



Published in final edited form as:

Hear Res. 2016 March ; 333: 108–117. doi:10.1016/j.heares.2015.12.008.

## Acoustic richness modulates the neural networks supporting intelligible speech processing

Yune-Sang Lee<sup>1</sup>, Nam Eun Min<sup>1</sup>, Arthur Wingfield<sup>2</sup>, Murray Grossman<sup>1</sup>, and Jonathan E. Peelle<sup>3</sup>

<sup>1</sup>Department of Neurology and Penn Frontotemporal Degeneration Center, University of Pennsylvania, Philadelphia PA USA

<sup>2</sup>Volen National Center for Complex Systems, Brandeis University, Waltham MA USA

<sup>3</sup>Department of Otolaryngology, Washington University in St. Louis, St. Louis MO USA

### Abstract

The information contained in a sensory signal plays a critical role in determining what neural processes are engaged. Here we used interleaved silent steady-state (ISSS) functional magnetic resonance imaging (fMRI) to explore how human listeners cope with different degrees of acoustic richness during auditory sentence comprehension. Twenty-six healthy young adults underwent scanning while hearing sentences that varied in acoustic richness (high vs. low spectral detail) and syntactic complexity (subject-relative vs. object-relative center-embedded clause structures). We manipulated acoustic richness by presenting the stimuli as unprocessed full-spectrum speech, or noise-vocoded with 24 channels. Importantly, although the vocoded sentences were spectrally impoverished, all sentences were highly intelligible. These manipulations allowed us to test how intelligible speech processing was affected by orthogonal linguistic and acoustic demands. Acoustically rich speech showed stronger activation than acoustically less-detailed speech in a bilateral temporoparietal network with more pronounced activity in the right hemisphere. By contrast, listening to sentences with greater syntactic complexity resulted in increased activation of a left-lateralized network including left posterior lateral temporal cortex, left inferior frontal gyrus, and left dorsolateral prefrontal cortex. Significant interactions between acoustic richness and syntactic complexity occurred in left supramarginal gyrus, right superior temporal gyrus, and right inferior frontal gyrus, indicating that the regions recruited for syntactic challenge differed as a function of acoustic properties of the speech. Our findings suggest that the neural systems involved in speech perception are finely tuned to the type of information available, and that reducing the richness of the acoustic signal dramatically alters the brain's response to spoken language, even when intelligibility is high.

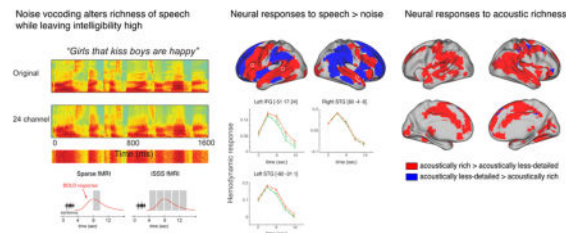
---

Please address correspondence to: Yune-Sang Lee, PhD, Department of Neurology, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, yslee@mail.med.upenn.edu. Jonathan Peelle, PhD, Department of Otolaryngology, Washington University in St. Louis, Saint Louis, MO 63110, peellej@ent.wustl.edu.

Conflicts of interest: The authors declare no competing financial interests.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Graphical Abstract



## Keywords

speech; acoustic; fMRI; language; executive function; listening effort; vocoding; hearing

## 1. Introduction

During everyday communication, the acoustic richness of speech sounds is commonly affected by many factors including background noise, competing talkers, or hearing impairment. Ordinarily, one might expect that when a speech input is lacking in sensory detail, greater processing resources would be needed for successful recognition of that signal (Rönnberg et al., 2013). Less certain, however, is the effect on neural activity when two intelligible speech signals are presented, but with one signal lacking in spectral detail—conceptually similar to what might be heard with a hearing aid or cochlear implant.

The acoustic quality of the speech signal has been of longstanding interest because acoustic details help convey paralinguistic information such as talker sex, age, or emotion (Gobl and Chasaide, 2003), as well as prosodic cues that can aid in spoken communication. We use the term acoustic richness instead of vocal quality to emphasize that changes to acoustic detail of the speech signal can arise from many sources. Although many behavioral studies have assessed speech perception by systematically manipulating voice quality (Chen and Loizou, 2011, 2010; Churchill et al., 2014; Loizou, 2006; Maryn et al., 2009), relatively little neuroimaging research has investigated the neural consequence of acoustic richness in intelligible speech. Here we examine how acoustic clarity affects the neural processing of intelligible speech. We focus on sentence comprehension, where the acoustic richness of the speech might interact with computational demands at the linguistic level.

Neuroanatomically, speech comprehension is supported in large part by a core network centered in bilateral temporal cortex (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009), frequently complemented by the left inferior frontal gyrus during sentence processing (Adank, 2012; Peelle, 2012). These regions are more active when listening to intelligible sentences than when hearing a variety of less intelligible control conditions (Davis and Johnsrude, 2003; Evans et al., 2014; Obleser et al., 2007; Rodd et al., 2012; Scott et al., 2000). There is increasing evidence that as speech is degraded to the point that its intelligibility is compromised listeners rely on regions outside of this core speech network, particularly in frontal cortex. Regions of increased activity during degraded speech processing include the cingulo-opercular network (Eckert et al., 2009; Erb et al., 2013;

Vaden et al., 2013; Wild et al., 2012) and premotor cortex (Davis and Johnsrude, 2003; Hervais-Adelman et al., 2012). The fact that these regions are more active for degraded speech than for acoustically rich speech suggests that listeners are recruiting additional cognitive resources to compensate for the loss of acoustic detail.

In these and related studies, however, acoustic richness and intelligibility are frequently correlated, such that the degraded speech has also been less intelligible. The relationship between intelligibility and acoustic richness makes it impossible to disentangle changes in neural processing due to reduced intelligibility from changes due to reduced acoustic richness. To address this issue, in the current study we used 24 channel vocoded speech that reduced the spectral detail of speech while allowing for good intelligibility. We refer to these stimuli as acoustically less-detailed speech because of the reduction of spectral resolution, compared to the acoustically rich original signal. Furthermore, to determine how resource demands for cognitive and auditory processes interact, we independently manipulated linguistic challenge by varying syntactic complexity. Because we have clear expectations for brain networks responding to syntactic challenge, including a syntactic manipulation also allowed us to validate the efficacy of our fMRI paradigm and data analysis approach.

One possibility is that, even when speech is intelligible, decreasing the acoustic richness of the speech signal would lead listeners to recruit a set of compensatory frontal networks. In this case we would expect increased frontal activity for acoustically less-detailed speech, which may be shared or different from that required to process syntactically complex material. An alternative possibility is that removing acoustic detail from otherwise intelligible speech would reduce the quality of the paralinguistic information (e.g., sex and age of the speaker) available to listeners, and thus limit the neural processing for non-linguistic information. In this case we would expect to observe reduced neural processing for acoustically less-detailed speech.

## 2. Material and Methods

### 2.1 Subjects

Twenty-six adults (age: 20–34 years, mean = 24.9 years; 12 females) were recruited from the University of Pennsylvania community. All reported themselves being right-handed native speakers of American English and in good health, with no history of neurological disorders or hearing difficulty. Based on pure tone audiometry, all participants' hearing acuity fell within a clinically normal range, with pure tone averages (1, 2, and 4 kHz) of 25 dB HL or less. Figure 1A shows individual audiometric profiles up to 8 kHz. All participants provided written consent as approved by the Human Subjects Institutional Review Board of the University of Pennsylvania and were paid for their participation.

