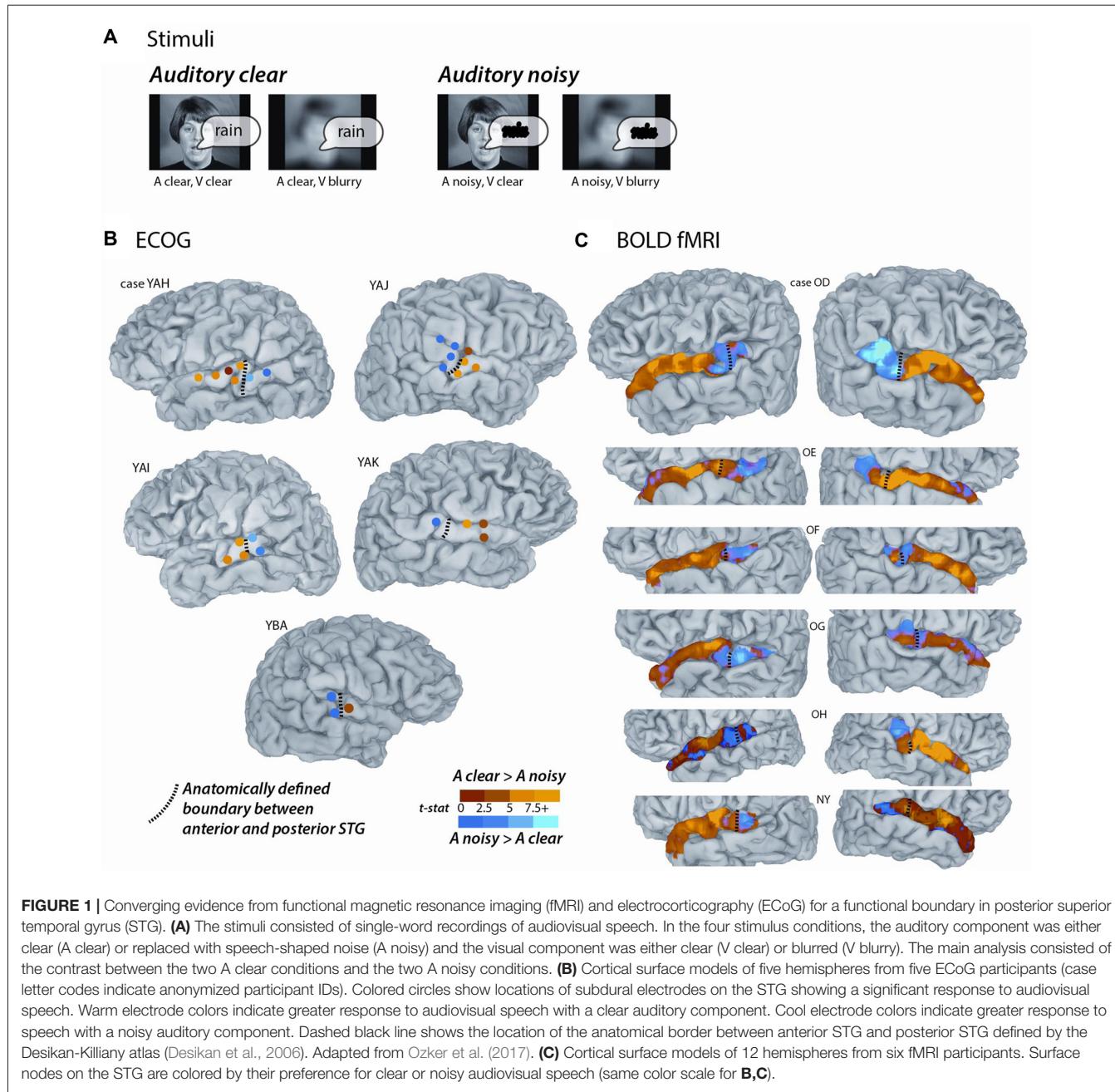


FIGURE 1 | Converging evidence from functional magnetic resonance imaging (fMRI) and electrocorticography (ECoG) for a functional boundary in posterior superior temporal gyrus (STG). **(A)** The stimuli consisted of single-word recordings of audiovisual speech. In the four stimulus conditions, the auditory component was either clear (A clear) or replaced with speech-shaped noise (A noisy) and the visual component was either clear (V clear) or blurred (V blurry). The main analysis consisted of the contrast between the two A clear conditions and the two A noisy conditions. **(B)** Cortical surface models of five hemispheres from five ECoG participants (case letter codes indicate anonymized participant IDs). Colored circles show locations of subdural electrodes on the STG showing a significant response to audiovisual speech. Warm electrode colors indicate greater response to audiovisual speech with a clear auditory component. Cool electrode colors indicate greater response to speech with a noisy auditory component. Dashed black line shows the location of the anatomical border between anterior STG and posterior STG defined by the Desikan-Killiany atlas (Desikan et al., 2006). Adapted from Ozker et al. (2017). **(C)** Cortical surface models of 12 hemispheres from six fMRI participants. Surface nodes on the STG are colored by their preference for clear or noisy audiovisual speech (same color scale for **B,C**).

of the original auditory speech (auditory-noisy). A parallel manipulation was performed on the visual component of the speech by replacing the original video with a blurred version. The two types of auditory speech and two types of visual speech were combined, resulting in four conditions (auditory-clear + visual-clear; auditory-clear + visual-blurred; auditory-noisy + visual-clear; auditory-noisy + visual-blurred). Schematic depictions of the stimuli are shown in **Figure 1A**, and the actual stimuli used in the experiments can be downloaded from <https://openwetware.org/wiki/Beauchamp:Stimuli>.

The face of the talker subtended approximately 15° of visual angle horizontally and vertically. For the ECoG participants,

stimuli were viewed on a 15" LCD monitor positioned at 57-cm distance from the participant and auditory stimuli were played through wall-mounted loudspeakers. For the fMRI participants, stimuli were viewed on a 32" BOLDview LCD Screen (Cambridge Research Systems) located behind the scanner and viewed through a mirror affixed to the head coil (approximately 77-cm from screen to participant). Auditory stimuli were presented through stereo insert headphones (Sensimetrics). For both ECoG and fMRI participants, sound levels were adjusted to a comfortable volume before the experiment. The duration of each video clip was 1.4 s and the duration of the auditory stimulus was 520 ms for the "rain" stimulus and 580 ms

for the “rock” stimulus. The auditory word onsets occurred after the video onset, 410 ms for “rain” and 450 ms for “rock”.

For the ECoG participants, from 32 to 56 repetitions of each condition were presented in random order. For the fMRI participants, 60 repetitions of each condition were presented in random order. Following each stimulus presentation, participants performed a two-alternative forced choice on the identity of the presented word.

Definition of Anterior and Posterior Superior Temporal Gyrus (STG)

Cortical surface models were constructed from the high-resolution T1-weighted anatomical MRI scans of ECoG and fMRI participants using FreeSurfer (Fischl et al., 2001). For ECoG participants, the post-implantation computed tomography (CT) scan showing the electrode locations was registered to the anatomical MRI to ensure accurate electrode localization.

Two atlases were used to parcellate the STG. The Destrieux atlas defines the entire STG using the “G_temp_sup-Lateral” label (lateral aspect of the STG; Destrieux et al., 2010). The Desikan-Killiany atlas (Desikan et al., 2006) applies a single “Superior Temporal” label to both the STG and the STS with an additional “Banks Superior Temporal” label for the posterior portion of the STS, with an anterior border defined by the posterior-most point of Heschl’s gyrus. We cleaved the Destrieux STG into an anterior STG portion and a posterior STG portion using the Heschl’s gyrus landmark defined by the Desikan-Killiany atlas (boundary shown as black dashed lines in **Figure 1**). The posterior STG is continuous with the supramarginal gyrus. Since the two atlases vary in their handling of this boundary, we manually defined the posterior boundary of the posterior STG as being just past the location where the gyrus begins its sharp turn upward into parietal lobe. All analyses were done only within single participants without any normalization or spatial blurring. In order to report the location of the anterior-posterior boundary in standard space, individual MRIs were aligned to the N27 brain (Holmes et al., 1998).

