

An Application of Univariate and Multivariate Approaches in fMRI to Quantifying the Hemispheric Lateralization of Acoustic and Linguistic Processes

Carolyn McGettigan*, Samuel Evans*, Stuart Rosen, Zarinah K. Agnew, Poonam Shah, and Sophie K. Scott

Abstract

■ The question of hemispheric lateralization of neural processes is one that is pertinent to a range of subdisciplines of cognitive neuroscience. Language is often assumed to be left-lateralized in the human brain, but there has been a long running debate about the underlying reasons for this. We addressed this problem with fMRI by identifying the neural responses to amplitude and spectral modulations in speech and how these interact with speech intelligibility to test previous claims for hemispheric asymmetries in acoustic and linguistic processes in speech perception. We used both univariate and multivariate analyses of the data, which enabled us to both identify the networks involved in processing these acoustic and linguistic factors and to test the significance of any apparent hemispheric asymmetries. We demonstrate bilateral activation of superior

temporal cortex in response to speech-derived acoustic modulations in the absence of intelligibility. However, in a contrast of amplitude-modulated and spectrally modulated conditions that differed only in their intelligibility (where one was partially intelligible and the other unintelligible), we show a left dominant pattern of activation in STS, inferior frontal cortex, and insula. Crucially, multivariate pattern analysis showed that there were significant differences between the left and the right hemispheres only in the processing of intelligible speech. This result shows that the left hemisphere dominance in linguistic processing does not arise because of low-level, speech-derived acoustic factors and that multivariate pattern analysis provides a method for unbiased testing of hemispheric asymmetries in processing. ■

INTRODUCTION

The question of hemispheric asymmetries in auditory processing, which might underlie a left hemispheric dominance in speech and language processing, has long been a popular topic for neuroscientific investigation (Boemio, Fromm, Braun, & Poeppel, 2005; Schönwiesner, Rubsamen, & von Cramon, 2005; Zatorre & Belin, 2001). Current theories posit, for example, differential processing of temporal versus spectral information (e.g., Zatorre & Belin, 2001) or differences in the preference for short versus long temporal integration windows (e.g., Poeppel, 2003) in the left and right temporal lobes. In parallel, a number of functional imaging studies of speech perception have identified responses to intelligibility in anterior sites on the STS, which, when contrasted with a complex acoustic control, are typically left lateralized or left dominant (Eisner, McGettigan, Faulkner, Rosen, & Scott, 2010; Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Narain et al., 2003; Scott, Blank, Rosen, & Wise, 2000). Some studies have investigated the acoustic basis for this pattern of lateralization using modifications to intelligible

speech (Obleser, Eisner, & Kotz, 2008). In contrast, others suggest that the prelexical processing of speech is not actually dominated by the left temporal lobe (Okada et al., 2010; Hickok & Poeppel, 2007).

Generally, the perception of acoustic structure is associated with bilateral cortical activation. For example, harmonic structure (Hall et al., 2002), frequency modulation (Hart, Palmer, & Hall, 2003; Hall et al., 2002), amplitude modulation (Hart et al., 2003; Giraud et al., 2000), spectral modulations (Thivard, Belin, Zilbovicius, Poline, & Samson, 2000), or dynamic spectral ripples (Langers, Backes, & van Dijk, 2003) all generate bilateral neural responses, with no evidence for asymmetry. However, these studies were not designed to necessarily test existing models of hemispheric asymmetries.

Several studies have directly tested the hypothesis that differences between how left and right temporal lobes respond to speech might reflect differential sensitivity to acoustic factors. An early neuroimaging study using PET (Belin, Zilbovicius, Crozier, Thivard, & Fontaine, 1998) employed stimuli with short (40 msec) and long (200 msec) “formant” transition times at the onset of sounds. Although speech-like, these stimuli did not form recognizable speech tokens. The analysis showed that both long and short formant transitions were processed

University College London

*These authors contributed equally to this work.

bilaterally in the superior temporal gyrus (STG). However, the direct comparison (long > short) gave activation in the right temporal lobe, whereas the opposite contrast (short > long) led to no significant activations. This study was widely interpreted as indicating a preference for rapid changes in the left temporal lobe, but it was in fact the right temporal lobe that responded preferentially to the long stimuli. The stimuli were constructed such that both the short and long transitions were associated with same offset sound, which meant that the overall duration of the stimuli covaried with the length of the onset transition. This makes it hard to determine whether the right STG preference is for slower spectral transitions or for longer sounds per se.

In a more recent fMRI study, Obleser and colleagues (2008) used noise-vocoded speech to examine the neural responses to changes in the spectral detail (i.e., number of channels) and the amount of amplitude envelope information within each channel (by varying the envelope smoothness), which was termed a temporal factor. They showed a greater response to amplitude envelope detail in the left than the right STG and a greater response to spectral detail on the right than on the left. However, within the left STG, the response to spectral detail was greater than the response to amplitude envelope detail. This finding is difficult to set within a proposed left hemisphere preference for “temporal” information. Likewise, their demonstration (following on from Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) that spectral detail was much more important to intelligibility than amplitude envelope detail would predict that it is the right temporal lobe that is predominantly associated with comprehension of the spoken word, a proposal at odds with the clinical literature.

Other studies have used stimuli that are not derived from speech to investigate potential hemispheric asymmetries in the neural response to acoustic characteristics. Zatorre and Belin (2001) varied the rate at which two short tones of different frequencies were repeated to create a “temporal” dimension and varied the size of the pitch changes between successive tones to create a “spectral” dimension (although as all the tones were sine tones, the instantaneous spectrum of these sounds would not have varied in complexity). These manipulations yielded bilateral activations in the dorsolateral temporal lobes. A direct comparison of the two conditions also showed bilateral activation, with the temporal stimuli leading to greater activation in bilateral Heschl’s gyri (HGs) and the spectral stimuli leading to greater activity in bilateral anterior STG fields. The parametric analysis showed a significantly greater slope fitted to the cortical response to temporal (rate) detail in the left anterior STG and to the pitch-varying detail in the right STG. However a direct comparison of the parametric effects of each kind of manipulation within each hemisphere was not reported.

Boemio and colleagues (2005) varied the way information in nonspeech sequences changed at different timescales by varying the duration of segments in the

sequence and the rate and extent of pitch change across the sequence. They found greater responses in the right STG as segment duration increased, consistent with a potential right hemisphere preference for items at longer timescales. However, there was no selective left hemisphere preference for the shorter-duration items. Schönwiesner and colleagues (2005) generated nonspeech sounds in which the spectral and temporal modulation densities were varied parametrically. Bilateral responses were seen to both manipulations, and as in Zatorre and Belin (2001), the authors compared the slopes of the neural responses to the temporal or spectral modulation density. They found a significant correlation of activation in right anterior STG with spectral modulation density and fitted a significant slope of activation in left anterior STG against temporal modulation density. However, as in the study by Zatorre and Belin (2001), they did not directly test for selectivity to either kind of information by comparing the activation to both temporal and spectral modulation density within either hemisphere.

Two recent studies presented a novel approach by investigating how fMRI signal correlated with different bandwidths of endogenous cortical EEG activity (Morillon et al., 2010; Giraud et al., 2007). Giraud et al. (2007) focused on relatively long (3–6 Hz) and short (28–40 Hz) temporal windows and found in two experiments a significantly greater correlation of the BOLD response in right auditory cortex with oscillatory activity in the 3–6 Hz frequency range, in line with the predictions of the Asymmetric Sampling in Time hypothesis (Poeppel, 2003). However, there was no significantly greater correlation in the left auditory fields with the 28–40 Hz frequency range, and in left lateral HG, the correlation with the 3–6 Hz temporal range was in fact greater than that with the 28–40 Hz range. This is not strong evidence in favor of a selective response to short timescale information in left auditory areas.

These kinds of studies are widely presented as indicating a clear difference in the ways that the left and right auditory cortices deal with acoustic information. However, the actual results are both more complex and more simple: A right hemisphere sensitivity to longer sounds (Boemio et al., 2005; Belin et al., 1998) and sounds with dynamic pitch variation (Zatorre & Belin, 2001) can be observed relatively easily, whereas a complementary sensitivity on the left to “temporal” or shorter-duration information is far more elusive (Boemio et al., 2005; Schönwiesner et al., 2005).

There remains the additional challenge that several of the more influential studies claiming hemispheric asymmetries have been based on responses to nonspeech sounds—it is not clear how these findings could be easily extrapolated to the question of natural speech processing. In response to this problem, the aim of the current fMRI study was to investigate the neural responses to acoustic modulations that are necessary and sufficient

for speech intelligibility; that is, modulations of amplitude and spectrum. By “intelligible,” we mean speech that can be fully recognized and understood: it is a term encompassing the phonetic, syntactic, and semantic representations and processes that contribute to speech comprehension.

Although no one acoustic cue determines the intelligibility of speech (Lisker, 1986), Remez, Rubin, Pisoni, and Carrell (1981) demonstrated that sentences comprising sine waves tracking the main formants (with the amplitude envelope intact) can be intelligible. This indicates that the dynamic amplitude and spectral characteristics of the formants in speech are sufficient to support speech comprehension. In the current experiment, we generated a 2×2 array of unintelligible conditions in which speech-derived modulations of formant frequency and amplitude were absent, applied singly or in combination, to explore neural responses to these factors and the extent to which any such responses are lateralized in the brain. To assess responses to intelligibility, we employed two dually modulated conditions—an unintelligible condition (forming part of the 2×2 array above) in which spectral and amplitude modulations came from two different sentences and an intelligible condition with matching spectral and temporal modulations that listeners could understand after a small amount of training. Importantly, naive subjects report hearing both of these conditions as sounding “like someone talking” (Rosen, Wise, Chadha, Conway, & Scott, in press) and that the unintelligible versions could not quite be understood. This lack of a strong low-level perceptual difference between the two conditions ensured that any neural difference would not result from any attentional imbalance, which may occur when people hear an acoustic condition that they immediately recognize as unintelligible.

A previous study in PET using the same stimulus manipulations (Rosen et al., in press) identified bilateral activation in left and right superior temporal cortex in response to acoustic modulations in the unintelligible conditions. The largest peak, in the right STG, showed a trend toward an additive response to the combination of spectral and amplitude modulations. In contrast, the comparison of intelligible and unintelligible condition generated peak activations in left STS. On the basis of these findings, the authors rejected the claim that specialized acoustic processing underlies the left hemisphere advantage for speech comprehension. However, the practical considerations of PET meant that this study was limited in statistical power and design flexibility, and the authors were not able to statistically compare responses in the left and right hemispheres.