### 2.2 Stimuli

Our experimental stimuli consisted of 96 6-word English sentences, half of which contained a subject-relative center-embedded clause and half an object-relative center-embedded clause (Peelle et al., 2010b, 2004). The syntactic manipulation was accomplished by rearranging word order, and thus lexical characteristics were identical across subject-relative

and object-relative sentences (e.g., subject-relative: “Kings that help queens are nice”; object-relative: “Kings that queens help are nice”). Each sentence contained a male and female character, but only one character performed an action in a particular sentence, allowing us to ask participants to name the gender of the character performing the action as a measure of comprehension. We used 24 base sentences, each of which appeared in 4 configurations (subject-relative with male or female actor, and object-relative with male or female actor).

Half of the sentences were presented unprocessed as originally recorded (acoustically rich speech), and half were noise-vocoded using 24 channels; the specific sentences presented as acoustically rich or vocoded speech were counterbalanced across participants. Noise vocoding allows the reduction of spectral detail while leaving temporal cues relatively intact (Shannon et al., 1995). Importantly for the current study, vocoding with 24 channels produces speech that is still fully intelligible. That is, word report is typically near-perfect (Faulkner et al., 2001), but lacking in spectral richness compared to normal speech. Our vocoding algorithm extracted the amplitude envelope by lowpass filtering half-wave rectified signal from each channel at 30 Hz. Sound files were lowpass filtered at 8 kHz and matched for root mean square (RMS) amplitude. The code for signal processing (*jp\_vocode.m*) is available from [https://github.com/jpeelle/jp\\_matlab](https://github.com/jpeelle/jp_matlab).

Together, our stimulus manipulations resulted in a  $2 \times 2$  factorial design, varying syntactic complexity (subject-relative, object-relative) and acoustic detail (acoustically rich, acoustically less-detailed). A subset of 24 experimental sentences was vocoded with a single channel to create an unintelligible control condition (Figure 1B).

### 2.3 MRI scanning

MRI data were collected using a 3 T Siemens Trio scanner (Siemens Medical System, Erlangen, Germany) equipped with an 8-channel head coil. Volume acquisition was angled approximately  $30^\circ$  away from the AC-PC line. Scanning began with acquisition of a T1-weighted structural volume using a magnetization prepared rapid acquisition gradient echo protocol [axial orientation, repetition time (TR) = 1620 ms, echo time (TE) = 3 ms, flip angle =  $15^\circ$ , field of view (FOV) =  $250 \times 188$  mm, matrix =  $256 \times 192$  mm, 160 slices, voxel resolution =  $0.98 \times 0.98 \times 1$  mm]. Subsequently, 4 runs of blood oxygenation level-dependent (BOLD) functional MRI scanning were performed (TR = 2000 ms, TE = 30 ms, flip angle =  $78^\circ$ , FOV =  $192 \times 192$  mm, matrix =  $64 \times 64$  mm, 32 slices, voxel resolution =  $3 \times 3 \times 3$  mm with 0.75 mm gap) using an interleaved silent steady state (ISSS) protocol (Schwarzbauer et al., 2006), in which 5 consecutive volumes were acquired in between 4 seconds of silence (Figure 1C). ISSS scanning is similar to standard sparse imaging (Hall et al., 1999) in that it allows presentation of auditory stimuli in relative quiet; however, with ISSS we are able to collect a greater number of images following each stimulus, maintaining steady-state longitudinal magnetization by continuing excitation pulses during the silent period. ISSS thus provides increased temporal resolution while avoiding the main problems of concurrent scanner noise during auditory presentation (Peelle, 2014). Here, each ISSS “trial” lasted 14 seconds: 4 seconds of relative quiet followed by 10 seconds (5 volumes) of data collection. Finally, a  $B_0$  mapping sequence was acquired at the end of the scanning (TR

= 1050 ms, TE = 4 ms, flip angle = 60°, FOV = 240 × 240 mm, matrix = 64 × 64 mm, 44 slices, slice thickness = 4 mm, voxel resolution = 3.8 × 3.8 × 4 mm).

## 2.4 Experimental Procedure

There were 4 fMRI runs per session. On each trial, participants were binaurally presented with either a spoken sentence (unprocessed or 24 channel vocoded) or unintelligible noise (1 channel vocoded speech) at a comfortable listening level, 1 second after the silent period of the ISSS began. Stimuli were presented using MRI-compatible high-fidelity insert earphones (Sensimetrics Model S14). For each sentence, participants were instructed to indicate the gender of the character performing the action via button press as quickly and accurately as possible. For the unintelligible noise stimuli, participants were told to press either of the buttons. Participants held the button box with both hands, using left and right hands for responses (an equal number of male and female responses were included in all runs). E-Prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA) was used to present stimuli and record accuracy and reaction times. In each run, there were 24 trials with spoken sentences (6 sentences × 2 syntactic constructions × 2 acoustic manipulations), 6 trials with 1 channel vocoded speech, and 6 trials of silence, for a total of 36 trials (8.4 min per run). The order of conditions within each run was randomized. Each of the 96 sentences was presented only once.

Prior to entering the scanner, participants received instructions and performed a practice session to ensure they understood the task. Once inside the scanner, but prior to scanning, participants confirmed intelligibility of 4 spoken sentences (two unprocessed and two 24 channel vocoded sentences) at the intensity to be used in the main experiment by correctly repeating each sentence as it was presented. These sentences were not included in the main stimulus set. All participants were able to repeat back the vocoded sentences accurately.

## 2.5 fMRI data analysis

Functional images for each participant were unwarped using the *prelude* and *flirt* routines from FSL version 5.0.5 (FMRIB Software Library, University of Oxford). The remaining preprocessing steps were performed using SPM12 (version 6225; Wellcome Trust Centre for Neuroimaging): images were realigned to the first image in the series, coregistered to each participant's structural image, normalized (with preserved voxel size) to MNI space using a transformation matrix generated during tissue class segmentation (Ashburner and Friston, 2005), and spatially smoothed with a 9 mm full-width half maximum (FWHM) Gaussian kernel. After preprocessing, the data were modeled using a finite impulse response (FIR) approach in which the response at each of the 5 time points following an event was separately estimated. We only analyzed trials that yielded correct behavioral responses. Additional regressors included 6 motion parameters and 4 run effects. High-pass filtering with a 128 sec cutoff was used to remove low-frequency noise. The typical first-order autoregressive modeling for temporal autocorrelation was turned off due to the discontinuous timeseries.

To assess main effects and their interaction, we performed a series of F tests using the parameter estimates for the 5 time points in a one-way repeated measures ANOVA. The FIR

analysis makes use of all time points collected, but does not indicate which condition showed a greater response. To interpret the direction of the effects identified by the F test, we used the integral of the positive portion of the response, or the summed positive response. That is, for each condition, we summed all non-negative parameter estimates, and used this single number as a measure of activation. The advantage of using the summed positive response is that it provides an indication of the overall direction of the effect without relying on assumptions about the shape or latency of the hemodynamic response. We also extracted the timecourse of activity for maxima to ensure the response shapes were physiologically plausible and consistent with the whole-brain comparisons. We then compared the timecourse of 4 sentence types using the time-to-peak measure using the Friedman nonparametric repeated measures ANOVA.

All whole-brain results were thresholded using a cluster-forming threshold of voxelwise  $p < .001$  (uncorrected) and a cluster-level threshold of  $p < .05$  (corrected) across the whole brain based on cluster extent and Gaussian random field theory (Friston et al., 1994; Worsley et al., 1992). The peak\_nii toolbox<sup>1</sup> was used as a complementary method of referring to the anatomical labeling of sub-peaks. Results were projected onto the Conte69 surface-based atlas using Connectome Workbench (Van Essen et al., 2012), or slices using MRICron<sup>2</sup> (Rorden and Brett, 2000). Unthresholded statistical maps are available at <http://neurovault.org/collections/571/> (Gorgolewski et al., 2015).