ECoG Experimental Design and Data Analysis

The ECoG dataset was previously published in Ozker et al. (2017). Experiments were conducted in the epilepsy monitoring unit of Baylor St. Luke’s Medical Center. Patients rested comfortably in their hospital beds while viewing stimuli presented on an LCD monitor mounted on a table and positioned at 57 cm distance from the participant. While the participants viewed stimulus movies, a 128-channel Cerebus amplifier (Blackrock Microsystems, Salt Lake City, UT, USA) recorded from subdural electrodes that consisted of platinum alloy discs (diameter 2.3 mm) embedded in a flexible silicon sheet with inter-electrode distance of 10 mm. An inactive intracranial electrode implanted facing the skull was used as a reference for recording. Signals were amplified, filtered and digitized at 2 kHz. Offline, common average referencing was used to

remove artifacts, and the data was epoched according to stimulus timing. Line noise was removed and spectral decomposition was performed using multitapers. The measure of neural activity was the broad-band high-gamma response (70–110 Hz) measured as the percent change relative to a pre-stimulus baseline window (500–100 ms before auditory stimulus onset). The high-gamma broadband response was used as it is the ECoG signal most closely associated with the rate of action potentials and the BOLD fMRI response (reviewed in Ray and Maunsell, 2011; Lachaux et al., 2012; Ojemann et al., 2013). Across patients, a total of 527 intracranial electrodes were recorded from. Of these, 55 were located on the STG. Twenty-seven of these showed a minimal level of stimulus-related activity, defined as significant high-gamma responses to audiovisual speech compared with prestimulus baseline ($p < 10^{-3}$, equivalent to 40% increase in stimulus power from baseline) and were included in the analysis.

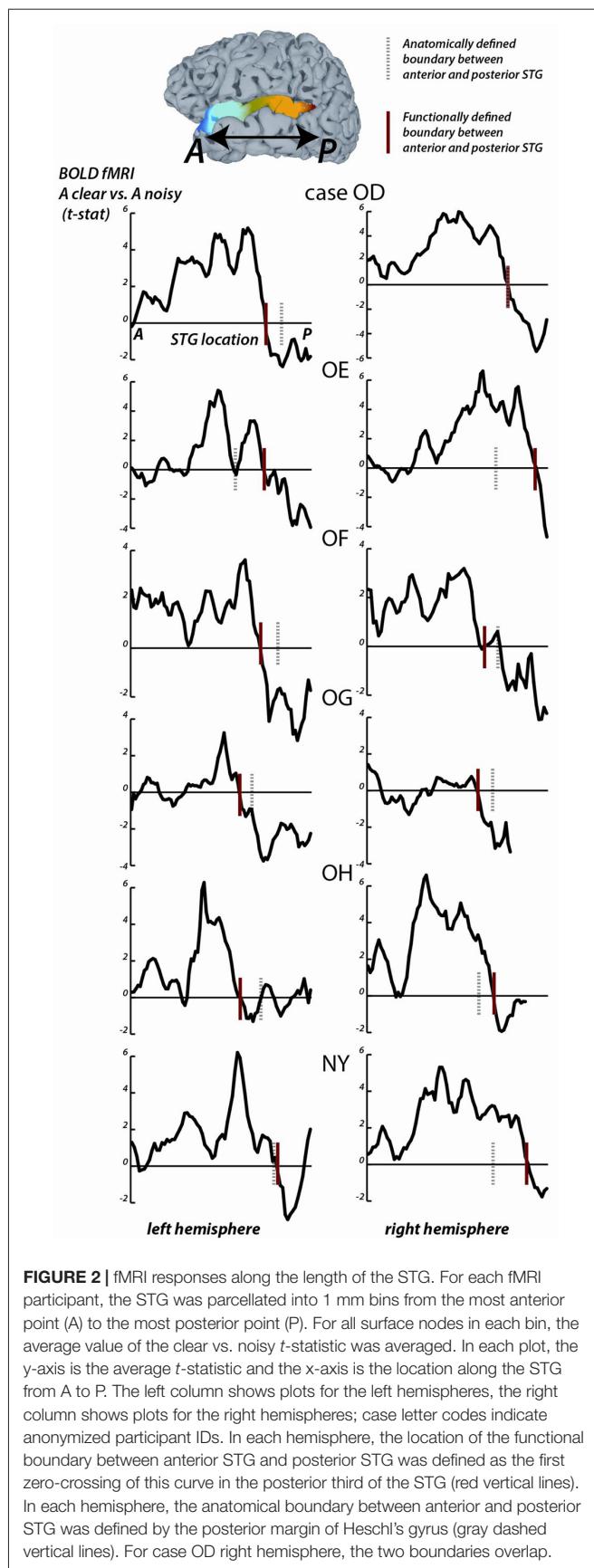
fMRI Experimental Design and Data Analysis

Experiments were conducted in the Core for Advanced MRI (CAMRI) at BCM using a 3 Tesla Siemens Trio MR scanner equipped with a 32-channel head gradient coil. BOLD fMRI data was collected using a multislice echo planar imaging sequence (Setsompop et al., 2012) with TR = 1500 ms, TE = 30 ms, flip angle = 72°, in-plane resolution of 2×2 mm, 69 2-mm axial slices, multiband factor: 3, GRAPPA factor: 2. fMRI data was analyzed using the afni_proc.py pipeline (Cox, 1996). Data was time shifted to account for different acquisition times for different slices; aligned to the first functional volume which was in turn aligned with the high-resolution anatomical; and rescaled so that each voxel had a mean of 100. No blurring or spatial normalization of any sort was applied to the EPI data. Five runs (scan series) were collected, each with 160 brain volumes (4 min duration). Each run contained 48 3-s trials, 12 for each stimulus condition, for a total of sixty repetitions of each condition. A rapid event-related design was used with fixation baseline occupying the remaining 96 s of each run, optimized with the scheduling algorithm optseq² (Dale et al., 1999).

A generalized linear model was used to model the fMRI time series independently for each voxel using the 3dDeconvolve function in AFNI. The model contained 10 regressors: six regressors of no interest generated by the motion correction process and four regressors of interest (one for each stimulus condition) using an exponential hemodynamic response function (HRF) generated with the 3dDeconvolve option “BLOCK(2,1)”. A general linear test with the values of “+1 +1 -1 -1” was used to find the *t*-statistic for the contrast between the two conditions with clear auditory speech and the two conditions with noisy auditory speech (data in **Figures 1, 2**). This contrast between auditory-clear and auditory-noisy was the main dependent measure in the analysis.

For the STG length analysis (**Figure 2**), unthresholded fMRI data in the form of the clear vs. noisy *t*-statistic was mapped to the cortical surface using the AFNI function 3dVol2Surf. The

²<https://surfer.nmr.mgh.harvard.edu/optseq>



options “-ave -f_steps 15” were used, resulting in a line between each node on the pial surface and the corresponding node on the smoothed white matter surface being subdivided into 15 equal segments, with the fMRI voxel values at each segment sampled and averaged. The entire STG was divided into 1 mm bins, from anterior to posterior, and the *t*-statistic at all nodes within each bin was averaged. For each hemisphere, a functional boundary was defined as the bin containing the first zero-crossing of the *t*-statistic (moving in an anterior-to-posterior direction) in the posterior third of the STG.