Neuroimaging research on speech has recently seen increasing use of multivariate pattern analysis (MVPA; Okada et al., 2010; Hickok, Okada, & Serences, 2009; Formisano, De Martino, Bonte, & Goebel, 2008). In the current study, we employed univariate and MVPA approaches, the latter specifically to compare the ability

of left and right temporal regions to classify stimuli according to their differences in acoustic properties and intelligibility.

METHODS

Participants

Twenty right-handed speakers of English (10 women; mean age = 25 years, range 19–35 years) took part in the study. All the participants had normal hearing and no history of neurological problems or difficulties with speech or language (self-reported). All were naive about the aims of the experiment and unfamiliar with the stimuli. The study was approved by the University College London Department of Psychology Ethics Committee.

Materials

All stimuli were based on sine wave versions of simple sentences. The stimuli were derived from a set of 336 semantically and syntactically simple sentences known as the Bamford–Kowal–Bench sentences (e.g., *The clown had a funny face*; Bench, Kowal, & Bamford, 1979). These were recorded in an anechoic chamber by an adult male speaker of Standard Southern British English (Brüel & Kjaer 4165 microphone, Naerum, Denmark, digitized at a 11.025-kHz sampling rate with 16-bit quantization).

The stimuli were based on the first two formant tracks only, as these were found to be sufficient for intelligibility (Rosen et al., in press). A semiautomatic procedure was used to track the frequencies and amplitudes of the formants every 10 msec. Further signal processing was conducted off-line in MATLAB (The Mathworks, Natick, MA). The construction of stimulus conditions followed a 2×2 design with factors spectral complexity (formant frequencies modulated vs. formants static) and amplitude complexity (amplitude modulated vs. amplitude static). To provide formant tracks that varied continuously over the entire utterance (e.g., such that they persisted through consonantal closures), the formant tracks were interpolated over silent periods using piecewise-cubic Hermite interpolation in log frequency and linear time. Static formant tracks were set to the median frequencies of the measured formant tracks, separately for each formant track. Similarly, static amplitude values were obtained from the median of the measured amplitude values larger than zero.

Five stimulus conditions were created, where S and A correspond to “spectral” and “amplitude” modulation, respectively. The subscript “ \emptyset ” indicates a steady/fixed state, whereas “mod” indicates a dynamic/modulated state.

- (i) $S_{\emptyset}A_{\emptyset}$, steady state formant tracks with fixed amplitude.
- (ii) $S_{\emptyset}A_{\text{mod}}$, steady state formant track with dynamic amplitude variation.
- (iii) $S_{\text{mod}}A_{\emptyset}$, dynamic frequency variation with fixed amplitude.

(iv) $S_{\text{mod}}A_{\text{mod}}$, dynamic frequency and amplitude variation but each coming from a different sentence, making the signal effectively unintelligible. Linear time scaling of the amplitude contours was performed as required to account for the different durations of the two utterances.

(v) $\text{int}S_{\text{mod}}A_{\text{mod}}$, the intelligible condition with dynamic frequency and amplitude variation taken from the same original sentence. These were created in the same way as (i)–(iv), but with less extensive hand correction (the interpolations for the unintelligible condition, $S_{\text{mod}}A_{\text{mod}}$, were particularly vulnerable to small errors in formant estimation because of the modulations being combined from different sentences).

Static formant tracks and amplitude values were set at the median frequency of the measured formants and amplitude values larger than zero, respectively.

Each stimulus was further noise-vocoded (Shannon et al., 1995) to enhance auditory coherence. For each item, the input waveform was passed through a bank of 27 analysis filters (sixth-order Butterworth) with frequency responses crossing 3 dB down from the pass-band peak. Envelope extraction at the output of each analysis filter was carried out using full-wave rectification and second-order Butterworth low-pass filtering at 30 Hz. The envelopes were then multiplied by a white noise, and each filtered by a sixth-order Butterworth IIR output filter identical to the analysis filter. The rms level from each output filter was set to be equal to the rms level of the original analysis outputs. Finally, the modulated outputs were summed together. The cross-over frequencies for both filter banks (over the frequency range of 70–5000 Hz) were calculated using an equation relating position on the basilar membrane to its best frequency (Greenwood, 1990). Figure 1 shows sample spectrograms from each of the five conditions (with auditory examples in the Supplementary Material).

The intelligibility of the modulated stimuli (i.e., excluding the $S_{\emptyset}A_{\emptyset}$ condition) was tested in 13 adult listeners by Rosen et al. (in press), using 10 items from each condition. The mean intelligibility scores were 61%, 6%, 3%, and 3% keywords correct for the $\text{int}S_{\text{mod}}A_{\text{mod}}$, $S_{\text{mod}}A_{\text{mod}}$, $S_{\text{mod}}A_{\emptyset}$, and $S_{\emptyset}A_{\text{mod}}$ conditions, respectively.

Design and Procedure

Behavioral Pretest

A behavioral test session was used to familiarize and train the participants with the $\text{int}S_{\text{mod}}A_{\text{mod}}$ condition. This ensured that all participants would be in “speech mode” during the scanning session (Dehaene-Lambertz et al., 2005)—that is, that they would actively listen for stimuli that they could understand.

Participants were informed that there would be a training phase to help them understand some of the stimuli they would hear in the scanner. They were then tested on

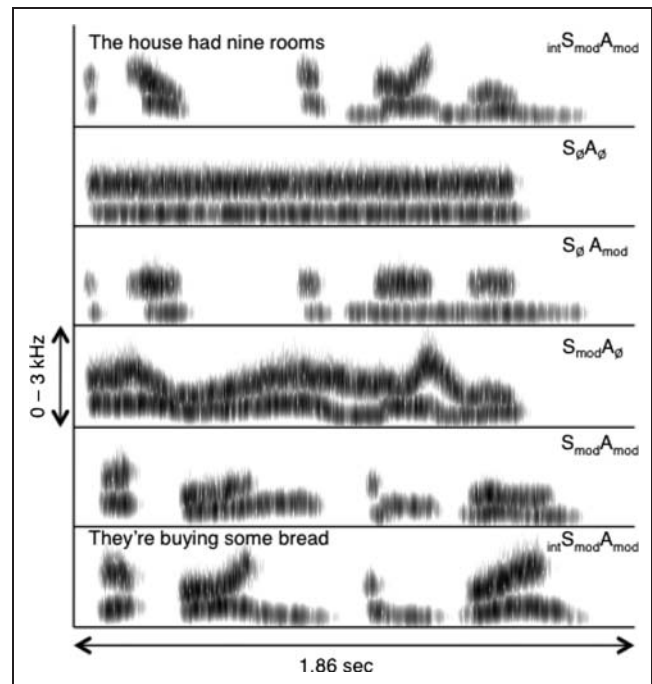


Figure 1. Example spectrograms from the five auditory conditions used in the experiment. The unintelligible condition $S_{\text{mod}}A_{\text{mod}}$ in the example was constructed using spectral modulations from “The house had nine rooms” and the amplitude envelope from “They’re buying some bread.” Darker shading indicates portions of greater intensity in the signal.

sentence report accuracy with items from the $\text{int}S_{\text{mod}}A_{\text{mod}}$ condition. A sentence was played for the participant over Sennheiser HD201 headphones (Sennheiser U.K., High Wycombe, Buckinghamshire, U.K.) and she or he was asked to repeat whatever she or he heard. Performance was graded according to the number of keywords the participant correctly identified. Each sentence had three key words. If the subject identified all three words, the tester provided positive feedback and moved on to the next sentence. If the participant was not able to identify one or more of the key words, the tester verbally repeated the sentence to the participant and played it again. This process was continued until the participant correctly repeated all the key words in five consecutive sentences or until 98 sentences were presented.

fMRI Experiment

Functional imaging data were acquired on a Siemens Avanto 1.5-T scanner (Siemens AG, Erlangen, Germany) with a 32-channel birdcage headcoil (which has been shown to significantly enhance signal-to-noise ratio for fMRI in the 1.5-T field; Parikh et al., 2011; Fellner et al., 2009). There were two runs of 150 echo-planar whole-brain volumes (repetition time = 9 sec, acquisition time [TA] = 3 sec, echo time = 50 msec, flip angle = 90°, 35 axial slices, 3 mm × 3 mm × 3 mm in-plane resolution).

A sparse-sampling routine was employed (Edmister, Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999), in which two stimuli from the same condition were presented sequentially during the silent period, with the onset of the first stimulus occurring presented 5.3 sec (with jittering of ± 500 msec) before acquisition of the next scan commenced.

In the scanner, auditory stimulation was delivered using MATLAB with the Psychophysics Toolbox extension (Brainard, 1997) via an amplifier and air conduction headphones (Etymotic, Inc., Elk Grove Village, IL) worn by the participant. In each functional run, the participant heard 50 stimuli from each of the five auditory conditions (two stimuli per trial). For the four unintelligible conditions, these 50 items were repeated in the second functional run. There were 100 distinct sentences in the $intS_{mod}A_{mod}$ condition (50 in each run). Participants were instructed to listen carefully to all the stimuli, with their eyes closed. They were told that they would hear some examples of the same sort used in the training phase, which they should try to understand. The stimuli were pseudorandomized to allow a relatively even distribution of the conditions without any predictable ordering effects. A silent baseline was included in the form of four miniblocks of five silent trials in each functional run. After the functional runs, a high-resolution T1-weighted anatomical image was acquired (HIREs MP-RAGE, 160 sagittal slices, voxel size = 1 mm^3).

Behavioral Posttest

The posttest comprised 80 sentences from the $intS_{mod}A_{mod}$ condition. Half of the items had been presented in the scanner, and half were novel exemplars. After each sentence, participants were asked to repeat what they heard. Speech perception accuracy was scored on-line, according to the number of key words correctly reported.

Analysis of fMRI Data

Univariate Analysis

Data were preprocessed and analyzed in SPM5 (Wellcome Trust Centre for Neuroimaging, London, UK.). Functional images were realigned and unwarped, coregistered with the anatomical image, normalized (with a voxel size of 3 mm^3) to Montreal Neurological Institute (MNI) stereotactic space using parameters obtained from a unified segmentation of the T1 anatomical image and smoothed using a Gaussian kernel of 8 mm FWHM. Event-related responses for each condition were modeled as a canonical hemodynamic response function, with event onsets modeled from the acoustic onset of the first auditory stimulus in each trial and with durations of 4 sec (the approximate duration of two sequential stimuli). For each session separately, each condition was modeled as a separate regressor in a general linear model. Six movement parameters (three translations and

three rotations) were included as regressors of no interest. At the first level (single-subject), con images were created to describe the main effect of amplitude modulation $[-1\ 1\ -1\ 1]$, the main effect of spectral modulation $[-1\ -1\ 1\ 1]$, and the interaction of the two factors $[-1\ 1\ 1\ -1]$. A contrast of intelligibility compared the two dually modulation conditions ($intS_{mod}A_{mod} > S_{mod}A_{mod}$). Second-level group analyses were carried out with separate one-sample *t* tests on the con images using SPM5 (the main effects and interaction were explored as *F* contrasts, and the intelligibility effect was described as a unidirectional *t* contrast). All second-level contrasts were thresholded at $p < .05$ (voxel-wise; family-wise-error corrected). Coordinates of peak activations were labeled using the SPM5 anatomy toolbox (Eickhoff et al., 2005). For activation plots, parameter estimates were extracted from spherical ROIs (4-mm radius) built around peak voxels using the MarsBaR toolbox in SPM (Brett et al., 2002).