To assess the laterality of our main results we calculated the lateralization index (LI) for each main effect using the LI toolbox<sup>3</sup> (Wilke and Lidzba, 2007). We first computed LI indices in the whole brain followed by LI indices separately for the temporal, parietal, and frontal lobes. The LI toolbox uses an iterative bootstrapping procedure to avoid thresholding effects. We did not use a lateralization cutoff (i.e., we included the entire hemisphere instead of selecting  $> 11$  mm from the midline), clustering, or variance weighting. Positive LI values reflect left lateralization.

## 5. Results

### 5.1 Acoustic analysis

We computed several acoustic measures on all sentence stimuli using Praat<sup>4</sup> (v 5.4.08) including root mean square amplitude (RMS), harmonic-to-noise ratio (HNR), and long-term average spectrum (LTAS). These data are shown in Table 1.

Because we matched our stimuli on RMS, RMS was numerically equivalent across condition by design, and thus we only statistically tested differences in LTAS and HNR. We summarized the LTAS by summing across frequency bands to provide a single number per sentence.

---

<sup>1</sup>[https://www.nitrc.org/projects/peak\\_nii](https://www.nitrc.org/projects/peak_nii)

<sup>2</sup><http://www.mccauslandcenter.sc.edu/mricro/mricron/install.html>

<sup>3</sup><https://www.medizin.uni-tuebingen.de/kinder/en/research/neuroimaging/software/>

<sup>4</sup><http://praat.org>



We performed factorial ANOVA analyses on the acoustic measures across our four experimental conditions (2: syntactic complexity  $\times$  2: acoustic richness). For HNR, there was a significant main effect of clarity,  $F_{(1,188)} = 3729$ ,  $p < 0.05$ , but not of syntax  $F_{(1,188)} = 1.53$ ,  $p = 0.22$ ; the interaction was also not significant,  $F_{(1,188)} = 2.24$ ,  $p = 0.14$ . For LTAS, there was also a main effect of clarity,  $F_{(1,188)} = 7661$ ,  $p < 0.05$ , and a marginal effect of syntax  $F_{(1,188)} = 3.86$ ,  $p = 0.051$ . The interaction between clarity and syntax was not significant,  $F_{(1,188)} = 0.58$ ,  $p = 0.45$ . These results indicate that noise vocoding selectively affected some acoustic characteristics of the sentences that our syntactic manipulation did not.

## 5.2 Behavioral results

Behavioral results from within the scanner are shown in Table 2. All subjects performed well on the task (mean accuracy = 94.5%,  $SD = 4.5\%$ ). To test the main effects of syntactic complexity and acoustic richness, and the interaction between the two, we performed  $2 \times 2$  factorial ANOVAs on logit-transformed accuracy and raw reaction time data for correct responses in R software (v 3.0.2). For accuracy, there was no significant effect of syntactic complexity,  $F_{(1,100)} = 0.80$ ,  $p = 0.36$ . The main effect of acoustic richness was marginally significant,  $F_{(1,100)} = 3.79$ ,  $p = 0.054$ , but there was no interaction effect,  $F_{(1,100)} = 0.43$ ,  $p = 0.52$ . For reaction time, we computed the duration between the critical time point when the sentence diverges to either subject-relative (e.g., the onset of “help” in “Kings that help queens are nice”) or object-relative (e.g., the onset of “queens” in “Kings that queens help are nice”) construction and the onset of the button press. There were no significant effects of syntactic complexity  $F_{(1,100)} = 2.17$ ,  $p = 0.14$ , acoustic richness,  $F_{(1,100)} = 1.07$ ,  $p = 0.3$ , or their interaction  $F_{(1,100)} = 0.03$ ,  $p = 0.87$ .

## 5.3 fMRI results

We first used an FIR model to identify regions exhibiting differential responses for noise and intelligible speech, shown in Figure 2. This comparison revealed significant clusters throughout the whole brain including the bilateral superior temporal lobes as well as the left frontal lobe, encompassing inferior and dorsolateral aspects of the inferior frontal gyrus (IFG).

We then extracted summed positive response metrics from significant voxels to better understand directionality for sentence vs. noise F-contrast map using parameter estimates across 5 time points for each condition. Figure 2A reveals that frontal and temporal cortices were more responsive to sentences than to noises, whereas parietal, cingulate, and prefrontal cortices were more responsive to noises than to sentences. Figure 2B shows summed positive response plots in some peak voxels within the left IFG, left superior temporal gyrus (STG) and superior temporal sulcus (STS), and right STG and STS. As can be seen, the activation profile resembles a canonical hemodynamic response function with the peak response at approximately 4 seconds.

Additionally, we performed an LI analysis on the FIR data, which produced a weighted mean index of 0.80 in the whole brain, indicating a left-lateralized response. To see whether this lateralization was driven by a particular region of the brain, we repeated the LI analysis

separately for various lobes, and found consistent results in frontal cortex (0.89) and temporal cortex (0.78), but opposite lateralization in parietal cortex (−0.75). The summed positive response analysis suggests that intelligible sentences yielded more extended clusters in left frontotemporal cortex, while unintelligible noise yielded more extended clusters in right parietal cortex.

Lastly, we examined the timing of the response in each of these regions, for each subject. To this end, we calculated the time bin at which the maximum positive response was observed, and used a Friedman nonparametric repeated measures ANOVA to compare timing across condition. We found a significant difference in timecourses among the 4 conditions in left STG only ( $\chi^2(3) = 10.7$ ,  $p = 0.01$ ), but post-hoc Wilcoxon Signed Rank Tests did not reach significance in any of the pairwise comparisons (the sum of rank assigned to the difference was either 0 or 4, which yielded non-significant p-value).

We next examined the main effects of syntactic complexity and acoustic richness by collapsing sentences on the basis of either syntactic or acoustic differences, shown in Figure 3 and listed in Tables 3 and 4 (the full list of all sub-peaks can be found in Supplementary Tables 1 and 2). The main effect of syntactic complexity was most evident along left STG, MTG, and frontal cortex (including IFG), although there were also effects in right inferior frontal and superior temporal cortex. Subsequently, we computed the summed positive response associated with each condition to assess the directionality. As expected, most regions showed greater activation in response to object-relative than subject-relative sentences (Figure 3A). These results are comparable to previous studies using the same sentences (Pelle et al., 2010b, 2004), helping to validate our approach to analyzing the ISSS fMRI data.

Testing the main effect of acoustic richness revealed large bilateral effects, including STS/STG and parietal cortex, as well as smaller activations in left parietal cortex and superior temporal cortex (Figure 3B). The summed positive response comparison revealed that most of the regions showing an effect of acoustic richness yielded greater activation in response to acoustically rich speech compared to speech that was reduced in spectral detail by vocoding. Only a few regions in right frontal cortex showed increased activation for acoustically less-detailed speech and no significant clusters were seen in left IFG, unlike what we observed for syntactic complexity (Figure 3A).

Visual inspection of Figure 3 suggests that the response to syntactic complexity is left lateralized, whereas the response to acoustic richness is more bilateral, or even right-lateralized. We again used the LI toolbox to quantify lateralization for these comparisons. The LI measure for the syntactic complexity comparison produced a weighted mean LI of 0.65 in the whole brain, indicating a left-lateralized response to spoken sentences. To see whether this lateralization was driven by a particular region of the brain, we repeated the LI analysis separately for various lobes, and found similar results (frontal cortex: 0.82; parietal cortex: 0.81; temporal cortex: 0.88). The same comparison for acoustic richness resulted in an LI of −0.74 in the whole brain (frontal cortex: −0.67; parietal cortex: −0.62; temporal cortex: −0.78), indicating right-lateralized activity.