To estimate the shape of the HRF without assumptions, a second model was constructed that used tent functions to estimate the amplitude of the response independently at each time point of the BOLD response. For the main fMRI experiment, the response window spanned 0–15 s after stimulus onset (11 time points at a TR of 1.5 s) using the 3dDeconvolve option “TENTzero(0,15,11)”, resulting in a model with 50 regressors (6 motion regressors and 44 regressors of interest).

To estimate the average BOLD fMRI HRF, anterior and posterior STG ROIs were created in each participant using as a boundary either the anatomical Heschl's gyrus boundary or the functional boundary defined by the STG length analysis. The anterior STG ROI contained all voxels from 0 mm to 30 mm anterior to the boundary and the posterior STG ROI contained all voxels from 0 mm to 15 mm posterior to the boundary. These values were chosen for consistency with the ECoG electrode locations, which ranged from 30 mm anterior to the anatomical/functional boundary to 15 mm posterior to it (**Figure 1B**). For correspondence with the ECoG electrode selection criteria (in which only electrodes that showed some response were included in the analysis) only voxels with an omnibus Full-F statistic of $F > 5$ ($q < 10^{-6}$) were included in the ROIs.

To directly compare the BOLD fMRI with the ECoG responses, the ECoG response were convolved with a double-gamma HRF with peak time = 6 s, undershoot time = 10 s and response-to-undershoot ratio = 4 (Lindquist et al., 2009). The only free parameter was a scale parameter that matched the amplitude of the predicted and actual fMRI responses; scale parameters that minimized the difference between the predicted and actual fMRI responses for each of the four curves were found using the Matlab function *fminbnd*.

Linear Mixed-Effects Models

Linear mixed-effect models (LMEs) were constructed using *R* with the *lme4* package. LMEs are similar to repeated-measure analysis of variances (ANOVAs) but have several advantages: LMEs are more statistically conservative, LMEs better handle missing observations and LMEs account for the correlation structure of the variables. The dependent measure for each LME was the % signal change from baseline. For the ECoG data, each electrode constituted an independent sample, and the responses to each stimulus condition were entered into the LME. For the fMRI data, each hemisphere constituted an independent sample, and the responses to each stimulus condition in the anterior and posterior STG ROIs in that hemisphere were entered into the LME. The fixed factors were location (anterior vs. posterior

STG), the presence or absence of auditory noise (auditory-clear vs. auditory-noisy) and the presence or absence of visual blur (visual-clear vs. visual-blurry). For each fixed factor, the LME estimated the significance of the effect and the magnitude of the effect relative to a baseline condition, which was always the response to auditory-clear, visual-clear speech in anterior STG.

Additional fMRI Data

Additional fMRI data was collected while participants were passively presented with 20-s excerpts from short stories (Aesop's fables) presented in auditory-only, visual-only and audiovisual versions, all recorded by the same female talker (Nath et al., 2011; Nath and Beauchamp, 2012). Data with these stimuli were collected in two runs, each with 180 brain volumes (4:30 duration); for one participant, only one run was collected. Each run contained nine blocks (20 s of stimulus, 10 s of fixation baseline) consisting of three blocks each (in random order) of the three types of stimuli. The HRF was estimated using a response window spanning 0–30 s after stimulus onset (21 time points) using “TENTzero(0,30,21)” for each of the three stimulus types. The beta-weights at the 4.5 s, 6 s and 7.5 s time points were averaged to create the estimate of response amplitude.

RESULTS

Participants were presented with audiovisual speech stimuli with a clear or noisy auditory component and a clear or blurry visual component (Figure 1A). In the ECoG dataset, electrodes implanted in different locations on the STG responded differently. Electrodes on the anterior STG responded more strongly to audiovisual speech with a clear auditory component while electrodes on posterior STG did not (Figure 1B). The posterior-most point of Heschl's gyrus has been proposed as a boundary dividing the STG/STS into anterior and posterior sections (Desikan et al., 2006; Ozker et al., 2017). All electrodes anterior to the boundary responded more strongly to clear speech while none of the electrodes posterior to the boundary did. The difference in response patterns was striking, even between electrodes that were only 10 mm apart, the closest possible distance in our recording array. For instance, in one participant

the response to clear speech of an anterior electrode was more than double its response to noisy speech ($138\% \pm 13\%$ vs. $49\% \pm 5\%$, mean across trials \pm SEM; unpaired *t*-test: $t_{109} = +6.2$, $p = 10^{-8}$) while the adjacent electrode, located 10 mm posterior across the boundary, responded less than half as much to clear speech as noisy speech ($38\% \pm 5\%$ vs. $89\% \pm 9\%$, $t_{109} = -4.5$, $p = 10^{-5}$). These effects were quantified across electrodes using a LME model (Table 2). There were three significant effects in the model: a main effect of location, driven by smaller responses in the posterior STG ($p = 0.01$, effect magnitude of 101%); a main effect of auditory noise, driven by weaker responses to noisy stimuli ($p = 10^{-13}$, magnitude 110%); and an interaction between location and auditory noise, driven by a larger response to noisy auditory stimuli in posterior STG ($p = 10^{-10}$, magnitude 141%).

To determine if a similar anterior-posterior STG boundary could be observed with fMRI, we scanned participants viewing the same stimuli as the ECoG participants and mapped the unthresholded statistical contrast of clear vs. noisy speech along the STG of each hemisphere (Figure 1C). Anterior STG responded more strongly to clear speech while posterior STG did not.

To quantify the location of the anterior-posterior boundary, the preference for clear vs. noisy audiovisual speech in unthresholded fMRI data was plotted in 1 mm bins along the entire anterior-to-posterior extent of the STG (Figure 2). The sign change in the *t*-statistic of the clear vs. noisy contrast was used to define a *functional* A-P boundary in the STG (red lines in Figure 2) for comparison with the *anatomical* A-P boundary in the STG defined by the posterior margin of Heschl's gyrus (black lines in Figure 2).

The mean anterior-to-posterior location of the fMRI-defined functional boundary in standard space was $y = -28$ mm (± 9 mm SD). The mean standard space location of the atlas-defined anatomical boundary in these participants was $y = -30$ mm (± 5 mm SD).

In some cases, the boundaries aligned remarkably well (inter-boundary distance of 1 mm, case OD right hemisphere) while in others they were farther apart (distance of 20 mm, case OE right hemisphere). There was no consistent anterior-to-posterior

TABLE 2 | Linear mixed-effects (LMEs) model of the response amplitude in anterior and posterior electrocorticography (ECoG) electrodes.

Fixed effects	Estimate	Std. Error	DF	t-value	p-value
Baseline	183.1	24.8	33.7	7.4	10^{-8}
Auditory noise (An)	-109.6	13.5	188	-8.1	10^{-13}
Posterior location × An	140.6	21.2	188	6.6	10^{-10}
Posterior location	-10.1	38.7	34.2	-2.6	0.01
Visual blur (Vb)	21.6	13.5	188	1.6	0.11
An × Vb	-13.3	19.1	188	-0.7	0.49
Posterior location × Vb	-8.9	21.2	188	-0.4	0.67
Posterior location × An × Vb	3.6	29.9	188	0.1	0.91

Results of an LME model of the ECoG response amplitudes, reprinted from Ozker et al. (2017). The fixed effects were the location of each electrode (Anterior vs. Posterior), the presence or absence of auditory noise (An) in the stimulus and the presence or absence of visual blur (Vb) in the stimulus. Electrodes and stimulus exemplar were included in the model as random factors. For each effect, the model estimates (in units of percent signal change) for that factor are shown relative to baseline, the response in anterior electrodes to clear audiovisual speech (AV stimulus condition). The baseline is shown first; all other effects are ranked by absolute t value. Significant effects are shown in bold. The significance of the baseline fixed effect is italicized because it was prespecified: only electrodes with significant amplitudes were included in the analysis.

