An application of univariate and multivariate approaches in fMRI to quantifying the hemispheric lateralization of acoustic and linguistic processes

Carolyn McGettigan,
University College London, Institute of Cognitive Neuroscience

Samuel Evans,
University College London, Institute of Cognitive Neuroscience

Stuart Rosen,
UCL

Zarinah Agnew,
University College London, Institute of Cognitive Neuroscience

Poonam Shah, and
UCL

Sophie Scott
University College London, Institute of Cognitive Neuroscience

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- 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 insula, inferior frontal cortex and superior temporal sulcus. Crucially, multivariate pattern analysis (MVPA) 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 due to low-level, speech-derived acoustic factors, and that MVPA provides a method for unbiased testing of hemispheric asymmetries in processing.

Keywords

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 (Zatorre & Belin, 2001; Boemio et al., 2005; Schönwiesner et al., 2005). Emergent theories posit, for example, differential processing of temporal vs. spectral information (e.g. Zatorre & Belin, 2001), or differences in the preference for short vs. long temporal integration windows (e.g. Poeppel, 2003) in 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 superior temporal sulcus (STS), which, when contrasted with a complex acoustic control, are typically left-lateralized, or left dominant (Scott et al., 2000; Narain et al., 2003; Eisner et al., 2010). Some studies have investigated the acoustic basis for this pattern of lateralization using modifications to intelligible speech (Obleser, Eisner & Kotz, 2008), while others suggest that the prelexical processing of speech is not actually dominated by the left temporal lobe (Hickok & Poeppel, 2007; Okada et al., 2010).

Generally, the perception of acoustic structure in sounds is associated with bilateral cortical activation. For example, the introduction of harmonic structure (Hall et al., 2002), frequency modulation (Hall et al., 2002; Hart et al., 2003), amplitude modulation (Hart et al., 2003), spectral modulations (Thivard et al., 2000) or dynamic spectral ripples (Langers et al., 2003) all show bilateral neural responses, with no evidence for asymmetry. However, these studies were not designed to necessarily test existing models of hemispheric asymmetries.

The concept that differences between the left and right temporal lobes in their response to speech might reflect differential sensitivity to acoustic factors has been tested more directly in several studies, with mixed success. An early neuroimaging study using positron emission tomography (PET; Belin et al, 1998) employed stimuli with short (40ms) and long (200ms) 'formant' transition times at the onset of sounds. Although speech-like, these stimuli did not form intelligible items. The analysis showed that while both long and short formant transitions were processed bilaterally in the superior temporal gyrus (STG), a direct comparison of the two conditions showed activation in the contrast *long > short* in the right temporal lobe, and 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, by showing a greater response to the long stimuli than to the short. 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 co-varied 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 smoothness of the amplitude envelopes), which was denoted a temporal factor. Their data did show 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, which is difficult to set within a proposed left-hemisphere preference for 'temporal' information. Likewise, their demonstration (following on from Shannon et al., 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 which would be 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 notes 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 across the dimension, and this condition may be better characterized as variation in the size and number of pitch changes). These manipulations yielded bilateral activations in the dorsolateral temporal lobes, and a direct comparison of the two conditions also showed bilateral activation, with the temporal stimuli leading to greater activation in bilateral Heschl's gyri, 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.

A more recent study (Boemio et al, 2005) varied the way information in non-speech sequences changed at different time scales, by varying both the duration of segments in the sequence and the rate and extent of pitch change across the sequence. The study found greater responses in the right STG as segment duration increased, consistent with a potential right hemisphere preference for items at longer time scales. However, there was no selective left hemisphere preference for the shorter duration items. Thus, as in Belin et al.'s study (1998), there was a right STG preference for longer sounds, but no converse sensitivity, that is, an enhanced response for shorter durations on the left.

Schönwiesner and colleagues (2005) also generated non-speech 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 was the case in the study by Zatorre and Belin (2001), they did not directly compare the activation to both temporal and spectral modulation density within either hemisphere as a test of a selective response to either kind of information. Certainly, in the case of the left anterolateral belt area, the neural responses to the two kinds of modulation were of comparable size.