Multivariate Analysis

Functional images were unwarped and realigned to the first acquired volume using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK.) Training and test examples from each condition were constructed from single volumes. The data were separated into training and test sets by functional run to ensure that training data did not influence testing (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). Linear and quadratic trends were removed, and the data *z*-scored within each run. A linear support vector machine (SVM) from the Spider toolbox (www.kyb.tuebingen.mpg.de/bs/people/spider/) was used to train and validate models (for background on SVMs, see Supplementary Methods and Results). The SVM used a hard margin and the Andre optimization. For each participant, the first classifier was trained on the first run and tested on the second, and vice versa, for the second classifier. For each participant, the overall performance for each classification was calculated by averaging the performance across the two classifiers. Three acoustic classifications compared the unintelligible modulated conditions with the unmodulated ($S_{\emptyset}A_{\emptyset}$) condition: $S_{\emptyset}A_{\emptyset}$ vs $S_{\emptyset}A_{mod}$, $S_{\emptyset}A_{\emptyset}$ versus $S_{mod}A_{\emptyset}$ and $S_{\emptyset}A_{\emptyset}$ versus $S_{mod}A_{mod}$. A further acoustic classification was run on the singly modulated conditions ($S_{\emptyset}A_{mod}$ vs $S_{mod}A_{\emptyset}$) to assess the discriminability of spectral versus amplitude modulations. Finally, an intelligibility classification run for $intS_{mod}A_{mod}$ vs $S_{mod}A_{mod}$. The classifications were performed for a number of subject-specific, anatomically defined ROIs. The Freesurfer image analysis suite (surfer.nmr.mgh.harvard.edu/) was used to perform cortical reconstruction and volumetric segmentation via an automated cortical parcellation of individual T1 images (Destrieux, Fischl, Dale, & Halgren, 2010). This generated subject-specific, left hemisphere, right hemisphere, and left-right combined anatomical ROIs for HG, middle

temporal gyrus (MTG) + STG (generated from parcellation of MTG, STG, and STS), and left and right hemisphere inferior occipital gyrus (IOG; included as a control site). These anatomically defined regions were included based on a priori hypotheses about key sites for intelligibility and acoustic processing of speech (Eisner et al., 2010; Obleser et al., 2007, 2008; Scott, Rosen, Lang, & Wise, 2006; Davis & Johnsrude, 2003; Scott et al., 2000), hence not contingent on the univariate results. See Supplementary Figure 1 for an example parcellation from the current study.

RESULTS

Behavioral Tests

Pretest

A stringent criterion of five consecutive correct responses (with 100% accuracy on keyword report) was used to ensure thorough prescan training on the $_{int}S_{mod}A_{mod}$ condition. For those participants who reached this criterion in the pretest, the mean number of trials to criterion was 46.6 ($SD = 17.7$). Four of the 20 fMRI participants did not reach this criterion within the list of 98 pretest items. However, as all participants achieved three consecutive correct responses within an average of only 23.5 trials ($SD = 16.4$), we were satisfied that all participants would understand a sufficient proportion of $_{int}S_{mod}A_{mod}$ items in the scanner to support planned intelligibility contrasts.

Posttest

The average accuracy across the whole posttest item set (calculated as the percentage of keywords correctly reported) was 67.2% ($SD = 8.5\%$), representing a mean improvement of 4.6% ($SD = 5.7\%$) on pretest scores (mean = 62.4%, $SD = 6.4\%$). This improvement was statistically significant ($t(19) = 3.615, p < .01$). There was no difference in accuracy between the old (67.0%) and new (67.5%) items ($p > .05$).

fMRI

Univariate Analysis

Main effects of acoustic modulations. Figure 2A shows the main effects of Spectral Modulation and Amplitude Modulation, as well as plots of parameter estimates of the four unintelligible conditions (compared with the silent baseline), taken from peak activation sites. For both contrasts, there was greater signal in bilateral mid-STG when the modulations were present than when they were absent. Activations in STG were larger in statistical height and spatial extent for the main effect of Spectral Modulations and included additional clusters in posterior

left STG and left precentral gyrus that also showed an enhanced response to the modulations. A single cluster, with its peak in left inferior frontal gyrus (IFG; pars triangularis) but the majority of its extent on middle frontal gyrus, showed less signal when amplitude modulation was present than when it was absent. Significant peak and subpeak voxels (more than 8 mm apart) for the main effects are listed in Table 1.

Interaction: Spectral \times Amplitude Modulation. We observed activation in bilateral mid-STG, with the overall peak in the left hemisphere ($[-63 -18 6]$). However, the plots of contrast estimates from the main peaks indicate a subadditivity of the two factors, that is, the difference in signal between $S_{\emptyset}A_{\emptyset}$ and the singly modulated conditions ($S_{\emptyset}A_{mod}$ or $S_{mod}A_{\emptyset}$) was larger than that between those singly modulated conditions and the $S_{mod}A_{mod}$ condition. These activations are shown in Figure 2B and are listed in Table 1.

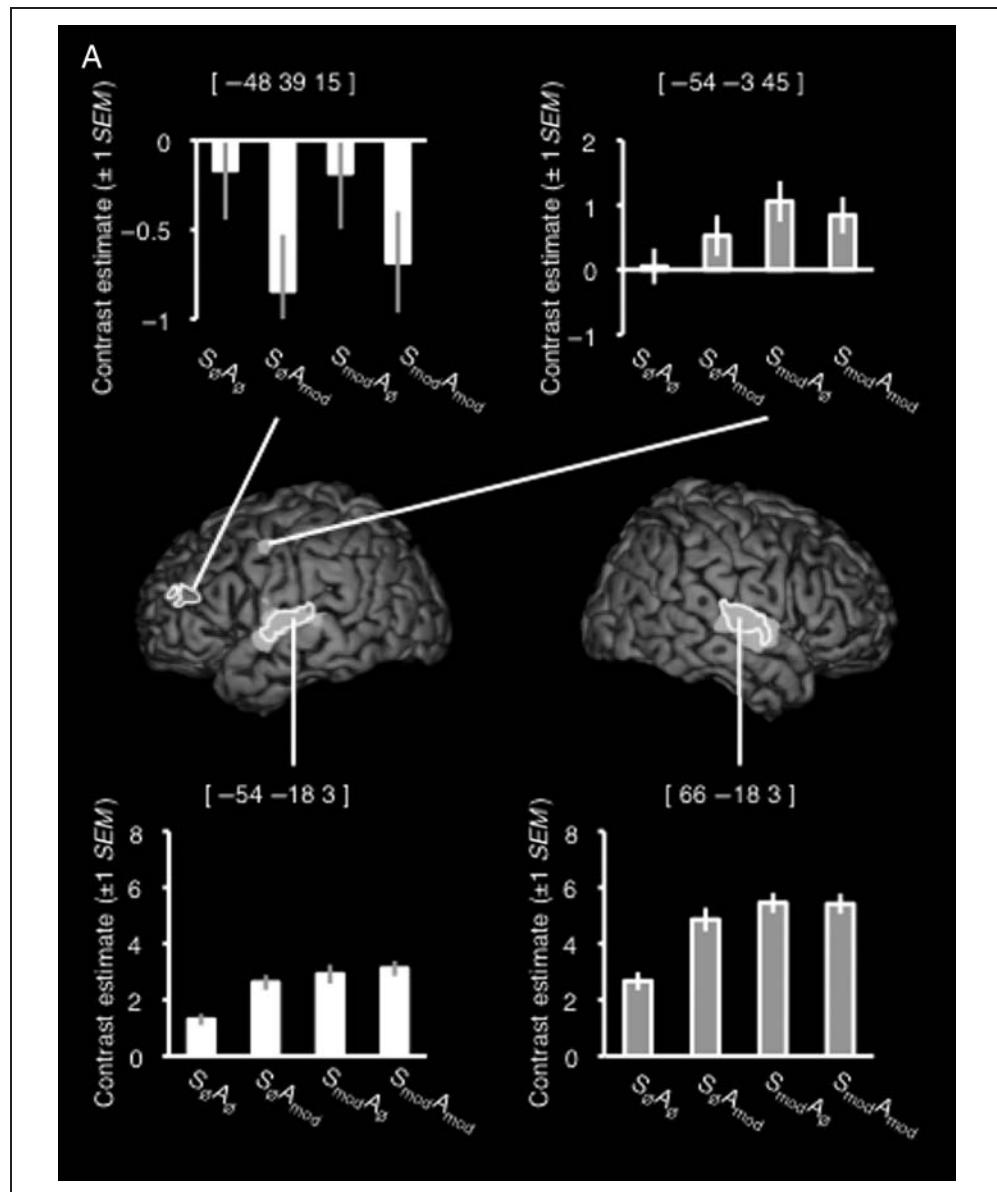
Effect of intelligibility. The contrast $_{int}S_{mod}A_{mod} > S_{mod}A_{mod}$ gave significant activation in bilateral STS and STG extending to the temporal pole, with the peak voxel in the left STS and a larger cluster extent in the left hemisphere. A single cluster of activation with its peak in the pars orbitalis of the IFG extended medially to the left anterior insula. There was also activation in a medial portion of the right IFG. Figure 3 shows the results of this contrast, with plots of contrast estimates for all five conditions compared with rest—significant peaks and subpeaks are listed in Table 2.

Multivariate Pattern Analysis

Two participants were excluded from the multivariate analyses because of unsuccessful cortical parcellation.