Finally, we tested the interaction between syntactic complexity and acoustic richness, shown in Figure 4A. This comparison yielded significant clusters in right posterior STG, left supramarginal gyrus (SMG), and right IFG. Among these regions, right posterior STG showed the most robust interaction effect (Table 5 and Supplementary Table 3). To understand the basis of the interaction, we extracted the data for the four sentence conditions from the summed positive response images. As shown in Figure 4B, the interaction occurred due to the fact that the established syntactic effect was attenuated in acoustically less-detailed speech. While this general pattern was consistent throughout the regions that we specified, we also observed two qualitatively distinct patterns of interaction: In the posterior temporal regions, we found greater activation for object-relative sentences in the acoustically rich condition compared to the acoustically less-detailed condition, although there was no difference between acoustically rich and acoustically less-detailed conditions for the subject-relative sentences. In the right inferior frontal gyrus, we observed greater activation in the acoustically less-detailed condition than the rich condition for the subject-relative sentences, although the object-relative sentences had similar levels of activation for the acoustically rich and less-detailed conditions.

## 6. Discussion

Human vocal communication relies not only on the linguistic content of speech (the words spoken and their grammatical arrangement), but also on paralinguistic aspects including speaker sex, identity, and emotion (McGettigan, 2015). When speech is degraded, the loss of acoustic clarity can differently affect these two complementary aspects of communication. Our goal in the current study was to examine whether reduced acoustic richness with preserved intelligibility affected neural processing of speech information.

Our behavioral results confirmed that the acoustic manipulation we used had little impact on the successful comprehension of both syntactically simpler and syntactically more complex sentences, with accuracy above 92% correct in all conditions. However, the neural substrate for speech comprehension was differentially responsive to changes in acoustic richness and syntactic complexity. Together, our findings demonstrate that the neural systems engaged during speech comprehension are sensitive to the amount of acoustic detail even when speech is fully intelligible, and that areas showing this sensitivity differ from those that support the processing of computationally demanding linguistic structures. These patterns of activation suggest that two distinct but interacting mechanisms contribute to successful auditory sentence comprehension.

### 6.1 A core frontotemporal network for speech comprehension

Regions showing an increased response to spoken sentences relative to noise were largely comparable to previous studies of spoken language comprehension. Compared to unintelligible control stimuli, listening to intelligible sentences resulted in increased activity in a large-scale frontotemporal network. These regions included bilateral temporal cortex and left IFG, which are consistently activated in response to spoken sentences (Adank, 2012), and reflect the combination of acoustic and linguistic processing required to extract meaning from connected speech. In the current study, this network was active for both sentence types compared to 1 channel vocoded speech, and also for more complex object-

relative sentences compared to subject-relative sentences. These findings were expected given the large literature on the role of left IFG in processing syntactically complex sentences (Friederici, 2011; Peelle et al., 2010a). The fact that regions of the speech network respond preferentially as a function of syntactic complexity suggests activity is due in part to language processing (as opposed to general acoustic processing or task effects). As with several previous studies, sentence-related processing was stronger on the left, including larger portions of posterior STG and MTG, this lateralization likely reflecting processes related to the linguistic complexity of sentence stimuli (McGettigan and Scott, 2012; Peelle, 2012).

We also observed increased activity in left inferior temporal cortex for intelligible speech compared to 1 channel vocoded speech. Although inferior temporal cortex is not typically emphasized in descriptions of speech networks, activity here is sometimes observed during sentence comprehension (Davis et al., 2011; Rodd et al., 2010), which may be related to the observation of increased activation in inferior temporal cortex during resolution of the meaning of semantically ambiguous words (Rodd et al., 2012, 2005).

The increased activity we observed for syntactically complex speech fits well with prior reports using these same materials (Hassanpour et al., 2015; Peelle et al., 2010b) and other studies manipulating grammatical challenge (Cooke et al., 2002; Rodd et al., 2010; Tyler et al., 2010). Although we did not find significant behavioral effects of syntactic complexity, this is not particularly surprising in the scanner environment. In the context of past behavioral and imaging results (Cooke et al., 2002; Gibson et al., 2013; Just et al., 1996; Peelle et al., 2010b) and our current imaging results, we see consistent evidence in support of the increased challenge of object-relative sentences.

It is important to note that in our study listeners needed to make an explicit decision following every sentence. This task resulted in additional cognitive demands related to decision-making and response selection that may be less evident in passive listening. Some of the activity we see in prefrontal cortex when comparing sentences to the noise condition may have been due in part to these metalinguistic task demands. However, the contrast comparing object-relative and subject-relative sentences involved similar decision-making for both types of sentences, yet we observed increased prefrontal activation for the object-relative sentences compared to the subject-relative sentences, suggesting that prefrontal activation is related at least in part to the linguistic demands of the sentences.

Thus, when listeners heard intelligible speech they activated a core speech network. Greater activation for object-relative compared to subject-relative sentences is consistent with a linguistic contribution of this frontotemporal network to auditory sentence comprehension. These findings are consistent with previous reports, and provide a context within which to consider the effects of acoustic richness.

## 6.2 Speech processing modulated by acoustic richness

The current study advances extant work by addressing the important question of whether acoustic richness modulates the activity of regions recruited during intelligible speech processing. Our acoustically less-detailed condition was designed to degrade the acoustic

richness of the speech signal without compromising intelligibility, allowing us to disentangle effects of intelligibility and acoustic detail that are frequently confounded. Noise vocoding necessarily changes a number of acoustic properties of the speech (e.g., spectral intensity, harmonic-to-noise ratio, center of gravity). Our goal was not to isolate which specific acoustic features affect neural processing, but to investigate the degree to which acoustic details matter when speech is highly intelligible.

We found more activity for acoustically rich speech relative to spectrally-impooverished speech in bilateral parietal, temporal, and frontal cortices, with these differences stronger in the right hemisphere. These findings suggest that the spectral content of the acoustically rich speech allowed listeners to process information not present in the vocoded speech. A natural question, then, is what might this increased right-lateralized activity reflect?

The functional role of the right hemisphere during speech processing is still a matter of debate (McGettigan and Scott, 2012). A longstanding view is that right temporal cortex preferentially processes spectral information (Obleser et al., 2008; Zatorre and Belin, 2001). In the context of speech comprehension, paralinguistic information (e.g., speaker identity, prosody, or emotion) is typically conveyed by modulating the spectral details (Kyong et al., 2014). The fact that we observed less activation in response to the vocoded-but-intelligible speech compared to acoustically rich speech suggests that computations on paralinguistic information were limited. This finding has important implications for listeners with hearing impairment or who are using assistive listening devices such as hearing aids or cochlear implants: Even in cases where intelligibility is high, acoustic richness may modulate other aspects of speech comprehension that go beyond linguistic content.

It is important to note that we also observed some regions that showed more activity for acoustically less-detailed speech than acoustically rich speech. Although all sentences were intelligible, listeners may have required increased effort in order to decode the content of the vocoded speech, consistent with evidence from dual-task and pupillometry studies suggesting that processing vocoded speech requires cognitive processes even when intelligible (Pals et al., 2013; Winn et al., 2015). Such effort may have been associated with executive resource allocation evidenced by increased frontal activation, as other regions of frontal cortex have been associated with effortful listening in previous studies, albeit with speech that was not fully intelligible (Davis and Johnsrude, 2003; Hervais-Adelman et al., 2012).