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

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

The aim of the current fMRI study was to investigate the neural responses to acoustic modulations that are necessary and sufficient for intelligibility; that is, modulations of amplitude and spectrum. While no one acoustic cue determines the intelligibility of speech (Lisker, 1986), Remez et al. (1981) demonstrated that sentences comprising sine-waves tracking the main formants in speech (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, naïve subjects report hearing both of these conditions as sounding 'like someone talking' (Rosen et al, in revision), 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 revision) identified bilateral activation in left and right superior temporal cortex in response to acoustic modulations in the unintelligible conditions, with the largest peak in the right STG showing the greatest 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: Formisano et al., 2008; Hickok et al. 2009; Okada et al., 2010). In the current study, we employed both a univariate and MVPA approach, 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. In this study, we present a study of acoustic and intelligibility processing in speech where the results of a univariate analysis are both supported and enhanced by multivariate pattern analysis.

Methods

Participants

20 right-handed speakers of English (10 female, aged 18-40 years) took part in the study. All the participants had normal hearing and reported no history of neurological problems, nor any problems with speech or language. All were naïve about the aims of the experiment and unfamiliar with the experimental stimuli. The study was approved by the UCL 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-Kowel-Bench (BKB) 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, 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 revision). A semi-automatic procedure was used to track the frequencies and amplitudes of the formants every 10ms. Further signal processing was conducted offline 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). In order 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 'Ø' indicates a steady/fixed state while 'mod' indicates a dynamic/modulated state.

i. $S_{\emptyset}A_{\emptyset}$, steady state formant tracks with fixed amplitude.

ii. $S_{\emptyset}A_{mod}$, steady state formant track with dynamic amplitude variation.

iii. $S_{mod}A_{\emptyset}$, dynamic frequency variation with fixed amplitude.

iv. $S_{mod}A_{mod}$, 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. $_{int}S_{mod}A_{mod}$, 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, $S_{mod}A_{mod}$, were particularly vulnerable to small errors in formant estimation due to 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 (each a 6th-order Butterworth) with frequency responses crossing 3 dB down from the pass-band peak. Envelope extraction at the output of each analysis filter was done using full-wave rectification and 2nd-order Butterworth low-pass filtering at 30 Hz. The envelopes were then multiplied by a white noise, and each filtered by a 6th-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 crossover frequencies for both filter banks (over the frequency range 70-5000 Hz) were calculated using an equation relating position on the basilar membrane to its best frequency (Greenwood, 1990). Figure 1 shows example spectrograms from each of the five conditions.

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 revision), using 10 items from each condition. The mean intelligibility scores were 61%, 6%, 3% and 3% keywords correct for the $_{int}S_{mod}A_{mod}$, $S_{mod}A_{mod}$, $S_{mod}A_{\emptyset}$, and $S_{\emptyset}A_{mod}$ conditions, respectively.

Design and Procedure

Behavioral Pre-test—A behavioral test session was used to familiarize and train the participants with the $_{int}S_{mod}A_{mod}$ 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 sentence report accuracy using items from the $_{int}S_{mod}A_{mod}$ condition. A sentence was played for the participant over Sennheiser HD201 headphones (Sennheiser U.K., High Wycombe, Buckinghamshire, UK) and s/he was asked to repeat whatever s/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 complete sentence to the participant and played it back to him/her once. 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 Tesla 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 Tesla field: Fellner et al., 2009; Parikh et al., 2011). There were two runs of 150 echo-planar whole-brain volumes (TR = 9 seconds, TA = 3 seconds, TE = 50 ms, flip angle 90 degrees, 35 axial slices, 3mm × 3mm × 3mm in-plane resolution. A sparse-sampling routine was employed (Edmister et al., 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 seconds (with jittering of +/− 500 ms) 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 worn by the participant (Etymotic Inc., Elk Grove Village, IL, USA). In each functional run, the participant heard 50 stimuli from each of the five auditory conditions (2 stimuli per trial). For the four unintelligible conditions, these 50 items were repeated in the second functional

run, while participants heard 100 distinct sentences in the $_{int}S_{mod}A_{mod}$ 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 order of presentation of stimuli from the different conditions was pseudorandomized to allow a relatively even distribution of the conditions across the run without any predictable ordering effects. A silent baseline was included in the form of four silent miniblocks in each functional run, each comprising five silent trials. After the functional run was complete, a high-resolution T1-weighted anatomical image was acquired (HIRes MP-RAGE, 160 sagittal slices, voxel size = 1 mm^3).