Acoustic classifications. Figure 4A–C shows boxplots of group classification accuracy by ROI for each of the classifications $S_{\emptyset}A_{\emptyset}$ versus $S_{\emptyset}A_{mod}$, $S_{\emptyset}A_{\emptyset}$ versus $S_{mod}A_{\emptyset}$, and $S_{\emptyset}A_{\emptyset}$ versus $S_{mod}A_{mod}$. Performance in each classification was tested against a chance performance of 0.5, using the one-sided Wilcoxon signed rank test,¹ with a corrected significance level of $p < .008$ (correcting for six ROIs in each classification). All temporal ROIs performed significantly better than chance ($p < .001$). The left IOG, which was included as a control site, performed significantly better than chance in the $S_{\emptyset}A_{\emptyset}$ versus $S_{mod}A_{mod}$ classification (signed rank statistic $w = 21, p = .0043$), although still quite poorly (median: 54%). For all other classifications, the IOG performed no better than chance. A second analysis comparing left and right homologs of each ROI showed that performance was equivalent between the left and right hemispheres for all three classifications, for all ROI pairs ($p > .017$ —significance level corrected

Figure 2. (A) Activation peaks and extents for the main effect of spectral modulation (gray) and amplitude modulation (white). Plots show the contrast estimates (± 1 SEM) for each condition taken from the peak voxel in each contrast. (B) Activation extent for the interaction of spectral and amplitude modulation. Plots show the contrast estimates (± 1 SEM) for each condition taken from ROIs (4-mm radius) built around local peaks. All images are shown at a corrected (family-wise error) height threshold of $p < .05$ and a cluster extent threshold of 40 voxels. Coordinates are given in MNI stereotactic space.



for three left–right comparisons in each classification; paired, two-sided Wilcoxon signed rank tests).²

To compare processing of spectral and amplitude modulations directly, a fourth acoustic classification of S_0A_{mod} versus $S_{mod}A_0$ was performed on the temporal ROIs only (as the IOG showed chance performance in classifying these conditions against S_0A_0). Figure 4D shows boxplots of the group performance, which was significantly better than chance in all ROIs (one-sided Wilcoxon signed rank tests, corrected significance level, $p < .017$: left HG $p < .005$; right HG, left STG + MTG, right STG + MTG $p < .001$). However, there were no differences between hemispheres ($p > .025$; paired, two-sided Wilcoxon signed rank).

It was possible that there may still be within-hemisphere preferences for one modulation type. Therefore, a further statistical comparison was made within hemispheres for

the classifications S_0A_0 versus S_0A_{mod} and S_0A_0 versus $S_{mod}A_0$ in left HG, right HG, left STG + MTG, and right STG + MTG (paired, two-sided Wilcoxon signed rank tests with corrected significance level $p < .013$; Figure 5). This showed that the classification of spectral modulations was significantly more accurate than the classification of amplitude modulations in left HG ($w = 20.5, p = .005$), right HG ($w = 14, p = .002$), and right STG + MTG ($w = 26, p = .010$). The difference in left STG + MTG was significant at an uncorrected alpha of .05 ($w = 30.5, p = .017$).

An alternative way to assess hemispheric differences in the multivariate data is to extract classifier weights from bilateral ROIs and explore their distribution across the two hemispheres (see Supplementary Material for an explanation of the weight vector). The top 30% positive and negative weights (according to magnitude and including only those in the top 30% for both cross-validated runs)

Figure 2. (continued)

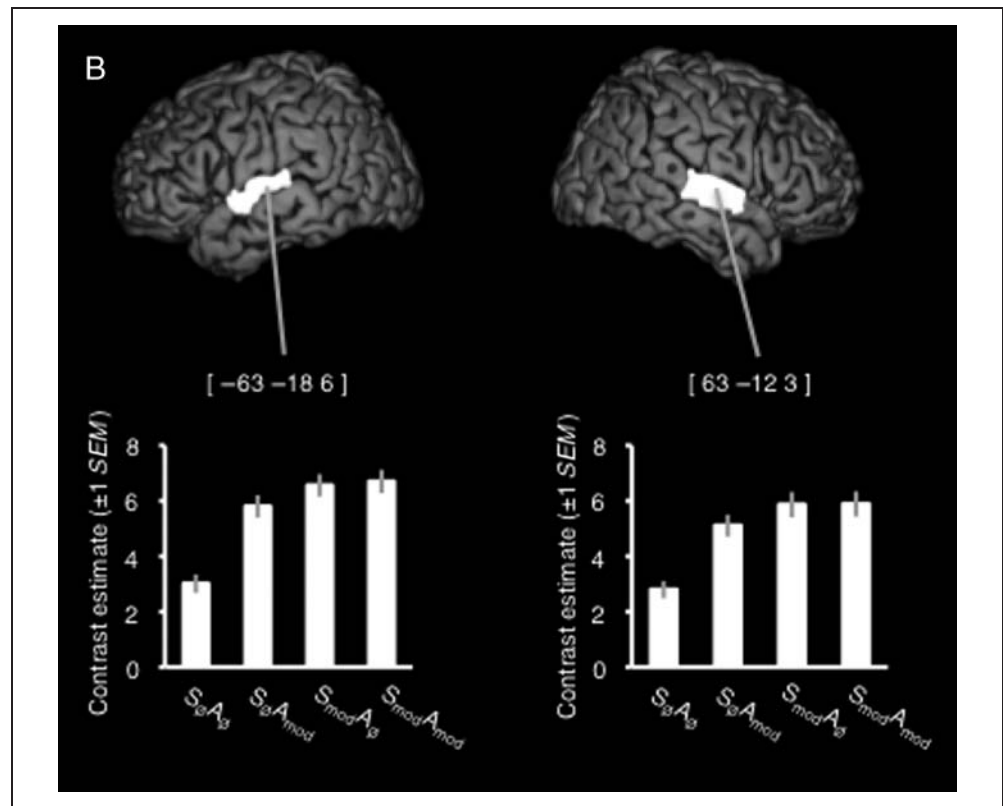
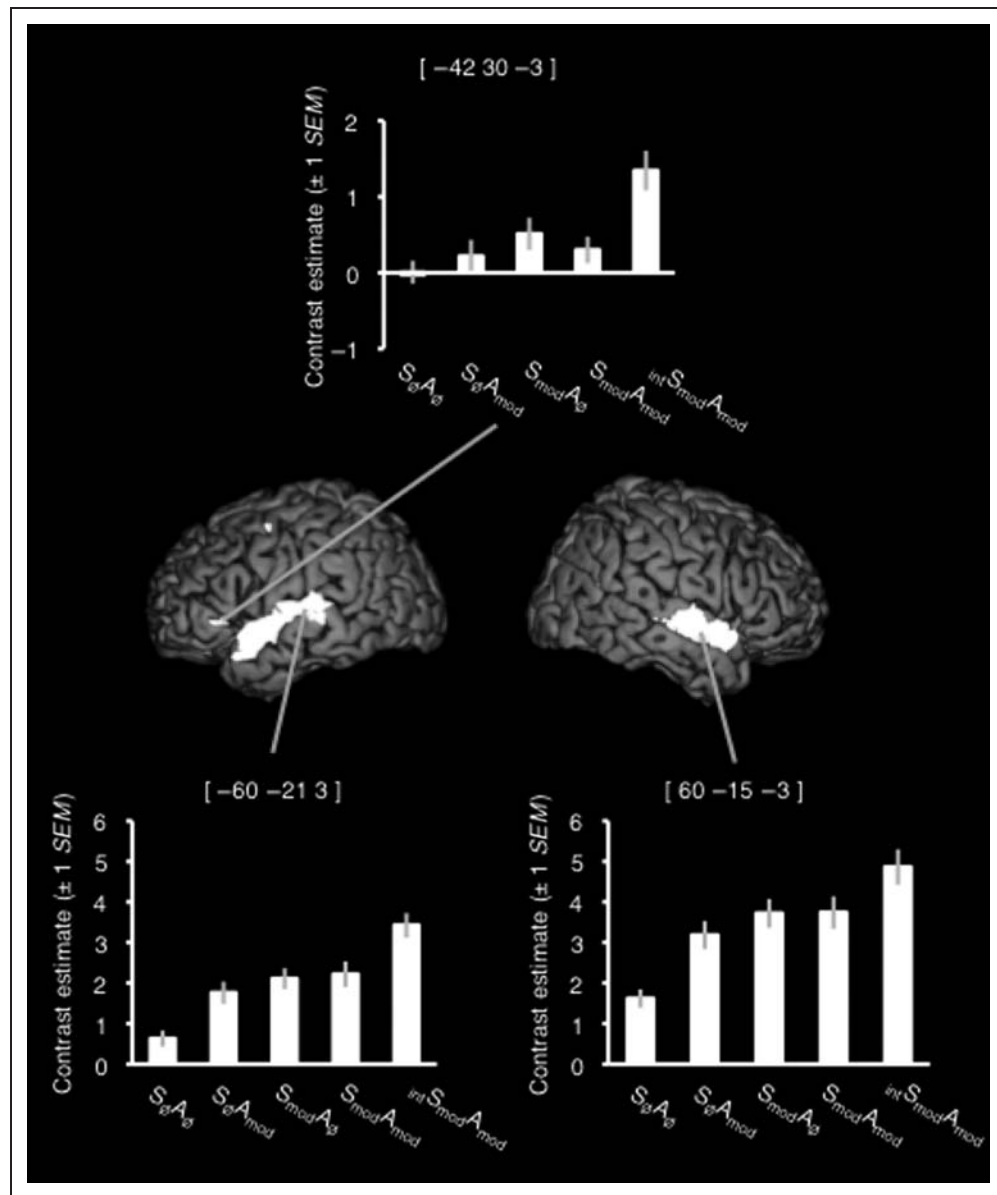


Table 1. Peak and Subpeak (if More Than 8 mm Apart) Activations from the Contrasts of Acoustic Effects in the Univariate Analysis

Contrast	No. of Voxels	Region	Coordinates				
			x	y	z	F	z
Main effect of Spectral Modulation	149	Right STG	66	-18	3	207.65	6.69
	175	Left STG	-60	-12	3	188.21	6.56
			-66	-33	6	75.63	5.34
	2		-48	-39	18	62.80	5.08
Main effect of Amplitude Modulation	4		-54	-3	45	57.65	4.96
	89	Left STG	-54	-18	3	158.12	6.34
			-60	-12	3	136.10	6.14
	48	Right STG	63	-12	3	108.79	5.84
Interaction of Spectral and Amplitude Modulation			63	-21	9	73.54	5.30
			63	-3	-6	58.42	4.97
	10	Left IFG (pars triangularis)	-48	39	15	66.62	5.16
	117	Left STG	-63	-18	6	192.77	6.59
			-57	-6	-3	125.07	6.03
	112	Right STG	63	-12	3	161.00	6.36
			66	-27	6	89.96	5.58
			57	0	-6	59.43	5.00

Figure 3. Activation in the contrast of $intS_{mod}A_{mod} > S_{mod}A_{mod}$. Plots show the contrast estimates ($\pm 1 SEM$) for each condition taken from ROIs (4-mm radius) built around local peaks. The image is shown at a corrected (family-wise error) height threshold of $p < .05$ and a cluster extent threshold of 40 voxels. Coordinates are given in MNI stereotactic space.



within bilateral HG and STG + MTG ROIs were extracted from the acoustic contrast $S_{mod}A_{\emptyset}$ versus $S_{\emptyset}A_{mod}$. By examining the classifier weights, we could ascertain which voxels contributed most to the classification and whether these voxels exhibited a relative increase in signal to spectral or amplitude modulation in the support vectors. Classification in both ROIs was significantly better than chance (one-sided signed rank Wilcoxon; bilateral STG + MTG: $w = 0$, $p < .001$; bilateral HG: $w = 27.5$, $p < .01$). Weights were visualized for both STG + MTG and HG ROIs in native space (see Figure 6 for weights in five representative participants). Negative weights (red) represent an increase in signal to $S_{mod}A_{\emptyset}$, and positive weights (blue) represent an increase to $S_{\emptyset}A_{mod}$. Both weight categories appeared well distributed within and between the hemispheres, suggesting a lack of hemispheric preference for modulation type. This was confirmed by comparing the number of positive and negative

weights within each hemisphere—using a two-sided signed rank test, there was no significant difference for the HG or the STG + MTG ROIs ($p > .05$; Figure 7).