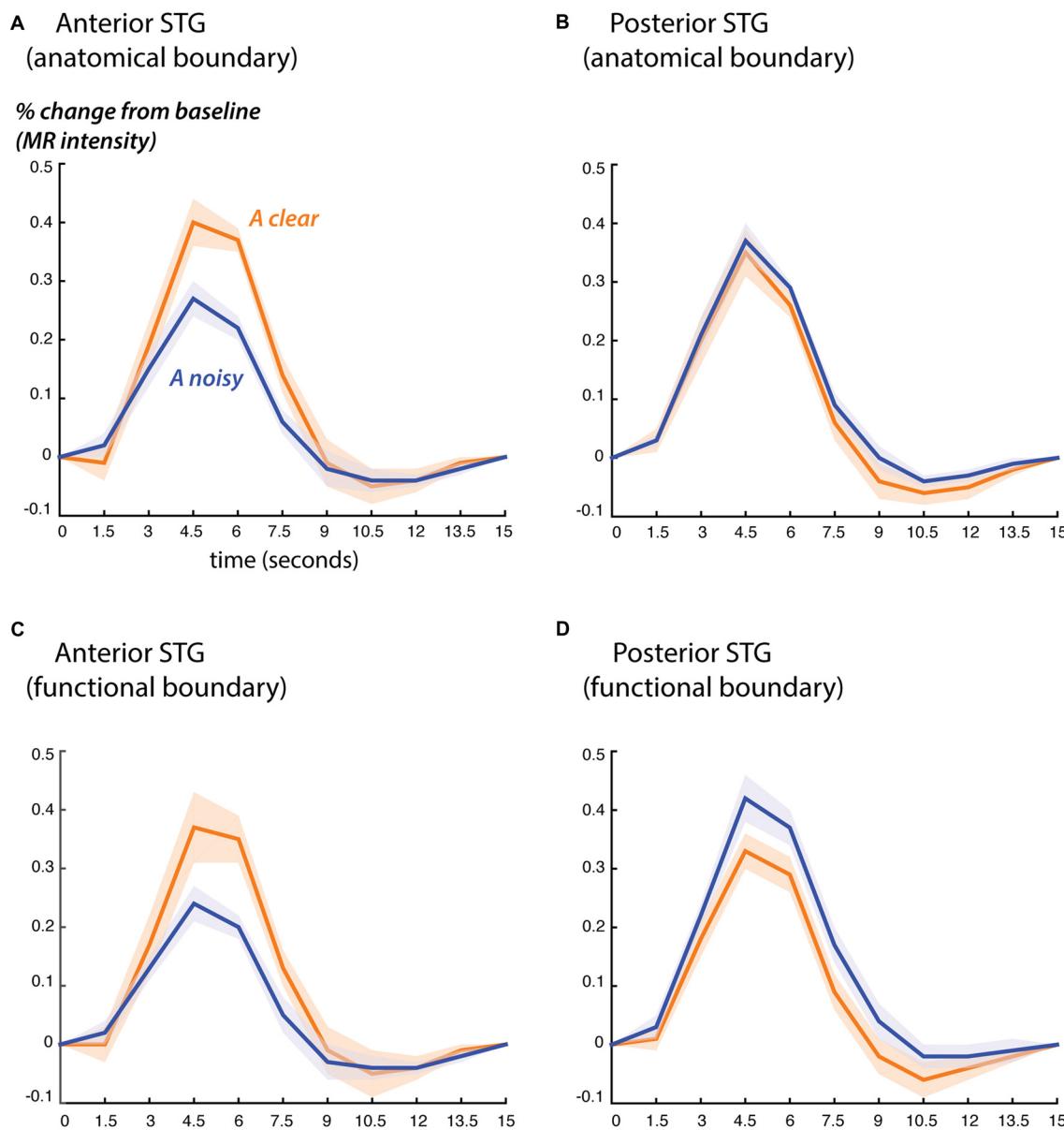


FIGURE 3 | Blood oxygen level dependent (BOLD) fMRI responses to clear and noisy audiovisual speech in anterior and posterior STG. **(A)** The average fMRI response for anterior STG was created by selecting all voxels in each hemisphere that were located from 0 mm to 30 mm anterior to the anatomical boundary defined by Heschl's gyrus and that showed a significant response to any stimulus. Responses shown for audiovisual speech with a clear auditory component (blue) and a noisy auditory component (orange). Lines show mean, shaded regions show SEM across participants. **(B)** The average fMRI response for posterior STG was created by selecting all responsive voxels in each hemisphere that were located from 0 mm to 15 mm posterior to the anatomical boundary defined by Heschl's gyrus. **(C)** Average BOLD fMRI response in the anterior STG, defined as all responsive voxels located from 0 mm to 30 mm anterior to the functional boundaries defined as shown in **Figure 2**. **(D)** BOLD fMRI response in the posterior STG, defined as all responsive voxels located from 0 mm to 15 mm posterior to the functional boundary.

difference between the anatomical and functional boundaries, resulting in a small mean distance between them ($y = -28$ vs. -30 , $t_{11} = 0.2$, $p = 0.8$).

The location of the anatomical and functional boundaries in the fMRI participants were similar to that of the anatomical boundary in the ECoG participants, located at $y = -27$ mm (± 2 mm SD); the 1 cm spacing of the ECoG electrodes did not allow a separate estimate of the functional boundary.

As in the ECoG data, the fMRI transition between clear and noisy speech preferring cortex happened over a short cortical distance. For instance, in participant OD's left hemisphere, the t -statistic of the clear vs. noisy contrast changed from $t = +5$ to $t = -2$ within 10 mm.

Next, we examined the fMRI response profiles on either side of the anatomical and functional boundaries (Figure 3).

We classified the STG from 0 mm to 30 mm anterior to each boundary as “anterior” and the STG from 0 mm to 15 mm posterior to each boundary as “posterior”. These values were chosen for consistency with the ECoG electrode locations, which ranged from 30 mm anterior to the boundary to 15 mm posterior to it.

Both the anatomical and functionally-defined STG ROIs showed the characteristic BOLD response, with a positive peak between 4 s and 6 s and a negative post-undershoot at the 10.5 s time point. The responses were robust, with a peak between 0.2% and 0.4% for a single audiovisual word. While the anterior STG responded more strongly to clear speech, the posterior STG did not. **Table 3** shows the results of the LME on the fMRI response amplitudes using an anatomically-defined border between anterior and posterior STG. The same three significant effects were observed as in the ECoG LME: a main effect of ROI location driven by smaller responses in the posterior STG ($p = 0.01$, magnitude 0.07%); a main effect of auditory noise driven by weaker responses to noisy auditory stimuli ($p = 10^{-6}$, magnitude 0.14%); and an interaction between these factors driven by larger responses to noisy auditory stimuli in posterior STG ($p = 10^{-4}$, magnitude 0.16%). The same effects were observed in an LME using the functionally-defined border between anterior and posterior STG (**Table 4**) with a significant main effect of auditory noise and interaction between auditory noise and location. However, the functionally-defined LME should be interpreted with caution as the border definition incorporated fMRI data, potentially biasing the model.