Behavioral Post-test—The post-test consisted of 80 sentences from the $_{int}S_{mod}A_{mod}$ condition presented over Sennheiser HD201 headphones, half of which had been presented in the scanner and half of which were novel exemplars from the same condition. After each sentence, participants were asked to repeat what they heard. Speech perception accuracy was scored online, 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 using parameters obtained from segmentation of the anatomical image, and smoothed using a Gaussian kernel of 8mm FWHM. Event-related responses for each event type were modeled as a canonical haemodynamic response function, with event onsets modeled from the acoustic onset of the first auditory stimulus in each trial and with durations of 4 seconds (the approximate duration of 2 sequential stimuli). Each condition was modeled as a separate regressor in a generalized linear model (GLM). Six movement parameters (3 translations, 3 rotations) were included as regressors of no interest. At the first level (single-subject), T-contrast images were created for the comparison of each of the five auditory conditions with the silent baseline. The T images for the four unintelligible conditions from each participant were entered into a random-effects, 2×2 repeated-measures ANOVA group model with factors Spectral Modulation (present or absent) and Amplitude Modulation (present or absent). A second random effects model formed an ANOVA with one factor, Condition, which included single-subject T-images from all five auditory conditions. All second-level contrasts were thresholded at $p < .05$ (family-wise-error corrected), with a cluster extent threshold of 40 voxels. Coordinates of peak activations were labeled using the SPM5 anatomy toolbox (Eickhoff et al., 2005).

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 et al., 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 (http://www.kyb.tuebingen.mpg.de/bs/people/spider/) was used to train and validate models. The SVM used a hard margin and the Andre optimization. A linear SVM is a discriminant function that attempts to fit a hyperplane separating data observed in different experimental conditions (classes). When applied to fMRI data, a SVM attempts to fit a linear boundary that maximizes the distance between the most similar training examples from each class - the support vectors - within a multidimensional space with as many dimensions as voxels. The performance of the classifier is validated by evaluating the success of this classification boundary (defined by the training data) in predicting the class of previously unseen data examples. In the current study, for each

participant, the first classifier was trained on the first run and tested on the second, and vice versa for the second classifier. The overall performance for each classification was calculated by averaging the performance across the two classifiers for each participant. Three acoustic classifications were performed for each participant: $S_\emptyset A_\emptyset$ vs $S_\emptyset A_{mod}$, $S_\emptyset A_\emptyset$ vs $S_{mod}A_\emptyset$ and $S_\emptyset A_\emptyset$ vs $S_{mod}A_{mod}$. An intelligibility classification was also run, for $_{int}S_{mod}A_{mod}$ vs $S_{mod}A_{mod}$. The classifications were performed separately for a number of subject-specific, anatomically-defined regions-of-interest (ROIs). The Freesurfer image analysis suite (http://surfer.nmr.mgh.harvard.edu/) was used to perform cortical reconstruction and volumetric segmentation via an automated cortical parcellation of individual participants' T1 images (Destrieux et al., 2010). This generated subject-specific, left- and right-hemisphere anatomical ROIs for Heschl's Gyrus (HG), Middle+Superior Temporal Gyrus (MTG+STG; generated from parcellation of MTG, STG and STS) and Inferior Occipital Gyrus (IOG; included as a control ROI). These anatomically-defined regions were included based on *a priori* hypotheses about the key sites for intelligibility and acoustic processing of speech (Davis & Johnsrude 2003; Eisner et al., 2010; Obleser et al., 2007, 2008; Scott et al., 2000, 2006), and hence not contingent on the results of the univariate analyses.

Results

Behavioral Tests

Pre-test—A stringent criterion performance of 5 consecutive correct responses (with 100% accuracy on keyword report) was used to ensure a thorough training on the $_{int}S_{mod}A_{mod}$ condition before the scan. For those participants who reached this criterion in the pre-test, 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 pre-test items. However, as all participants achieved 3 consecutive correct responses within an average of only 23.5 trials (SD 16.4), with six participants reaching this threshold within the first 6 items, we were satisfied that all participants would be able to understand a sufficient proportion of the $_{int}S_{mod}A_{mod}$ items in the scanner to support our planned intelligibility contrasts.

Post-test—The average accuracy across the whole post-test 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 pre-test 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 (Fully Factorial ANOVA)—Figure 2a shows the activation extents from a group F-contrast exploring the main effects of spectral modulation, and amplitude modulation, as well as plots of effect size across the four unintelligible conditions taken from the peak activations in each contrast. For both contrasts, there was greater signal in bilateral STG when the modulations were present than when they were absent. For the main effect of spectral modulation, the peak activation was located on STG, anterior and lateral to Heschl's gyrus in both hemispheres, although both clusters encompassed portions of Heschl's and posterior areas of STG. For modulations of the amplitude envelope, bilateral clusters on STG were more focal, and located lateral to Heschl's gyrus.