### 6.3 The interaction of acoustic and linguistic processing

We found significant interaction effects in circumscribed regions of temporal, parietal, and frontal cortices bilaterally. As can be seen in Figure 4B, the well-established effect of syntactic complexity was attenuated by loss of acoustic richness, despite preserved intelligibility. This result may be supported by previous behavioral evidence, in which older adults with hearing loss (i.e., acoustic degradation) showed poorer performance on processing syntactically different aural sentences (Wingfield et al., 2006). That is, neural resources may have been allocated to abstracting speech information in acoustically degraded signals (i.e., sensory processing), which in turn may have exacerbated syntactic processing (cognitive processing). Although this interpretation needs to be corroborated by

future experiments, present data suggest that these regions may play a key role in dynamic allocation of neural resources to handle simultaneous acoustic and linguistic challenge. Obleser et al. (2011) also varied spectral detail and linguistic complexity, presenting sentences at different levels of spectral detail (vocoded with 8, 16, or 32 channels) and different levels of word-order scrambling. A key difference is that Obleser and colleagues did not include a normal speech condition. Although the specific stimuli differ from our current study, some parallels can be drawn. First, similar to our current results, the authors found bilateral increases in activity for the 32 channel vocoded speech compared to 8 channel vocoded speech. These were less extensive than in the current study, which may be due in part to the fact that 32 channel vocoded speech is still acoustically impoverished compared to the rich, natural speech that we used. Second, their results suggested that the effects of acoustic richness affected the distribution of the clusters responding to linguistic complexity. This latter finding is broadly consistent with our current results, in which we showed a significant interaction between acoustic richness and syntactic complexity. Thus, although the specific patterns of activity depend on the details of the stimuli employed, there is increasing consensus that even when speech is fully intelligible listeners' brains respond differently as a function of the acoustic and linguistic processing that must be done.

## 6.4 Conclusions

Our results show that the brain dynamically adjusts to speech characteristics in order to maintain a high level of comprehension. We found neural responses that differed as functions of syntactic complexity, acoustic richness, and their interaction, all while speech remained fully intelligible. These findings highlight the adaptive nature of neural processing for spoken language, and emphasize that dimensions other than intelligibility play a critical role in the neural processes underlying human vocal communication.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by NIH grants R01AG038490, R01AG019714, P01AG017586, R01NS044266, P01AG032953, and the Dana Foundation. We thank Ethan Kotloff for his help on data sorting and analyses, the radiographers at the University of Pennsylvania for their help with data collection, and our volunteers for their participation.

## References

- Adank P. Design choices in imaging speech comprehension: an Activation Likelihood Estimation (ALE) meta-analysis. *NeuroImage*. 2012; 63:1601–1613.10.1016/j.neuroimage.2012.07.027 [PubMed: 22836181]
- Ashburner J, Friston KJ. Unified segmentation. *NeuroImage*. 2005; 26:839–851.10.1016/j.neuroimage.2005.02.018 [PubMed: 15955494]
- Chen F, Loizou PC. Predicting the intelligibility of vocoded speech. *Ear Hear*. 2011; 32:331–338.10.1097/AUD.0b013e3181ff3515 [PubMed: 21206363]
- Chen F, Loizou PC. Contribution of Consonant Landmarks to Speech Recognition in Simulated Acoustic-Electric Hearing. *Ear Hear*. 2010; 31:259–267.10.1097/AUD.0b013e3181c7db17 [PubMed: 20081538]

- Churchill TH, Kan A, Goupell MJ, Ihlefeld A, Litovsky RY. Speech Perception in Noise with a Harmonic Complex Excited Vocoder. *J Assoc Res Otolaryngol*. 2014; 15:265–278.10.1007/s10162-013-0435-7 [PubMed: 24448721]
- Cooke A, Zurif EB, DeVita C, Alsop D, Koenig P, Detre J, Gee J, Pinãngo M, Balogh J, Grossman M. Neural basis for sentence comprehension: Grammatical and short-term memory components. *Hum Brain Mapp*. 2002; 15:80–94.10.1002/hbm.10006 [PubMed: 11835600]
- Davis MH, Ford MA, Kherif F, Johnsrude IS. Does semantic context benefit speech understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. *J Cogn Neurosci*. 2011; 23:3914–3932.10.1162/jocn\_a\_00084 [PubMed: 21745006]
- Davis MH, Johnsrude IS. Hierarchical Processing in Spoken Language Comprehension. *J Neurosci*. 2003; 23:3423–3431. [PubMed: 12716950]
- Eckert MA, Menon V, Walczak A, Ahlstrom J, Denslow S, Horwitz A, Dubno JR. At the heart of the ventral attention system: The right anterior insula. *Hum Brain Mapp*. 2009; 30:2530–2541.10.1002/hbm.20688 [PubMed: 19072895]
- Erb J, Henry MJ, Eisner F, Obleser J. The Brain Dynamics of Rapid Perceptual Adaptation to Adverse Listening Conditions. *J Neurosci*. 2013; 33:10688–10697.10.1523/JNEUROSCI.4596-12.2013 [PubMed: 23804092]
- Evans S, Kyong JS, Rosen S, Golestani N, Warren JE, McGettigan C, Mourão-Miranda J, Wise RJS, Scott SK. The Pathways for Intelligible Speech: Multivariate and Univariate Perspectives. *Cereb Cortex*. 2014; 24:2350–2361.10.1093/cercor/bht083 [PubMed: 23585519]
- Faulkner A, Rosen S, Wilkinson L. Effects of the number of channels and speech-to-noise ratio on rate of connected discourse tracking through a simulated cochlear implant speech processor. *Ear Hear*. 2001; 22:431–438. [PubMed: 11605950]
- Friederici AD. The Brain Basis of Language Processing: From Structure to Function. *Physiol Rev*. 2011; 91:1357–1392.10.1152/physrev.00006.2011 [PubMed: 22013214]
- Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC. Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp*. 1994; 1:210–220.10.1002/hbm.460010306 [PubMed: 24578041]
- Gibson E, Bergen L, Piantadosi ST. Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proc Natl Acad Sci*. 2013; 110:8051–8056.10.1073/pnas.1216438110 [PubMed: 23637344]
- Gobl C, Chasaide AN. The Role of Voice Quality in Communicating Emotion, Mood and Attitude. *Speech Commun*. 2003; 40:189–212.10.1016/S0167-6393(02)00082-1
- Gorgolewski KJ, Varoquaux G, Rivera G, Schwarz Y, Ghosh SS, Maumet C, Sochat VV, Nichols TE, Poldrack RA, Poline JB, Yarkoni T, Margulies DS. NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front Neuroinformatics*. 2015; 9:8.10.3389/fninf.2015.00008
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp*. 1999; 7:213–223. [PubMed: 10194620]
- Hassanpour MS, Eggebrecht AT, Culver JP, Peelle JE. Mapping cortical responses to speech using high-density diffuse optical tomography. *NeuroImage*. 2015; 117:319–326.10.1016/j.neuroimage.2015.05.058 [PubMed: 26026816]
- Hervais-Adelman AG, Carlyon RP, Johnsrude IS, Davis MH. Brain Regions Recruited for the Effortful Comprehension of Noise-Vocoded Words. *Lang Cogn Process*. 2012; 27:1145–1166.10.1080/01690965.2012.662280
- Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci*. 2007; 8:393–402.10.1038/nrn2113 [PubMed: 17431404]
- Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR. Brain Activation Modulated by Sentence Comprehension. *Science*. 1996; 274:114–116.10.1126/science.274.5284.114 [PubMed: 8810246]
- Kyong JS, Scott SK, Rosen S, Howe TB, Agnew ZK, McGettigan C. Exploring the Roles of Spectral Detail and Intonation Contour in Speech Intelligibility: An fMRI Study. *J Cogn Neurosci*. 2014; 26:1748–1763.10.1162/jocn\_a\_00583 [PubMed: 24568205]