Intelligible versus unintelligible stimuli. Figure 8 shows boxplots of group classification accuracy by ROI for the classification of $intS_{mod}A_{mod}$ versus $S_{mod}A_{mod}$ stimuli. Performance was significantly better than chance for the left and right STG + MTG ROIs (both: $w = 0$, $p < .0001$; one-tailed Wilcoxon signed rank tests, corrected significance level $p < .008$). The left IOG performed poorly at 54% but this was significantly better than chance ($w = 29.5$, $p = .0073$). Performance in left and right HG did not survive the corrected threshold of $p < .008$ but was significant at an uncorrected level of $p < .05$ in both hemispheres (left: $w = 21$, $p = .023$; right: $w = 42.5$, $p = .030$). All other ROIs performed at chance in this classification. The comparison of left and

right homologs of each ROI showed that performance was equivalent between the left and right hemispheres for HG and IOG but was significantly greater in the left STG + MTG than in its right hemisphere homolog ($w = 24.5$, $p = .014$; paired, two-sided Wilcoxon signed rank tests, corrected significance level of $p < .017$).

Previous data offer no basis to hypothesize that the IOG would be involved in speech processing. Post hoc inspection of the univariate analysis revealed IOG activation, at a lowered threshold, for the contrast $S_{\text{mod}}A_{\text{mod}} > \text{int}S_{\text{mod}}A_{\text{mod}}$. It is possible that this region is involved in default-network processes and the observed activation profiles reflect task-related deactivation of this region.

The top 30% positive and negative classifier weights were extracted from the intelligibility contrast using a bilateral STG + MTG mask. The classification was highly significant ($w = 0$, $p < .001$; one-sided Wilcoxon). Classifier weights were extracted within native space and are shown for five representative subjects in Figure 9 (violet = *intelligible*, yellow = *unintelligible*). Classifier weights characteristic of an increase in signal to intelligible and unintelligible sounds were well distributed within and across both hemispheres. However, when positive and negative weights were counted and compared, there was a significantly larger number of voxels characterizing a response to intelligible speech in the left ($w = 7$, $p < .001$) and a larger number responding to unintelligible sounds in the right ($w = 32$, $p = .0198$; Wilcoxon two-sided signed rank with corrected level $p < .025$; see Figure 10).

DISCUSSION

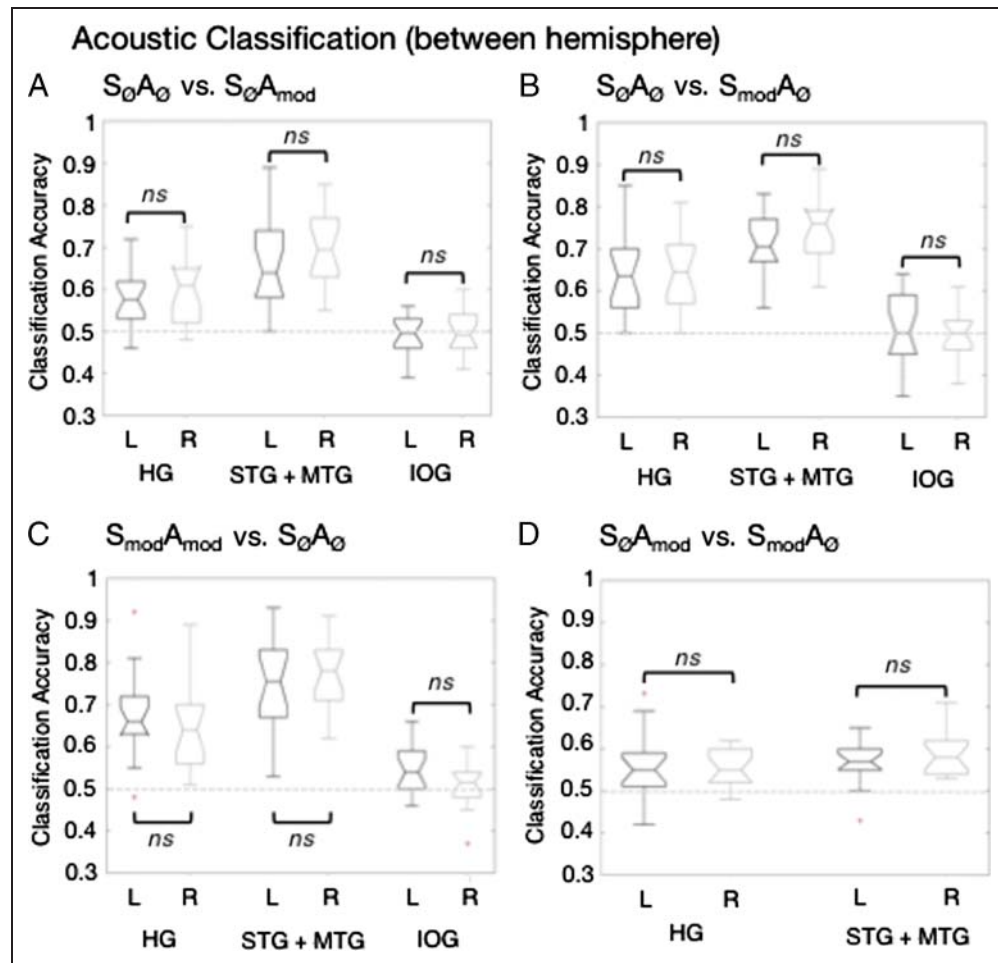
The current study demonstrates that the acoustic processing of speech-derived modulations of spectrum and amplitude generates bilaterally equivalent activation in superior temporal cortex, for unintelligible stimuli. It is only when the modulations generate an intelligible percept that a left dominant pattern of activation in STS/STG and frontal regions emerges. Using multivariate pattern analysis, we demonstrate statistical equivalence between the left and right hemispheres for the processing of acoustic modulations, but a significant left hemispheric advantage for the decoding of intelligibility in STG + MTG (incorporating the STS). The latter result supports the extensive clinical data associating damage to left hemisphere structures with lasting speech comprehension deficits and stands in contrast to recent work, making strong claims for bilateral equivalence in the representation of intelligibility in speech (e.g., Okada et al., 2010). Here we assume that intelligibility includes all stages in the comprehension of a sentence, over and above the early acoustic processing of the speech, and we stress that the intelligibility responses we see include the acoustic-phonetic, semantic, and syntactic processes and representations, which contribute to the comprehension of speech.

The neural correlates of spectral and amplitude modulations in the unintelligible conditions of the current experiment (examined as main effects) showed areas of considerable overlap in portions of the STG bilaterally.

Table 2. Peak and Subpeak (if More Than 8 mm Apart) Activations from the Contrast of Intelligibility in the Univariate Analysis

Contrast	No. of Voxels	Region	Coordinates			T	z
			x	y	z		
Intelligible > Unintelligible	251	Left STS	-60	-21	-3	13.01	6.53
			-57	-12	-3	11.79	5.27
			-54	0	-9	11.40	6.19
			-63	-33	6	9.35	5.66
			-51	9	-12	9.35	5.66
			-48	12	-21	8.32	5.34
	154	Right STS	60	-15	-3	11.00	6.09
			60	-6	-6	10.74	6.03
			54	9	-18	9.04	5.57
	28	Left IFG (pars orbitalis)	51	-27	3	7.34	4.99
			-42	30	-3	7.30	4.98
	3	Right IFG (pars triangularis)	-30	27	0	7.29	4.97
			33	30	3	7.26	4.96
4	Left precentral gyrus	-54	-51	9	6.99	4.86	
1	Left precentral gyrus	-48	0	51	6.52	4.67	

Figure 4. Box plots of group classification performance on (A) $S_{\emptyset}A_{\emptyset}$ versus $S_{\emptyset}A_{mod}$, (B) $S_{\emptyset}A_{\emptyset}$ versus $S_{mod}A_{\emptyset}$, (C) $S_{\emptyset}A_{\emptyset}$ versus $S_{mod}A_{mod}$, and (D) $S_{\emptyset}A_{mod}$ versus $S_{mod}A_{\emptyset}$ in the anatomically defined ROIs. Annotations indicate the result of pairwise comparisons (paired, two-sided Wilcoxon signed rank tests) of performance in left and right ROIs. * = significant at a corrected level of $p < .017$ (A–C) or $p < .025$ (D); L = left hemisphere; R = right hemisphere.



Previous studies exploring the neural correlates of amplitude envelope and spectral modulation have observed similar bilateral patterns of activation in the dorsolateral temporal lobes (Obleser et al., 2008; Boemio et al., 2005;

Hart et al., 2003; Langers et al., 2003). Inspection of the peak activations in Table 1 shows that the cluster extents and statistical heights of the local peaks are largely similar across the hemispheres. Rosen et al. (in press) found

Figure 5. Box plots of group classification performance for within-hemisphere comparison of Spectral and Amplitude modulation processing (i.e., $S_{\emptyset}A_{\emptyset}$ versus $S_{\emptyset}A_{mod}$ and $S_{\emptyset}A_{\emptyset}$ versus $S_{mod}A_{\emptyset}$). Annotations indicate the result of pairwise comparisons (paired, two-sided Wilcoxon signed rank tests) of performance in the two classifications. * = significant at a corrected level of $p < .013$; L = left hemisphere; R = right hemisphere.