Behavioral Data

Participants’ ability to identify the word presented in each trial was at ceiling for auditory-clear stimuli (mean 99%) and lower for auditory-noisy stimuli (mean 78%). This performance difference could drive differences in brain responses, if, for instance, error monitoring circuits were more engaged during poorer performance. To address this concern, we compared responses across stimulus conditions with identical auditory stimuli but differing performance. Accuracy for noisy auditory speech was better when paired with clear visual speech than when paired with blurry visual speech (87% vs. 69%, $p = 0.004$). A performance explanation predicts differing brain responses to these two types of stimuli (due to differing performance) while an auditory stimulus explanation predicts no difference (since the two stimuli have identical auditory components of the stimuli). LMEs on the brain responses did not show a significant difference between these conditions ($p = 0.11$ for ECoG, $p = 0.31$ for fMRI) suggesting that performance differences do not account for the observed differences in brain responses to clear and noisy auditory speech.

Comparison Between fMRI and ECoG

BOLD fMRI and ECoG provide different measures of neural activity. BOLD fMRI is an indirect measure, with a time scale of seconds and an amplitude scale of image brightness increases relative to baseline. ECoG is a direct measure of neural activity, with a time scale of milliseconds and an amplitude scale of spectral power increases relative to baseline. To compare the fMRI and ECoG responses, we accounted for these differences,

TABLE 3 | LME model of the fMRI response amplitude in superior temporal gyrus (STG) regions defined by anatomical boundary.

Fixed effects	Estimate	Std. Error	DF	t-value	p-value
Baseline	0.40	0.03	35.9	15.4	$<10^{-16}$
Auditory noise (An)	-0.14	0.03	84	-5.4	10^{-6}
Posterior location × An	0.16	0.04	84	4.2	10^{-4}
Posterior location	-0.07	0.03	84	-2.6	0.01
Visual blur (Vb)	-0.03	0.03	84	-1.0	0.31
Posterior location × Vb	-0.02	0.04	84	-0.6	0.52
Posterior location × An × Vb	0.01	0.05	84	0.2	0.83
An × Vb	0.00	0.04	84	0.0	0.97

Results of an LME model of the fMRI response amplitude in STG regions defined by an anatomical boundary. The fixed effects were the location of each region (Anterior vs. Posterior), the presence or absence of auditory noise (An) in the stimulus and the presence or absence of visual blur (Vb) in the stimulus. STG ROIs from right and left hemisphere across six subjects were included in the model as random factor. For each effect, the model estimate (in units of % signal change) for that factor is shown relative to baseline, the response in anterior STG ROI to clear audiovisual speech. The baseline is shown first, all other effects are ranked by absolute t-value. Significant effects are shown in bold. The significance of the baseline fixed effect is italicized because it was prespecified.

TABLE 4 | LME model of the fMRI response amplitude in STG regions defined by functional boundary.

Fixed effects	Estimate	Std. Error	DF	t-value	p-value
Baseline	0.37	0.03	38.32	10.8	10^{-13}
Posterior location × An	0.21	0.05	84	4.1	10^{-4}
Auditory noise (An)	-0.14	0.04	84	-3.8	10^{-4}
Posterior location	-0.04	0.04	84	-1.1	0.28
Visual blur (Vb)	-0.02	0.04	84	-0.7	0.49
Posterior location × Vb	-0.02	0.05	84	-0.4	0.66
Posterior location × An × Vb	0.03	0.07	84	0.4	0.69
An × Vb	-0.01	0.05	84	-0.1	0.92

Results of an LME model of the fMRI response amplitude in STG regions defined by the functionally-defined boundary between anterior and posterior STG. Significant effects are shown in bold. The significance of the baseline fixed effect is italicized because it was prespecified.

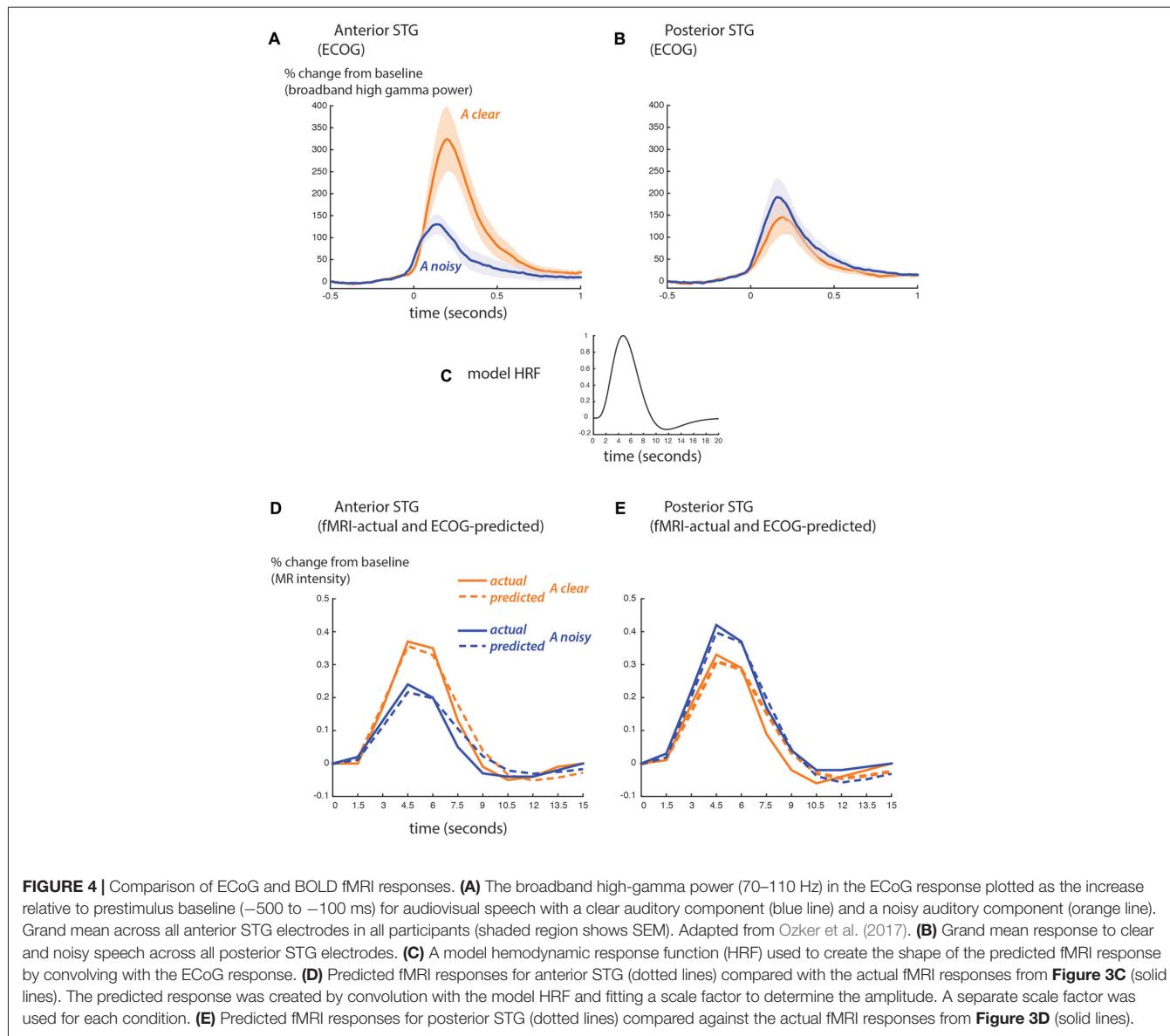


FIGURE 4 | Comparison of ECoG and BOLD fMRI responses. **(A)** The broadband high-gamma power (70–110 Hz) in the ECoG response plotted as the increase relative to prestimulus baseline (−500 to −100 ms) for audiovisual speech with a clear auditory component (blue line) and a noisy auditory component (orange line). Grand mean across all anterior STG electrodes in all participants (shaded region shows SEM). Adapted from Ozker et al. (2017). **(B)** Grand mean response to clear and noisy speech across all posterior STG electrodes. **(C)** A model hemodynamic response function (HRF) used to create the shape of the predicted fMRI response by convolving with the ECoG response. **(D)** Predicted fMRI responses for anterior STG (dotted lines) compared with the actual fMRI responses from **Figure 3C** (solid lines). The predicted response was created by convolution with the model HRF and fitting a scale factor to determine the amplitude. A separate scale factor was used for each condition. **(E)** Predicted fMRI responses for posterior STG (dotted lines) compared against the actual fMRI responses from **Figure 3D** (solid lines).