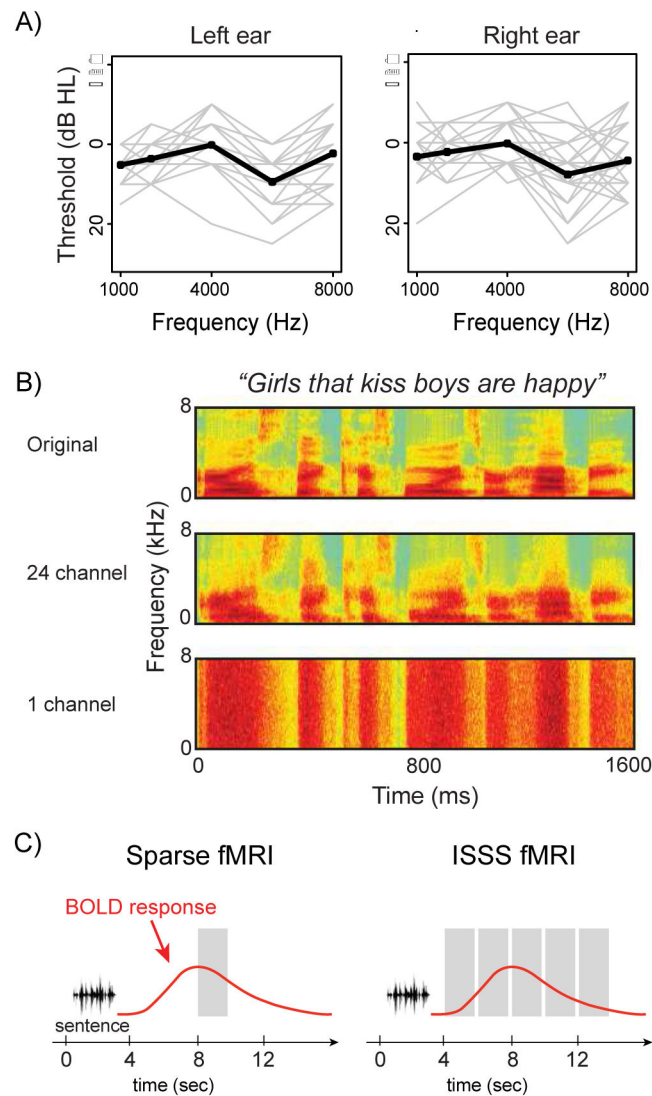
- Loftus GR, Masson ME. Using confidence intervals in within-subject designs. *Psychon Bull Rev.* 1994; 1:476–490.10.3758/BF03210951 [PubMed: 24203555]
- Loizou, PC. Speech Processing in Vocoder-Centric Cochlear Implant. In: Møller, AR., editor. *Advances in Oto-Rhino-Laryngology.* S. KARGER AG; Basel: 2006. p. 109-143.
- Maryn Y, Roy N, Bodt MD, Cauwenberge PV, Corthals P. Acoustic measurement of overall voice quality: A meta-analysis. *J Acoust Soc Am.* 2009; 126:2619–2634.10.1121/1.3224706 [PubMed: 19894840]
- McGettigan C. The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication. *Front Hum Neurosci.* 2015; 9:129.10.3389/fnhum.2015.00129 [PubMed: 25852517]
- McGettigan C, Scott SK. Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends Cogn Sci.* 2012; 16:269–276.10.1016/j.tics.2012.04.006 [PubMed: 22521208]
- Obleser J, Eisner F, Kotz SA. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci Off J Soc Neurosci.* 2008; 28:8116–8123.10.1523/JNEUROSCI.1290-08.2008
- Obleser J, Meyer L, Friederici AD. Dynamic assignment of neural resources in auditory comprehension of complex sentences. *NeuroImage.* 2011; 56:2310–2320.10.1016/j.neuroimage.2011.03.035 [PubMed: 21421059]
- Obleser J, Wise RJS, Dresner MA, Scott SK. Functional Integration across Brain Regions Improves Speech Perception under Adverse Listening Conditions. *J Neurosci.* 2007; 27:2283–2289.10.1523/JNEUROSCI.4663-06.2007 [PubMed: 17329425]
- Pals C, Sarampalis A, Baskent D. Listening effort with cochlear implant simulations. *J Speech Lang Hear Res JSLHR.* 2013; 56:1075–1084.10.1044/1092-4388(2012/12-0074) [PubMed: 23275424]
- Peelle JE. Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Brain Imaging Methods.* 2014; 8:253.10.3389/fnins.2014.00253
- Peelle JE. The hemispheric lateralization of speech processing depends on what “speech” is: a hierarchical perspective. *Front Hum Neurosci.* 2012; 6:309.10.3389/fnhum.2012.00309 [PubMed: 23162455]
- Peelle JE, Johnsrude I, Davis MH. Hierarchical processing for speech in human auditory cortex and beyond. *Front Hum Neurosci.* 2010a; 5:110.3389/fnhum.2010.00051
- Peelle JE, McMillan C, Moore P, Grossman M, Wingfield A. Dissociable patterns of brain activity during comprehension of rapid and syntactically complex speech: Evidence from fMRI. *Brain Lang.* 2004; 91:315–325.10.1016/j.bandl.2004.05.007 [PubMed: 15533557]
- Peelle JE, Troiani V, Wingfield A, Grossman M. Neural Processing during Older Adults' Comprehension of Spoken Sentences: Age Differences in Resource Allocation and Connectivity. *Cereb Cortex.* 2010b; 20:773–782.10.1093/cercor/bhp142 [PubMed: 19666829]
- Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci.* 2009; 12:718–724.10.1038/nn.2331 [PubMed: 19471271]
- Rodd JM, Davis MH, Johnsrude IS. The Neural Mechanisms of Speech Comprehension: fMRI studies of Semantic Ambiguity. *Cereb Cortex.* 2005; 15:1261–1269.10.1093/cercor/bhi009 [PubMed: 15635062]
- Rodd JM, Johnsrude IS, Davis MH. Dissociating Frontotemporal Contributions to Semantic Ambiguity Resolution in Spoken Sentences. *Cereb Cortex.* 2012; 22:1761–1773.10.1093/cercor/bhr252 [PubMed: 21968566]
- Rodd JM, Longe OA, Randall B, Tyler LK. The functional organisation of the fronto-temporal language system: Evidence from syntactic and semantic ambiguity. *Neuropsychologia.* 2010; 48:1324–1335.10.1016/j.neuropsychologia.2009.12.035 [PubMed: 20038434]
- Rönnerberg J, Lunner T, Zekveld A, Sörqvist P, Danielsson H, Lyxell B, Dahlström Ö, Signoret C, Stenfelt S, Pichora-Fuller MK, Rudner M. The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Front Syst Neurosci.* 2013; 7:31.10.3389/fnsys.2013.00031 [PubMed: 23874273]
- Rorden C, Brett M. Stereotaxic display of brain lesions. *Behav Neurol.* 2000; 12:191–200. [PubMed: 11568431]