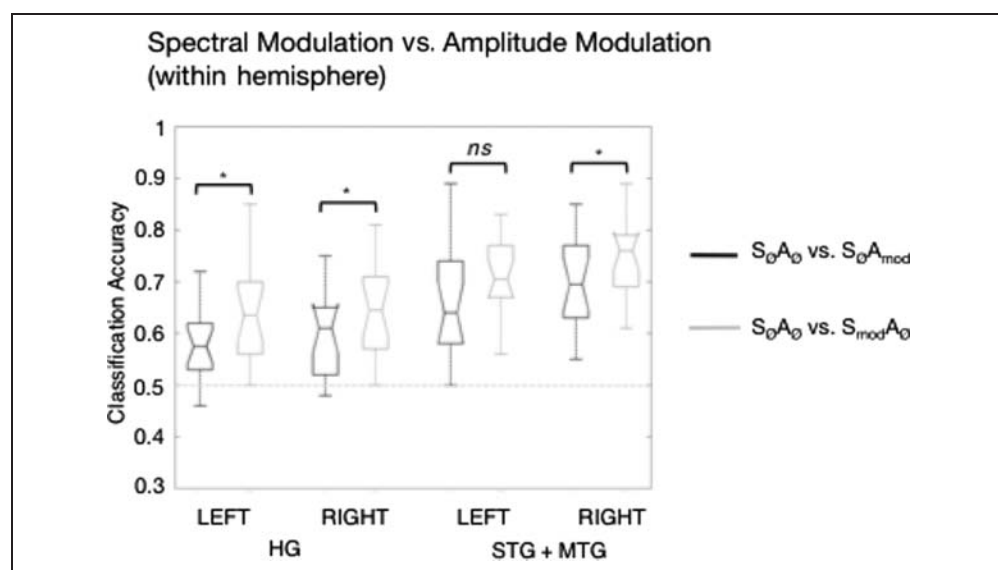
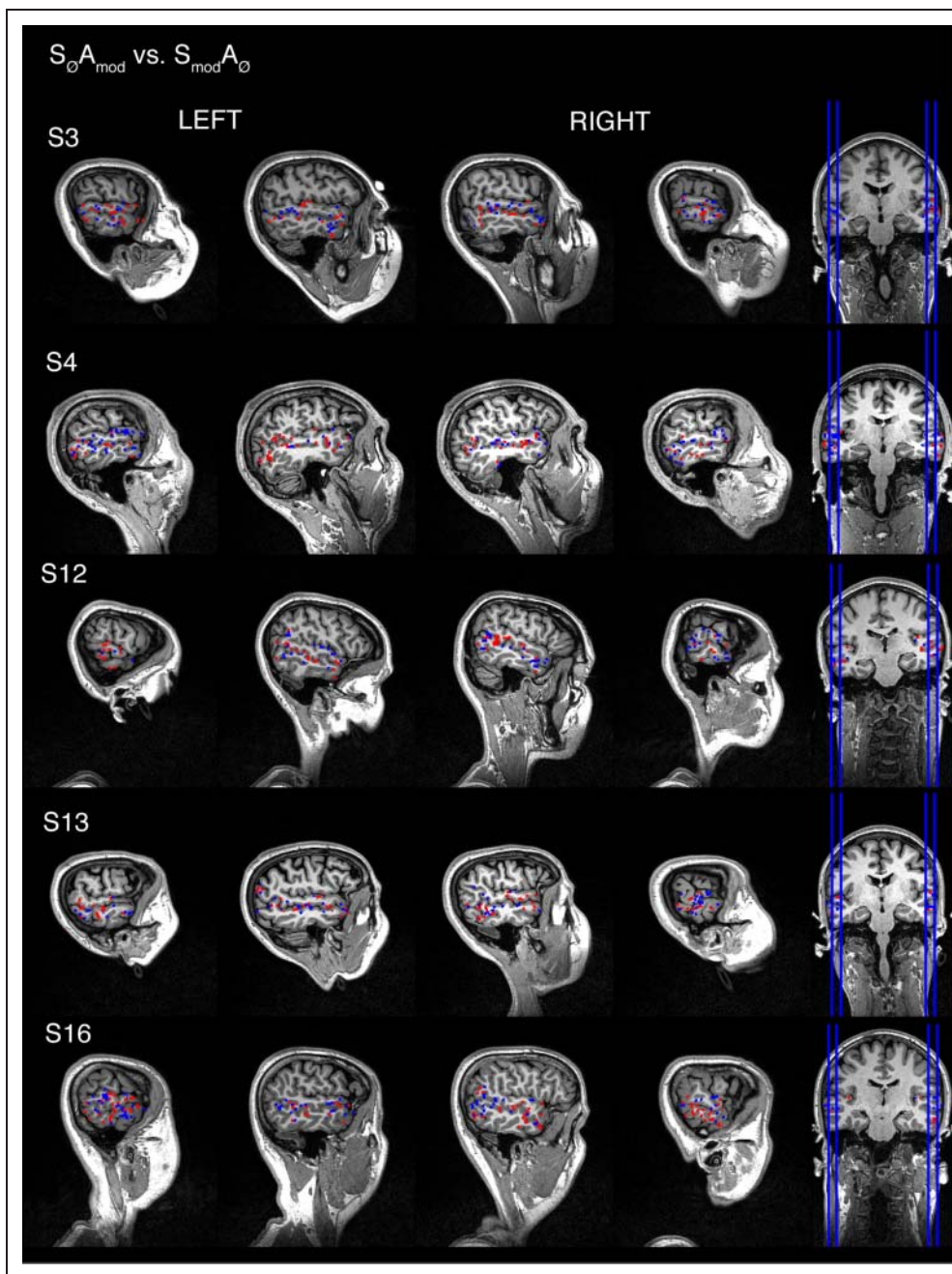


Figure 6. Classifier weights shown in native space for five representative subjects for the acoustic classification $S_{\emptyset}A_{\text{mod}}$ versus $S_{\text{mod}}A_{\emptyset}$. Red voxels = $S_{\emptyset}A_{\text{mod}}$; blue = $S_{\text{mod}}A_{\emptyset}$.

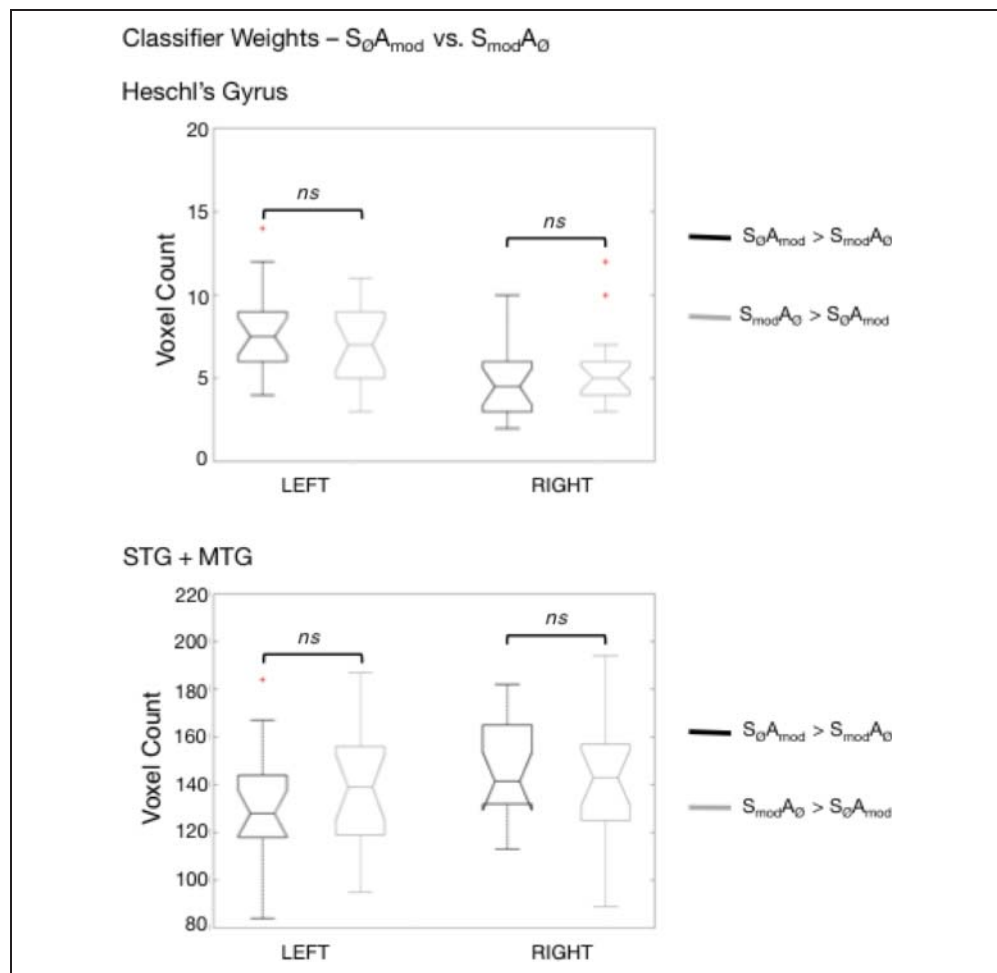


similar equivalence in the extent of activation between left and right STG for the processing of the unintelligible conditions. However, they found that a site on right STG indicated a strong additive profile, with a much greater response to the $S_{\text{mod}}A_{\text{mod}}$ condition than to the conditions with only one type of modulation ($S_{\emptyset}A_{\text{mod}}$ and $S_{\text{mod}}A_{\emptyset}$). The current study offered no evidence for a truly additive response anywhere in the regions responsive to acoustic modulations. This may reflect design differences. Rosen et al. employed a blocked design in PET, in which listeners were exposed to around 1 min of stimulation from a single condition during each scan. This may have allowed for a slow emergence of a greater response to the $S_{\text{mod}}A_{\text{mod}}$

condition than the immediate responses measured in the current, event-related design. Nonetheless, the univariate analyses in neither study offer support for a left-specific specialization for either amplitude or spectral modulations derived from natural speech.