beginning with the time scale. **Figure 4** shows the ECoG responses from the STG, reprinted from Figure 4 of Ozker et al. (2017). Increases in the high-gamma broadband signal began less than 100 ms after auditory speech onset, and peaked at about 200 ms after auditory speech onset. To convert the directly-recorded neural activity measured with ECoG to the indirect and much slower measure of neural activity provided by BOLD fMRI, the ECoG responses were convolved with a standard HRF (**Figure 4C**) and downsampled from 1 ms resolution to a temporal resolution of 1.5 s, the repetition time (TR) of the fMRI data. This created a predicted fMRI response (based on the measured ECoG responses) on the same time scale and time base as the actual fMRI response. The second obstacle was the different amplitude scales of the responses. ECoG amplitude is measured in % change in the high-gamma broadband signal relative to pre-stimulus fixation baseline, while fMRI signal

amplitude is measured in % intensity increase of the EPI images relative to fixation baseline. A separate scale factor was calculated for each condition in order to generate the best fit between the predicted and actual fMRI responses.

Figure 4D shows the predicted-from-ECoG responses and actual fMRI responses based on the functional boundary between anterior and posterior STG. The shape of the responses was similar, as demonstrated by a high correlation coefficient between predicted and actual responses (anterior STG: 0.98 for auditory-clear, 0.96 for auditory-noisy; posterior STG: 0.97 for auditory-clear, 0.99 for auditory-noisy). The average across scale factor across conditions for the amplitude conversion was 612 ECoG% per BOLD%, meaning that a peak ECoG response of 612% was equivalent to a BOLD fMRI response of 1%. The scale factors were identical for auditory-clear and auditory-noisy conditions in posterior STG (476) but were markedly

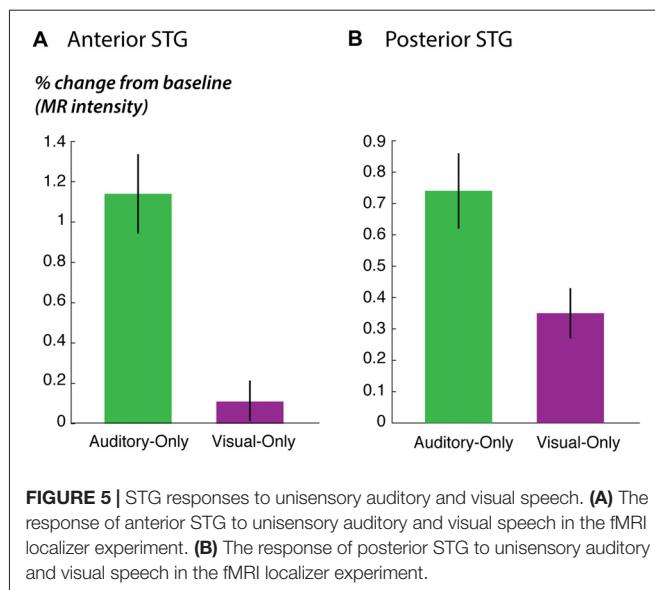


FIGURE 5 | STG responses to unisensory auditory and visual speech. **(A)** The response of anterior STG to unisensory auditory and visual speech in the fMRI localizer experiment. **(B)** The response of posterior STG to unisensory auditory and visual speech in the fMRI localizer experiment.

higher in anterior STG, especially for auditory-clear audiovisual words (909 for auditory-clear and 588 for auditory-noisy). This reflected the fact that in the ECoG data, the anterior STG response to clear speech was more than twice as large as the response to auditory-noisy speech (300% vs. 110%) while in fMRI, the anterior STG responded more strongly to clear speech than to noisy speech but the difference was less pronounced (0.37% vs. 0.24%).

In ECoG and fMRI, we observed distinct patterns of responses to clear and noisy speech in anterior and posterior STG. A possible explanation for these results is that anterior STG is unisensory auditory cortex, rendering it susceptible to auditory noise added to speech, while posterior STG is multisensory auditory-visual cortex, allowing it to compensate for auditory noise using visual speech information. This explanation predicts that posterior STG should show stronger responses to visual speech than anterior STG. To test this explanation, we took advantage of the fact that the fMRI participants viewed additional stimuli consisting of short stories presented in unisensory visual, unisensory auditory and audiovisual versions. The response to these stories was calculated in STG ROIs defined using the functional boundary (Figure 5). This analysis was unbiased because the functional boundary was created using fMRI data from auditory-clear and auditory-noisy words, completely independent of the story stimuli.

As predicted, the response to the unisensory visual story stimuli was significantly stronger in posterior STG than in anterior STG (posterior vs. anterior: 0.4% vs. 0.1%, $p = 0.02$ for visual speech). This could not be explained by an overall difference in responsiveness; in fact, for the other story stimuli, there was a trend towards weaker responses in posterior STG (posterior vs. anterior: 0.7% vs. 1.1%, $p = 0.09$ for unisensory auditory speech; 1.3% vs. 1.7%; $p = 0.07$ for audiovisual speech), consistent with the weaker responses in posterior STG to single audiovisual words observed in the main experiment (effect of posterior location in Table 3).

DISCUSSION

We measured neural activity in the human STG using two very different techniques: directly, using surface electrodes implanted in ECoG participants with epilepsy, or indirectly, using the BOLD response in fMRI participants who were healthy controls. Both ECoG and fMRI participants viewed the same clear and noisy audiovisual speech stimuli and performed the same speech recognition task. Both techniques demonstrated a sharp functional boundary in the STG. Cortex responded more strongly to clear audiovisual speech on the anterior side of the boundary while on the posterior side of the boundary it did not. For both techniques, the boundary was located at a similar location in standard space ($y = 30$ mm) and the transition between the two functional zones happened within 10 mm of anterior-to-posterior distance along the STG.

In both fMRI and ECoG patients, an anatomical boundary set at the most posterior point of Heschl's gyrus provided a reasonable proxy for the functional boundary. This is important because unlike the fMRI or ECoG data needed to locate the functional boundary, the structural MRI scan needed to locate the anatomical boundary is easily obtainable (for instance, in the examination of patients with brain lesions). While primary visual and auditory cortex are easily localizable using anatomical landmarks, it has proven to be much more of a challenge to find landmarks for association areas (Weiner and Grill-Spector, 2011; Witthoft et al., 2014).

Multisensory integration provides a conceptual framework for understanding these results. When noisy auditory speech is presented, auditory information alone is insufficient for perception, and auditory-speech regions in anterior STG respond with diminished intensity. Visual speech information can compensate for noisy auditory speech (Sumby and Pollack, 1954; Bernstein et al., 2004; Ross et al., 2007), but this requires recruitment of multisensory brain regions capable of combining the auditory and visual speech information to restore intelligibility. While both anterior and posterior STG responded to audiovisual speech, data from the fMRI localizer experiment showed that posterior STG responded more strongly to visual-only speech than anterior STG, supporting the idea that posterior STG is a multisensory area capable of combining auditory and visual speech.