- Schwarzbauer C, Davis MH, Rodd JM, Johnsrude I. Interleaved silent steady state (ISSS) imaging: A new sparse imaging method applied to auditory fMRI. *NeuroImage*. 2006; 29:774–782.10.1016/j.neuroimage.2005.08.025 [PubMed: 16226896]
- Scott SK, Blank CC, Rosen S, Wise RJS. Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*. 2000; 123:2400–2406.10.1093/brain/123.12.2400 [PubMed: 11099443]
- Tyler LK, Shafto MA, Randall B, Wright P, Marslen-Wilson WD, Stamatakis EA. Preserving Syntactic Processing across the Adult Life Span: The Modulation of the Frontotemporal Language System in the Context of Age-Related Atrophy. *Cereb Cortex*. 2010; 20:352–364.10.1093/cercor/bhp105 [PubMed: 19505991]
- Vaden KI, Kuchinsky SE, Cute SL, Ahlstrom JB, Dubno JR, Eckert MA. The Cingulo-Opercular Network Provides Word-Recognition Benefit. *J Neurosci*. 2013; 33:18979–18986.10.1523/JNEUROSCI.1417-13.2013 [PubMed: 24285902]
- Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens TEJ, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss SW, Della Penna S, Feinberg D, Glasser MF, Harel N, Heath AC, Larson-Prior L, Marcus D, Michalareas G, Moeller S, Oostenveld R, Petersen SE, Prior F, Schlaggar BL, Smith SM, Snyder AZ, Xu J, Yacoub E. The Human Connectome Project: A data acquisition perspective. *NeuroImage, Connectivity*. 2012; 62:2222–2231.10.1016/j.neuroimage.2012.02.018
- Wild CJ, Yusuf A, Wilson DE, Peelle JE, Davis MH, Johnsrude IS. Effortful Listening: The Processing of Degraded Speech Depends Critically on Attention. *J Neurosci*. 2012; 32:14010–14021.10.1523/JNEUROSCI.1528-12.2012 [PubMed: 23035108]
- Wilke M, Lidzba K. LI-tool: A new toolbox to assess lateralization in functional MR-data. *J Neurosci Methods*. 2007; 163:128–136.10.1016/j.jneumeth.2007.01.026 [PubMed: 17386945]
- Wingfield A, McCoy SL, Peelle JE, Tun PA, Cox CL. Effects of Adult Aging and Hearing Loss on Comprehension of Rapid Speech Varying in Syntactic Complexity. *J Am Acad Audiol*. 2006; 17:487–497.10.3766/jaaa.17.7.4 [PubMed: 16927513]
- Winn MB, Edwards JR, Litovsky RY. The Impact of Auditory Spectral Resolution on Listening Effort Revealed by Pupil Dilation. *Ear Hear*. 2015; 36:e153–165.10.1097/AUD.0000000000000145 [PubMed: 25654299]
- Worsley KJ, Evans AC, Marrett S, Neelin P. A three-dimensional statistical analysis for CBF activation studies in human brain. *J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab*. 1992; 12:900–918.10.1038/jcbfm.1992.127
- Zatorre RJ, Belin P. Spectral and Temporal Processing in Human Auditory Cortex. *Cereb Cortex*. 2001; 11:946–953.10.1093/cercor/11.10.946 [PubMed: 11549617]

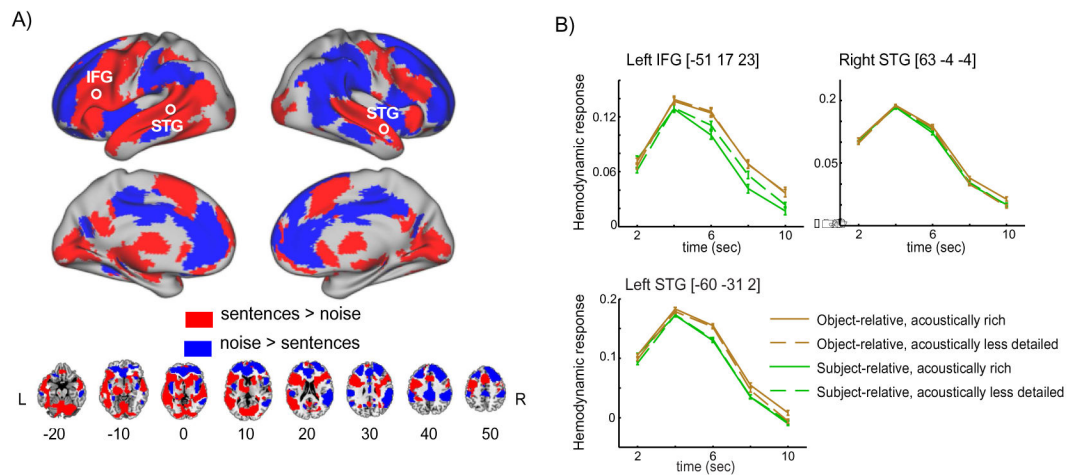
### Highlights

- We examined the consequences of acoustic richness and syntactic complexity on the speech network using Interleaved silent steady-state (ISSS) fMRI.
- Acoustic (Spectral) degradation of speech signal results in reduced activation of right fronto-temporal network despite high intelligibility.
- Syntactic complexity yields upregulation of a left frontotemporal network.



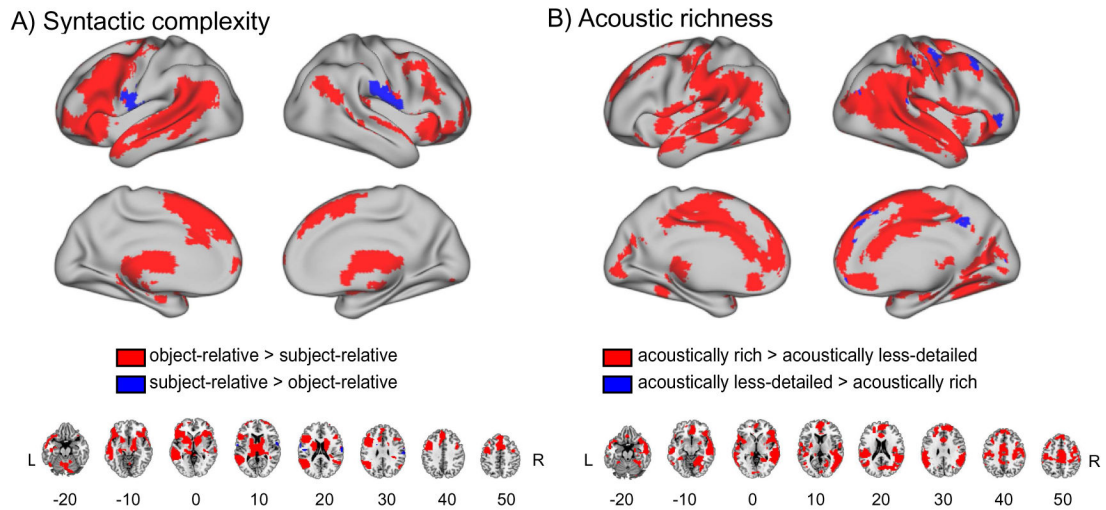
**Figure 1.**

**A.** Pure-tone hearing acuity for participants' left and right ears. Individual listeners' profiles are shown in gray lines, with the group mean in black. **B.** Spectrograms of a representative sentence in the three acoustic conditions tested: unprocessed speech (acoustically rich), vocoded with 24 channels (acoustically less detailed but fully intelligible), or vocoded with 1 channel (unintelligible). **C.** Schematic comparison between a traditional sparse fMRI protocol and the ISSS protocol used in the current study.