Further investigation of the responses to acoustic modulation using multivariate pattern analysis offered no evidence of hemispheric asymmetries in the classification of unintelligible modulated stimuli from the basic $S_{\emptyset}A_{\emptyset}$ condition nor for classification of the two single-modulated conditions. Previous approaches to calculating laterality effects in functional neuroimaging data (Josse, Kherif, Flandin, Seghier, & Price, 2009; Obleser

Figure 7. Box plots of group voxel counts for the top 30% of positive and negative weights in acoustic classification $S_{\text{G}}A_{\text{mod}}$ versus $S_{\text{mod}}A_{\text{G}}$. Annotations indicate the result of pairwise comparisons (paired, two-sided Wilcoxon signed rank tests) of voxel counts in left and right ROIs. * = significant at a corrected level of $p < .013$ (four comparisons).



et al., 2008; Boemio et al., 2005; Schönwiesner et al., 2005; Zatorre & Belin, 2001) have included voxel counting (which is dependent on statistical thresholding), flipping the left hemisphere to allow subtractive comparisons with the right (which can be confounded by anatomical differences between the hemispheres), and use of ROIs of arbitrary size and shape (which are often generated non-independently, based on activations observed in the same study; furthermore, the sensitivity of the mean signal to subtle differences between conditions is compromised in large ROIs (Poldrack, 2007)). The SVM approach in the current study works well with large numbers of voxels because the SVMs minimize classification error whilst taking into account model complexity. This reduces overfitting, ensuring good generalization from training to test data. Indeed, SVMs have previously been shown to be generally robust to ROI size, showing similar levels of performance regardless of whether 300 or 3000 voxels are used (Misaki, Kim, Bandettini, & Kriegeskorte, 2010). In our study, the data were not subject to prior thresholding, and the use of anatomical ROIs avoided issues of arbitrariness in ROI size and shape. Thus, using improved methods for the detection of hemispheric

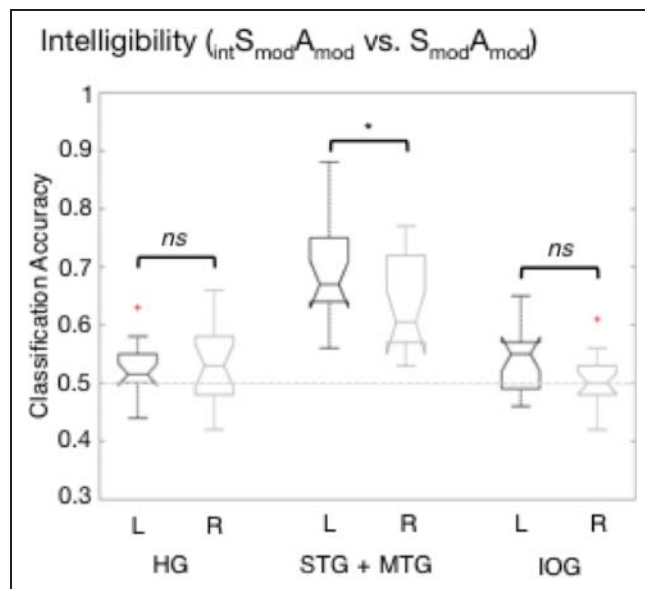
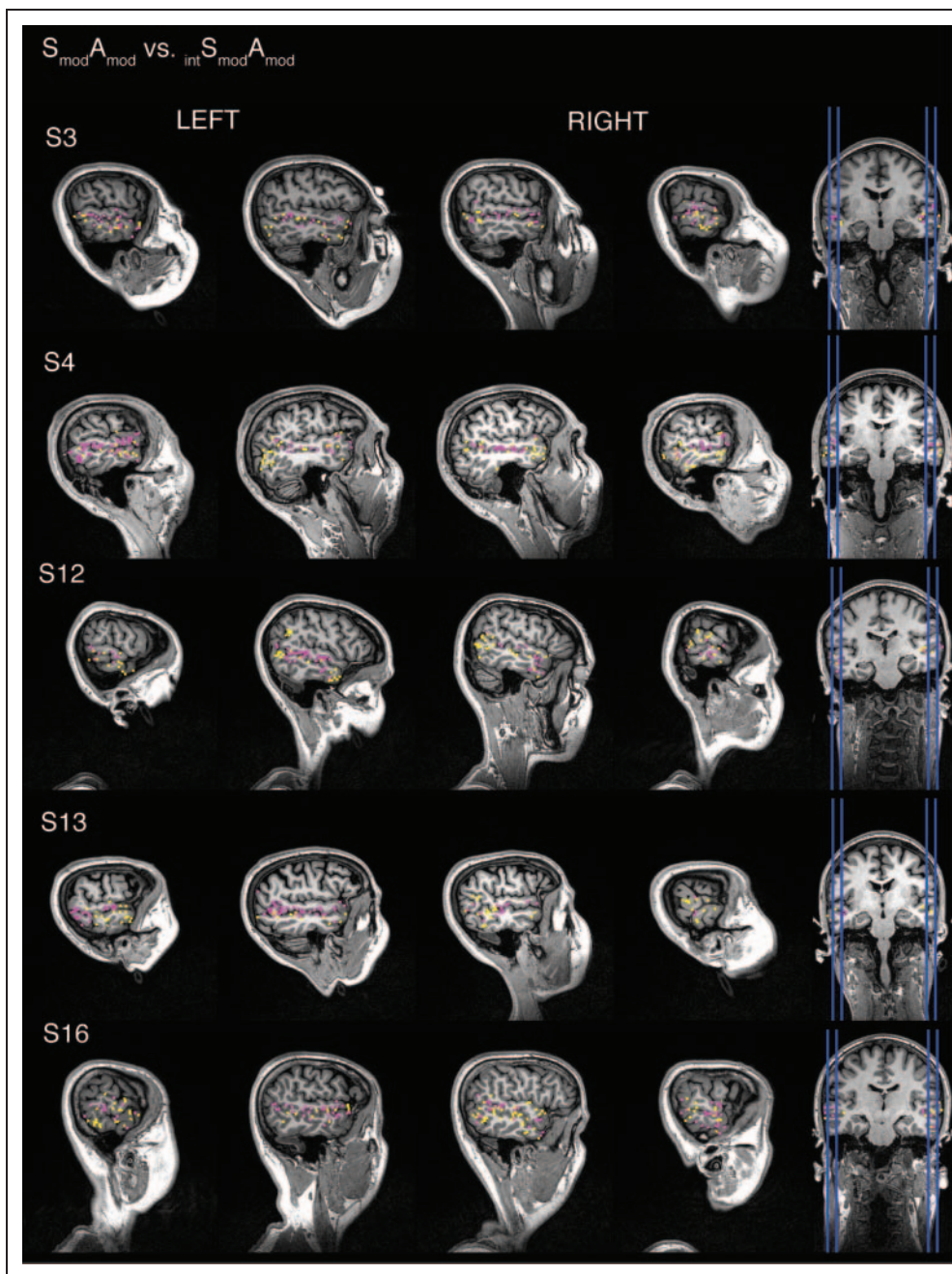


Figure 8. Box plots of group classification performance for the intelligibility contrast $\text{int} S_{\text{mod}}A_{\text{mod}}$ versus $S_{\text{mod}}A_{\text{mod}}$. Annotations indicate the result of pairwise comparisons (paired, two-sided Wilcoxon signed rank tests) of performance in the two classifications. * = significant at a corrected level of $p < .017$; L = left hemisphere; R = right hemisphere.

Figure 9. Classifier weights shown in native space for five representative subjects for the acoustic classification $S_{mod}A_{mod}$ versus $S_{int}S_{mod}A_{mod}$. Violet = $S_{int}S_{mod}A_{mod}$; blue = $S_{mod}A_{mod}$.



asymmetries, our findings stand in contrast to previous claims for subtle left–right differences in preference for temporal and spectral information (Boemio et al., 2005; Zatorre & Belin, 2001). Moreover, the significant within-ROI advantage for the classification of spectral modulations compared with amplitude modulations in bilateral HG and STG + MTG indicates that the lack of evidence for hemispheric asymmetries was not because of insufficient sensitivity in the MVPA.

Several neuroimaging studies have identified peak responses to intelligible speech in left STS that was either strongly lateralized activation (Eisner et al., 2010; Narain et al., 2003; Scott et al., 2000) or rather more bilaterally

distributed along both left and right STS (Awad, Warren, Scott, Turkheimer, & Wise, 2007; Scott et al., 2006; Davis & Johnsrude, 2003). In the former cases, the strong left lateralization was observed in direct subtractive contrasts with unintelligible control conditions that were well matched in complexity to the intelligible speech. However, in some studies, there were clear perceptual differences between the intelligible and unintelligible conditions. This may have made it easier for participants to ignore or attend less closely to those stimuli they knew to be unintelligible. In the current experiment, the intelligible and unintelligible stimuli were constructed to be very similar, acoustically and perceptually, and participants typically describe the

$S_{\text{mod}}A_{\text{mod}}$ stimuli as sounding like someone speaking, but with no sense of intelligibility (Rosen et al., in press). In the absence of clear perceptual differences between these two modulated conditions, listeners should have attended equivalently to stimuli from these two categories. Our univariate analysis revealed a left dominant response to intelligible speech, in STG/STS, IFG, insula, and premotor cortex, with the implication that a left lateralized response to speech depends neither on acoustic sensitivities nor on attentional differences.

Central to our study was the formal statistical comparison of performance in the left and right hemispheres using MVPA, which showed a significant left hemisphere advantage for the processing of intelligible speech in STG + MTG. Crucially, this affords a simpler and more convincing means of addressing the question of hemispheric asymmetries in speech processing than has been seen in other studies. For example, Okada et al. (2010) used multivariate classification data to argue for bilateral equivalence for the processing of speech intelligibility in superior temporal cortex. However, their conclusions were drawn without directly comparing the raw classification performances across hemispheres. We statistically compared classifier accuracy across left and right ROIs and compared the distribution of classifier weights within hemispheres. Both approaches gave consistent results, which we believe add considerable enhancement to the findings of the univariate analysis.

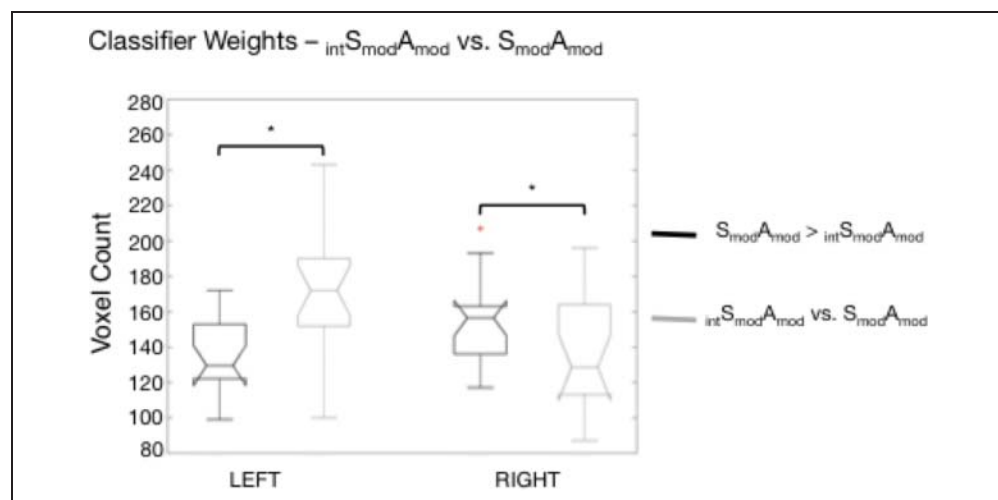
Can our results be reconciled with the studies suggesting an acoustic basis for the leftward dominance in language processing? As we describe in the Introduction, many studies have made strong claims for preferential processing of temporal features or short integration windows in the left hemisphere, but a truly *selective* response to such acoustic properties has never been clearly demonstrated. In contrast, many of these same studies have been able to demonstrate convincing right hemisphere selectivity for properties of sounds including longer durations and pitch variation. Although we have taken a

different approach by creating speech-derived stimuli with a specific interest in the modulations contributing to intelligibility, our finding of no specific leftward preference for these modulations (in the absence of intelligibility) is consistent with the previous literature.