The neural code in posterior STG is hinted at by a recent study, which found that a region of posterior STG and STS (similar to the posterior STG region described in the present manuscript) responded more strongly to silent videos of faces making mouth movements compared to silent videos of faces making eye movements (Zhu and Beauchamp, 2017). The same region responded strongly to unisensory auditory speech, with a greater amplitude for vocal than non-vocal sounds. Interestingly, as statistical thresholds were increased to select voxels with a greater preference for visual mouth movements, response to unisensory auditory speech increased, suggesting that at a single voxel level, small populations of neurons code for mouth movements and speech sounds,

the two components of audiovisual speech (Bernstein et al., 2011). This cross-modal correspondence in neural coding of multisensory cues is exactly as predicted by computational models of multisensory integration (Beck et al., 2008; Magnotti and Beauchamp, 2017).

There is a substantial body of evidence showing that posterior STS is a cortical hub for multisensory integration, responding to both auditory and visual stimuli including faces and voices, letters and voices and recordings and videos of objects (Calvert et al., 2000; Foxe et al., 2002; Beauchamp et al., 2004; van Atteveldt et al., 2004; Miller and D'Esposito, 2005; Reale et al., 2007). The finding that the adjacent cortex in posterior STG is also important for multisensory integration has several ramifications. In a transcranial magnetic stimulation (TMS) study, integration of auditory and visual speech (as indexed by the McGurk effect) was disrupted with TMS targeted at the posterior STS (Beauchamp et al., 2010). The present results suggest that posterior STG may also have played a role in the observed disruption, and raise the possibility that electrical brain stimulation of STG in ECoG patients can increase our understanding of multisensory speech perception as it has for visual perception (Murphrey et al., 2008; Rangarajan and Parvizi, 2016).

While there have been numerous previous fMRI studies of noisy and clear audiovisual speech (e.g., Callan et al., 2003; Sekiyama et al., 2003; Bishop and Miller, 2009; Stevenson and James, 2009; Lee and Noppeney, 2011; McGettigan et al., 2012), none described a sharp boundary in the response patterns to clear and noisy speech within the STG. A likely explanation is that many of the previous studies use spatial filtering or blurring as a preprocessing step in their fMRI data analysis pipeline and reported only group average data, which introduces additional blurring due to inter-subject anatomical differences, especially for commonly-used volume-based templates. Combined, these two spatial blurring steps could easily eliminate sharp boundaries present in fMRI data. For instance, blurring eliminates the otherwise robust observation of functional specialization for different object categories in visual cortex (Tyler et al., 2003). Another possible explanation for the failure of previous studies to observe the boundary is the common practice of reporting responses only at the location of activation peaks, rather than examining the entire extent of the activation. Anterior and posterior STG form a continuous region of active cortex, with the strongest activation in anterior STG. Therefore, only reporting responses from a single peak STG location (which would almost certainly fall in anterior STG) would camouflage the very different pattern of activity in posterior STG.

Implications for ECoG and fMRI

While the primary goal of our study was not a comparison of the two methodologies, there was good correspondence between the actual fMRI signal and the fMRI signal predicted from our measure of ECoG amplitude, the broadband high-gamma response in the window from 70 Hz to 110 Hz. This is consistent with mounting evidence that the high-frequency broadband signal in ECoG is a good match for the fMRI signal

(reviewed in Ojemann et al., 2013). Other ECoG measures, such as the narrowband gamma response (30–80 Hz) or the narrowband alpha response, may characterize neuronal synchrony rather than level of neuronal activity, and hence are poorly correlated with the BOLD signal (Hermes et al., 2017).

A reassuring finding from the present study is that we observed similar patterns of responses between ECoG patients with epilepsy and healthy controls viewing the same stimuli and performing the same task. This provides data to partially mitigate persistent concerns that ECoG patients may have different brain organization than healthy controls, reducing the generalizability of the results of ECoG studies. A related concern is the small sample size typical of many ECoG studies. Our ECoG dataset compared anterior and posterior STG responses in five individual hemispheres. Our fMRI dataset more than doubled this sample size (to 12 hemispheres) and the continuous sampling of the fMRI voxel grid provided more statistical power to identify the location of the anterior-posterior border. Increasing the sample size would allow additional characterization of individual variability in the anterior-posterior border.

One minor discrepancy between the ECoG and fMRI results was a larger relative amplitude for the favored stimuli in ECoG. For instance, anterior STG showed a nearly three-fold difference in the response amplitude to clear vs. noisy audiovisual speech (300% vs. 110%). The difference in fMRI was in the same direction but much smaller (0.37% vs. 0.24%). We attribute this to the ability of ECoG electrodes to sample small populations of highly-selective neurons, while the BOLD fMRI response spatially sums over larger populations of neurons, mixing more and less selective signals. This same pattern has been observed in other studies comparing fMRI with ECoG. For instance, in a study of the fusiform face area, the BOLD signal evoked by faces was approximately double that evoked by non-face objects while the broadband high-gamma amplitude was triple or more for the same contrast (Parvizi et al., 2012).

AUTHOR CONTRIBUTIONS

MO and MSB designed and conducted the experiments, analyzed the data and wrote the manuscript. DY assisted with ECoG experiments.

FUNDING

This work was funded by Veterans Administration Clinical Science Research and Development Merit Award Number 1I01CX000325-01A1 and National Institutes of Health (NIH) R01NS065395.

ACKNOWLEDGMENTS

We thank the Core for Advanced MRI (CAMRI) and Johannes Rennig for assistance with fMRI data collection and Ping Sun, DY and Bill Bosking for assistance with ECoG.