**Figure 2.**

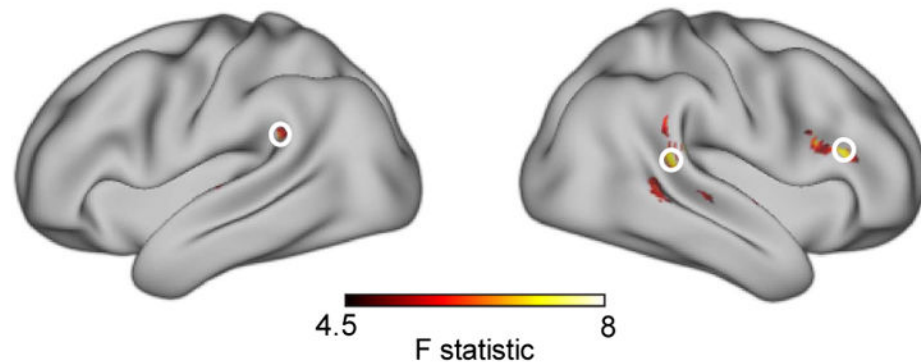
**A.** Brain regions discriminating between intelligible sentence materials and unintelligible noise (1 channel vocoded speech). Significant voxels were identified using a whole-brain FIR model, subsequently color-coded according to the sign of the direct comparison of summed positive responses between conditions. **B.** fMRI intensities extracted from peak voxels located in left IFG, left STG, and right STG (circled in the rendering view). The x-axis depicts time relative to sentence onset. Error bars represent one standard error of the mean with between-subjects variance removed, suitable for within-subjects comparisons (Loftus and Masson, 1994).



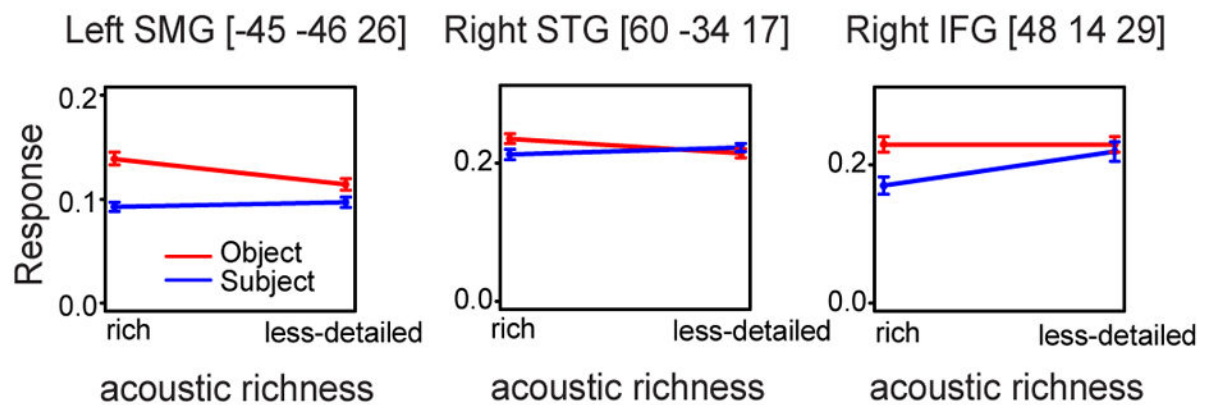
**Figure 3.**

**A.** Main effect of syntactic complexity comparing subject-relative sentences to object-relative sentences. **B.** Main effect of acoustic richness comparing acoustically rich speech (unprocessed) to intelligible but acoustically less-detailed speech (24 channel noise vocoded). As in Figure 2, significant voxels were identified using a whole-brain FIR model, subsequently color-coded according to the sign of the direct comparison of summed positive response metrics between groups.

## A) Interaction between syntax and acoustic manipulations



## B)



**Figure 4.**

**A.** Cortical regions exhibiting a significant interaction between syntax and acoustic manipulations. **B.** Parameter estimates from the summed positive response analyses extracted from peak voxels located in left SMG, right STG, and right IFG (circled in the rendering above) for object-related and subject-related sentences. Error bars represent one standard error of the mean with between-subjects variance removed.



**Table 1**Mean ( $\pm$  SD) acoustic measures of sentence stimuli

	Acoustically rich		Acoustically less-detailed	
	subject relative	object relative	subject relative	object relative
Root mean square amplitude	0.07 (0.0)	0.07 (0.0)	0.07 (0.0)	0.07 (0.0)
Long-term average spectrum	1.6 (2.3)	2.3 (1.5)	-20.3 (1.5)	-20.0 (1.5)
Harmonic to noise ratio	10.8 (1.2)	10.4 (0.7)	1.3 (0.7)	1.4 (0.7)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**Mean ( $\pm$  SD) behavioral performance during the fMRI experiment

	Acoustically rich		Acoustically less-detailed	
	subject relative	object relative	subject relative	object relative
Accuracy (%correct)	96.6 (4.6)	94.8 (5.5)	93.7 (5.9)	92.9 (7.6)
Reaction time (ms)	2256 (586)	2422 (679)	2367 (585)	2574 (727)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Maxima for main effects of syntactic complexity

Region Name	Z-val	MNI Coordinates			volume of cluster (ul)
		x	y	z	
L posterior aspect of STS	>7.72	-60	-55	14	159840
L thalamus	>7.72	-12	-13	8	
L orbitofrontal cortex	>7.72	-48	17	20	
R insula	6.50	33	29	2	
R cerebellum	>7.72	18	-73	-28	78867
L cerebellum	7.72	-39	-70	-34	
L cerebellum	7.63	-9	-79	-28	
L SMA	>7.72	-3	17	53	25488
L superior frontal gyrus	5.95	-6	38	44	
L ACC	5.33	-9	20	32	
R precentral gyrus	5.81	39	-1	50	5400
R IFG (pars triangularis)	4.73	48	23	23	
R precentral gyrus	4.54	45	5	38	
L postcentral gyrus	5.01	-60	-7	14	1431
L Rolandic Operculum	3.89	-48	-13	20	
R superior frontal gyrus	4.94	24	65	14	2160
R superior frontal gyrus	4.54	42	56	2	
R superior frontal gyrus	3.99	36	59	14	
R postcentral gyrus	4.93	63	-7	17	2592
R supramarginal gyrus	4.90	66	-19	26	
R postcentral gyrus	3.85	63	-4	29	

L=left; R=right; STG=superior temporal gyrus; MTG=middle temporal gyrus

ACC=anterior cingulate cortex; SMA=supplementary motor area; IFG=inferior frontal gyrus

Table 4

Maxima for main effects of acoustic richness

Region Name	MNI Coordinates			volume of cluster ( <i>af</i> )	
	z-val	x	y		z
R middle part of MTG	>7.58	63	-25	-7	251505
R orbitofrontal cortex	7.53	51	29	-7	
R angular gyrus	7.13	54	-49	29	
L Hippocampus	5.08	-12	-34	2	2214
L fusiform cortex	3.87	-27	-46	-16	1647
L parahippocampus	3.86	-30	-43	-7	
L fusiform cortex	3.83	-30	-43	-13	

Table 5

Maxima for interaction between syntax and acoustic richness

Region Name	MINI Coordinates				volume of cluster ( <i>nl</i> )
	z-val	x	y	z	
R posterior STG	6.13	60	-34	17	7236
R supramarginal gyrus	4.64	63	-40	29	
R posterior MTG	4.37	60	-43	2	
R anterior MTG	5.92	-60	-4	-7	1836
R anterior STG	4.30	-63	-19	5	
R anterior MTG	4.15	-63	-10	-13	
L superior frontal gyrus	5.45	-9	41	41	2727
R IFG (pars opercularis)	5.22	48	14	29	4806
R IFG (pars triangularis)	5.08	51	26	20	4806
L supramarginal gyrus	5.19	-45	-46	26	2079
L supramarginal gyrus	3.57	-63	-40	26	
R anterior STG/STS	4.27	60	2	-1	1836