Responses in the left premotor cortex (including portions of the left IFG) have been previously implicated in studies of degraded speech comprehension as correlates of increased comprehension or perceptual learning (Osnes, Hugdahl, & Specht, 2011; Adank & Devlin, 2010; Eisner et al., 2010; Davis & Johnsrude, 2003). A recent study by Osnes et al. (2011) showed that, in a parametric investigation of increasing intelligibility of speech (where participants heard a morphed continuum from a noise to a speech sound), premotor cortex were engaged when speech was noisy, but still intelligible. This indicates that motor representations were engaged to assist in the performance of a “do-able” speech perception task. The anterior insula has previously been associated with speech production (Wise, Greene, Buchel, & Scott, 1999; Dronkers, 1996). The activation of these sites in the current study may suggest some form of articulatory strategy was used to attempt to understand the speech.

The combinatorial coherence of amplitude and spectral modulations may have formed the acoustic “gate” for progression to further stages of processing in frontal sites. For example, vowel onsets in continuous speech are associated with the relationship between amplitude and the spectral shape of the signal (Kortekaas, Hermes, & Meyer, 1996). Although the unintelligible $S_{\text{mod}}A_{\text{mod}}$ stimuli may sound like someone talking, the formant and envelope cues to events such as vowel onset may no longer be temporally coincident. Davis, Ford, Kherif, and Johnsrude (2011) investigated the interaction of semantic and acoustic properties in the perception of spoken sentences, using time-resolved, sparse fMRI. They observed that responses in temporal cortex showing an interaction of acoustic and semantic factors preceded those in inferior frontal cortex. The authors argue

Figure 10. Box plots of group voxel counts for the top 30% of positive and negative weights in acoustic classification $intS_{\text{mod}}A_{\text{mod}}$ versus $S_{\text{mod}}A_{\text{mod}}$. Annotations indicate the result of pairwise comparisons (paired, two-sided Wilcoxon signed rank tests) of voxel counts in left and right STG + MTG ROIs. * = significant at a corrected level of $p < .025$.



that there is no available neural evidence that higher-order properties of speech affect low-level perception in a truly “top–down” manner. However, there may still be an interaction between higher-order representations and incoming sensory information within the auditory/speech perception system (Poepfel & Monahan, 2011). For example, perceptual expectancies about the properties of acoustic patterns can be rapidly and unconsciously learnt, and this can directly affect the kinds of acoustic patterns to which we are sensitive (Stilp, Rogers, & Kluender, 2010). Further analysis of the current data set using dynamic causal modeling may allow us to address these issues.

The current study provides a timely advance in our understanding of hemispheric asymmetries for speech processing. Using speech-derived stimuli, we demonstrate bilateral equivalence in superior temporal cortex for the acoustic processing of unintelligible amplitude and spectral modulations and a left dominant pattern of activation for intelligible speech. Our multivariate analyses provide direct statistical evidence for a significant left hemisphere advantage in the processing of speech intelligibility. In conclusion, our data support a model of hemispheric specialization in which the left hemisphere preferentially processes intelligible speech, but not because of an underlying acoustic selectivity (Scott & Wise, 2004).

Acknowledgments

C. M. and S. E. contributed equally to the study. C. M., Z. K. A. and S. K. S. are funded by Wellcome Trust Grant WT074414MA awarded to S. K. S. S. E. is funded by an Economic and Social Research Council studentship. The authors would like to thank staff at the Birkbeck-University College London Centre for Neuroimaging for technical advice.

Reprint requests should be sent to Dr. Carolyn McGettigan, Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, WC1N 3AR, United Kingdom, or via e-mail: c.mcgettigan@ucl.ac.uk.

Notes

1. Nonparametric statistical tests were chosen as these make fewer assumptions about underlying distributions and are less susceptible to outliers (Demsar, 2006).
2. It is conceivable that the SVM failed to demonstrate differences in accuracy for the left versus right hemisphere classifiers because of the large size of ROIs. To explore this, the $S_{\emptyset A_{\emptyset}}$ versus $S_{\emptyset A_{\text{mod}}}$ and $S_{\emptyset A_{\emptyset}}$ versus $S_{\text{mod} A_{\emptyset}}$ classifications in STG + MTG were rerun using recursive feature elimination and SVM to test performance using subsets of the voxels in the ROIs. This indicated that the SVM incurred no loss of sensitivity because of the large size of the ROIs (see Supplementary Material).

REFERENCES

Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *Neuroimage*, *49*, 1124–1132.

- Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2007). A common system for the comprehension and production of narrative speech. *Journal of Neuroscience*, *27*, 11455–11464.
- Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., & Fontaine, A. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience*, *10*, 536–540.
- Bench, J., Kowal, A., & Bamford, J. (1979). The BKB (Bamford–Kowal–Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*, 108–112.
- Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*, 389–395.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.
- Brett, M., Anton, J., Valabregue, R., & Poline, J. (2002). Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain; June 2–6, 2002; Sendai, Japan. 2002. p. 497. Available on CD-ROM in NeuroImage, Vol 16, No. 2, Abstract 497.
- Davis, M. H., Ford, M. A., Kherif, F., & Johnsrude, I. S. (2011). Does semantic context benefit speech understanding through “top–down” processes? Evidence from time-resolved sparse fMRI. *Journal of Cognitive Neuroscience*, *23*, 3914–3932.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, *23*, 3423–3431.
- Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage*, *24*, 21–33.
- Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*, 1–30.
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyru and sulci using standard anatomical nomenclature. *Neuroimage*, *53*, 1–15.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, *384*, 159–161.
- Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, *7*, 89–97.
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*, *25*, 1325–1335.
- Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, *30*, 7179–7186.
- Fellner, C., Doenitz, C., Finkenzeller, T., Jung, E. M., Rennert, J., & Schlaier, J. (2009). Improving the spatial accuracy in functional magnetic resonance imaging (fMRI) based on the blood oxygenation level dependent (BOLD) effect: Benefits from parallel imaging and a 32-channel head array coil at 1.5 Tesla. *Clinical Hemorheology and Microcirculation*, *43*, 71–82.
- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science*, *322*, 970–973.
- Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, *56*, 1127–1134.
- Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., et al. (2000). Representation of the temporal

- envelope of sounds in the human brain. *Journal of Neurophysiology*, *84*, 1588–1598.
- Greenwood, D. D. (1990). A cochlear frequency-position function for several species—29 Years later. *Journal of the Acoustical Society of America*, *87*, 2592–2605.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping*, *3*, 213–223.
- Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., & Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *12*, 140–149.
- Hart, H. C., Palmer, A. R., & Hall, D. A. (2003). Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cerebral Cortex*, *13*, 773–781.
- Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, *101*, 2725–2732.
- Hickok, G., & Poeppel, D. (2007). Opinion—The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
- Josse, G., Kherif, F., Flandin, G., Seghier, M. L., & Price, C. J. (2009). Predicting language lateralization from gray matter. *Journal of Neuroscience*, *29*, 13516–13523.
- Kortekaas, R. W., Hermes, D. J., & Meyer, G. F. (1996). Vowel-onset detection by vowel-strength measurement, cochlear-nucleus simulation, and multilayer perceptrons. *Journal of the Acoustical Society of America*, *99*, 1185–1199.
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*, 535–540.
- Langers, D. R. M., Backes, W. H., & van Dijk, P. (2003). Spectrotemporal features of the auditory cortex: The activation in response to dynamic ripples. *Neuroimage*, *20*, 265–275.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*, 1621–1631.
- Lisker, L. (1986). “Voicing” in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech*, *29*, 3–11.
- Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *Neuroimage*, *53*, 103–118.
- Morillon, B., Lehongre, K., Frackowiak, R. S. J., Ducorps, A., Kleinschmidt, A., Poeppel, D., et al. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences, U.S.A.*, *107*, 18688–18693.
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, *13*, 1362–1368.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, *28*, 8116–8123.
- Obleser, J., Wise, R. J., Alex Dresner, M., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *Journal of Neuroscience*, *27*, 2283–2289.
- Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I., Saberi, K., et al. (2010). Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cerebral Cortex*, *20*, 2486–2495.
- Osnes, B., Hugdahl, K., & Specht, K. (2011). Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. *Neuroimage*, *54*, 2437–2445.
- Parikh, P. T., Sandhu, G. S., Blackham, K. A., Coffey, M. D., Hsu, D., Liu, K., et al. (2011). Evaluation of image quality of a 32-channel versus a 12-channel head coil at 1.5T for MR imaging of the brain. *American Journal of Neuroradiology*, *32*, 365–373.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time.” *Speech Communication*, *41*, 245–255.
- Poeppel, D., & Monahan, P. J. (2011). Feedforward and feedback in speech perception: Revisiting analysis by synthesis. *Language and Cognitive Processes*, *26*, 935–951.
- Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*, 67–70.
- Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech-perception without traditional speech cues. *Science*, *212*, 947–950.
- Rosen, S., Wise, R. J. S., Chadha, S., Conway, E., & Scott, S. K. (2011). Sense, nonsense and modulations: The left hemisphere dominance for speech perception is not based on sensitivity to specific acoustic features. *PLoS ONE*, *6*, e24672.
- Schönwiesner, M., Rubsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, *22*, 1521–1528.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain*, *123*, 2400–2406.
- Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. *Journal of the Acoustical Society of America*, *120*, 1075–1083.
- Scott, S. K., & Wise, R. J. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, *92*, 13–45.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.
- Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated complex acoustic properties. *Proceedings of the National Academy of Sciences, U.S.A.*, *107*, 21914–21919.
- Thivard, L., Belin, P., Zilbovicius, M., Poline, J. B., & Samson, Y. (2000). A cortical region sensitive to auditory spectral motion. *NeuroReport*, *11*, 2969–2972.
- Wise, R. J. S., Greene, J., Buchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet*, *353*, 1057–1061.
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*, 946–953.