REFERENCES

- Beauchamp, M. S., Lee, K. E., Argall, B. D., and Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41, 809–823. doi: 10.1016/s0896-6273(04)00070-4
- Beauchamp, M. S., Nath, A. R., and Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J. Neurosci.* 30, 2414–2417. doi: 10.1523/JNEUROSCI.4865-09.2010
- Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., et al. (2008). Probabilistic population codes for Bayesian decision making. *Neuron* 60, 1142–1152. doi: 10.1016/j.neuron.2008.09.021
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312. doi: 10.1038/35002078
- Bernstein, L. E., Auer, E. T., and Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Commun.* 44, 5–18. doi: 10.1016/j.specom.2004.10.011
- Bernstein, L. E., Jiang, J., Pantazis, D., Lu, Z. L., and Joshi, A. (2011). Visual phonetic processing localized using speech and nonspeech face gestures in video and point-light displays. *Hum. Brain Mapp.* 32, 1660–1676. doi: 10.1002/hbm.21139
- Bishop, C. W., and Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *J. Cogn. Neurosci.* 21, 1790–1805. doi: 10.1162/jocn.2009.21118
- Callan, D. E., Jones, J. A., Munhall, K., Callan, A. M., Kroos, C., and Vatikiotis-Bateson, E. (2003). Neural processes underlying perceptual enhancement by visual speech gestures. *Neuroreport* 14, 2213–2218. doi: 10.1097/00001756-200312020-00016
- Calvert, G. A., Campbell, R., and Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr. Biol.* 10, 649–657. doi: 10.1016/s0960-9822(00)00513-3
- Chang, E. F., Rieger, J. W., Johnson, K., Berger, M. S., Barbaro, N. M., and Knight, R. T. (2010). Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* 13, 1428–1432. doi: 10.1038/nn.2641
- Cheng, K., Waggoner, R. A., and Tanaka, K. (2001). Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron* 32, 359–374. doi: 10.1016/s0896-6273(01)00477-9
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014
- Crosse, M. J., Di Liberto, G. M., and Lalor, E. C. (2016). Eye can hear clearly now: inverse effectiveness in natural audiovisual speech processing relies on long-term crossmodal temporal integration. *J. Neurosci.* 36, 9888–9895. doi: 10.1523/JNEUROSCI.1396-16.2016
- Dale, A. M., Greve, D. N., and Burock, M. A. (1999). “Optimal stimulus sequences for event-related fMRI,” in *5th International Conference on Functional Mapping of the Human Brain* (Duesseldorf, Germany).
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010
- Fischl, B., Liu, A., and Dale, A. M. (2001). Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* 20, 70–80. doi: 10.1109/42.906426
- Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J. Neurophysiol.* 88, 540–543. doi: 10.1152/jn.2002.88.1.540
- Grill-Spector, K., and Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677. doi: 10.1146/annurev.neuro.27.070203.144220
- Hermes, D., Nguyen, M., and Winawer, J. (2017). Neuronal synchrony and the relation between the blood-oxygen-level dependent response and the local field potential. *PLoS Biol.* 15:e2001461. doi: 10.1371/journal.pbio.2001461
- Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., and Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* 22, 324–333. doi: 10.1097/00004728-199803000-00032
- Janszky, J., Jokeit, H., Heinemann, D., Schulz, R., Woermann, F. G., and Ebner, A. (2003). Epileptic activity influences the speech organization in medial temporal lobe epilepsy. *Brain* 126, 2043–2051. doi: 10.1093/brain/awg193
- Kramer, M. A., and Cash, S. S. (2012). Epilepsy as a disorder of cortical network organization. *Neuroscientist* 18, 360–372. doi: 10.1177/1073858411422754
- Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). *LmerTest: Tests in Linear Mixed Effects Models*. Available online at: <http://CRAN.R-project.org/Package/lmerTest>. R package version: 2.0–29.
- Lachaux, J. P., Axmacher, N., Mormann, F., Halgren, E., and Crone, N. E. (2012). High-frequency neural activity and human cognition: past, present and possible future of intracranial EEG research. *Prog. Neurobiol.* 98, 279–301. doi: 10.1016/j.pneurobio.2012.06.008
- Leaver, A. M., and Rauschecker, J. P. (2016). Functional topography of human auditory cortex. *J. Neurosci.* 36, 1416–1428. doi: 10.1523/JNEUROSCI.0226-15.2016
- Lee, H., and Noppeney, U. (2011). Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. *J. Neurosci.* 31, 11338–11350. doi: 10.1523/JNEUROSCI.6510-10.2011
- Lindquist, M. A., Meng Loh, J., Atlas, L. Y., and Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage* 45, S187–S198. doi: 10.1016/j.neuroimage.2008.10.065
- Magnotti, J. F., and Beauchamp, M. S. (2017). A causal inference model explains perception of the mcgurk effect and other incongruent audiovisual speech. *PLoS Comput. Biol.* 13:e1005229. doi: 10.1371/journal.pcbi.1005229
- McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., and Scott, S. K. (2012). Speech comprehension aided by multiple modalities: behavioural and neural interactions. *Neuropsychologia* 50, 762–776. doi: 10.1016/j.neuropsychologia.2012.01.010
- Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994
- Miller, L. M., and D’Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.* 25, 5884–5893. doi: 10.1523/JNEUROSCI.0896-05.2005
- Moerel, M., De Martino, F., and Formisano, E. (2014). An anatomical and functional topography of human auditory cortical areas. *Front. Neurosci.* 8:225. doi: 10.3389/fnins.2014.00225
- Murphy, D. K., Yoshor, D., and Beauchamp, M. S. (2008). Perception matches selectivity in the human anterior color center. *Curr. Biol.* 18, 216–220. doi: 10.1016/j.cub.2008.01.013
- Nath, A. R., and Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage* 59, 781–787. doi: 10.1016/j.neuroimage.2011.07.024
- Nath, A. R., Fava, E. E., and Beauchamp, M. S. (2011). Neural correlates of interindividual differences in children’s audiovisual speech perception. *J. Neurosci.* 31, 13963–13971. doi: 10.1523/JNEUROSCI.2605-11.2011
- Ojemann, G. A., Ojemann, J., and Ramsey, N. F. (2013). Relation between functional magnetic resonance imaging (fMRI) and single neuron, local field potential (LFP) and electrocorticography (ECOG) activity in human cortex. *Front. Hum. Neurosci.* 7:34. doi: 10.3389/fnhum.2013.00034
- Ozker, M., Schepers, I. M., Magnotti, J. F., Yoshor, D., and Beauchamp, M. S. (2017). A double dissociation between anterior and posterior superior temporal gyrus for processing audiovisual speech demonstrated by electrocorticography. *J. Cogn. Neurosci.* 29, 1044–1060. doi: 10.1162/jocn_a_01110
- Parvizi, J., Jacques, C., Foster, B. L., Witthoft, N., Rangarajan, V., Weiner, K. S., et al. (2012). Electrical stimulation of human fusiform face-selective regions distorts face perception. *J. Neurosci.* 32, 14915–14920. doi: 10.1523/JNEUROSCI.2609-12.2012
- Rangarajan, V., and Parvizi, J. (2016). Functional asymmetry between the left and right human fusiform gyrus explored through electrical brain stimulation. *Neuropsychologia* 83, 29–36. doi: 10.1016/j.neuropsychologia.2015.08.003
- Ray, S., and Maunsell, J. H. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* 9:e1000610. doi: 10.1371/journal.pbio.1000610

- Reale, R. A., Calvert, G. A., Thesen, T., Jenison, R. L., Kawasaki, H., Oya, H., et al. (2007). Auditory-visual processing represented in the human superior temporal gyrus. *Neuroscience* 145, 162–184. doi: 10.1016/j.neuroscience.2006.11.036
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., and Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* 17, 1147–1153. doi: 10.1093/cercor/bhl024
- Salmelin, R. (2007). Clinical neurophysiology of language: the MEG approach. *Clin. Neurophysiol.* 118, 237–254. doi: 10.1016/j.clinph.2006.07.316
- Sekiyama, K., Kanno, I., Miura, S., and Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neurosci. Res.* 47, 277–287. doi: 10.1016/s0168-0102(03)00214-1
- Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., and Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn. Reson. Med.* 67, 1210–1224. doi: 10.1002/mrm.23097
- Shahin, A. J., Kerlin, J. R., Bhat, J., and Miller, L. M. (2012). Neural restoration of degraded audiovisual speech. *Neuroimage* 60, 530–538. doi: 10.1016/j.neuroimage.2011.11.097
- Sohoglu, E., and Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proc. Natl. Acad. Sci. U S A* 113, E1747–E1756. doi: 10.1073/pnas.1523266113
- Stevenson, R. A., and James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage* 44, 1210–1223. doi: 10.1016/j.neuroimage.2008.09.034
- Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- Tang, C., Hamilton, L. S., and Chang, E. F. (2017). Intonational speech prosody encoding in the human auditory cortex. *Science* 357, 797–801. doi: 10.1126/science.aam8577
- Tyler, L. K., Bright, P., Dick, E., Tavares, P., Pilgrim, L., Fletcher, P., et al. (2003). Do semantic categories activate distinct cortical regions? Evidence for a distributed neural semantic system. *Cogn. Neuropsychol.* 20, 541–559. doi: 10.1080/02643290244000211
- van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282. doi: 10.1016/j.neuron.2004.06.025
- Weiner, K. S., and Grill-Spector, K. (2011). Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. *Neuroimage* 56, 2183–2199. doi: 10.1016/j.neuroimage.2011.03.041
- Witthoft, N., Nguyen, M. L., Golarai, G., LaRocque, K. F., Liberman, A., Smith, M. E., et al. (2014). Where is human V4? Predicting the location of hV4 and VO1 from cortical folding. *Cereb. Cortex* 24, 2401–2408. doi: 10.1093/cercor/bht092
- Zhu, L. L., and Beauchamp, M. S. (2017). Mouth and voice: a relationship between visual and auditory preference in the human superior temporal sulcus. *J. Neurosci.* 37, 2697–2708. doi: 10.1523/JNEUROSCI.2914-16.2017

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ozker, Yoshor and Beauchamp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.