Significant peak and sub-peak voxels (more than 8mm apart) for the Main Effects are listed in Table 1.

Interaction: Spectral x Amplitude Modulation—An F-contrast was run exploring the interaction of Spectral and Amplitude modulations. This was run to test for additive or super-additive responses to the intelligibility-relevant acoustic modulations (i.e. activation for the dually modulated $S_{mod}A_{mod}$ condition equivalent to, or exceeding, the the sum of the responses to the singly modulated $S_{\emptyset}A_{mod}$ and $S_{mod}A_{\emptyset}$ conditions). This contrast gave bilateral STG activation, with the overall peak voxel in the right hemisphere [66 −18 3]. However, the plots of contrast estimates from the main peaks indicate a sub-additivity of the two factors i.e. that 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 listed in Table 1.

Effect of Intelligibility (One-way ANOVA)—A T-contrast of $_{int}S_{mod}A_{mod} > S_{mod}A_{mod}$ gave significant activation in bilateral STS and STG, with the peak voxel in left STS and a larger cluster extent in the left hemisphere. A single cluster of activation extended from the left anterior insula to the pars triangularis and pars opercularis of left inferior frontal gyrus (IFG). Figure 3 shows the results of this contrast, with plots of contrast estimates for all five conditions compared with rest – significant peaks and sub-peaks are listed in Table 2.

Multivariate Pattern Analysis

Two participants were excluded from the multivariate analyses due to unsuccessful cortical parcellation.

Acoustic Classifications—The support vector machine was trained and tested on three acoustic classifications: $S_{\emptyset}A_{\emptyset}$ vs. $S_{\emptyset}A_{mod}$, $S_{\emptyset}A_{\emptyset}$ vs. $S_{mod}A_{\emptyset}$, $S_{\emptyset}A_{\emptyset}$ vs. $S_{mod}A_{mod}$. Figure 4a-c shows boxplots of group classification accuracy by ROI for each of the three classifications. Performance in each classification was tested against a chance performance of 0.5, using the one-sided Wilcoxon signed rank test, with a corrected significance level of $p < .008$ (correcting for six ROIs tested in each classification). All temporal ROIs performed significantly better than chance on the acoustic classifications ($p < .001$). The left IOG, which was included as a control site, performed significantly better than chance in the $S_{\emptyset}A_{\emptyset}$ vs. $S_{mod}A_{mod}$ 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 homologues 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 for 3 left-right comparisons in each classification); paired, two-sided Wilcoxon signed rank tests). Finally, in order to compare spectral and temporal processing within each hemisphere, performance on the $S_{\emptyset}A_{\emptyset}$ vs. $S_{\emptyset}A_{mod}$ and $S_{\emptyset}A_{\emptyset}$ vs. $S_{mod}A_{\emptyset}$ classifications was compared within left HG, right HG, left STG+MTG and right STG+MTG (paired, two-sided Wilcoxon signed rank tests with corrected significance level of $p < 0.013$). 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 ($w = 30.5$, $p = .017$).

Intelligible vs. Unintelligible stimuli—Figure 4d shows boxplots of group classification accuracy by ROI for the classification of $_{int}S_{mod}A_{mod}$ vs. $S_{mod}A_{mod}$ stimuli. Performance was significantly better than chance for the left and right STG+MTG ROIs (both: $w = 0$, p

< .0001; in one-tailed Wilcoxon signed rank tests with corrected significance level of p < .008). The left IOG, which was included as a control site, performed poorly at 54% but this was significantly better than chance (w = 29.5, p = .0073). Performance in left and right Heschl's gyrus 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 no different from chance in this classification. The comparison of left and right homologues of each ROI showed that performance was equivalent between the left and right hemispheres for Heschl's gyrus 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 contrasts $S_{mod}A_{mod} >_{int} S_{mod}A_{mod}$ and $S\emptyset A_\emptyset > S_{mod}A_{mod}$. It is possible that this region is involved in default-network processes and the observed activation profiles reflect task-related deactivation of this region.

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, insula and inferior frontal cortex 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 (Hickok and Poeppel, 2007; Okada et al., 2010).

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. Previous studies exploring the neural correlates of amplitude envelope and spectral modulation have observed similar bilateral patterns of activation in the dorsolateral temporal lobes (Hart et al., 2003; Langers et al., 2003; Boemio et al., 2005; Obleser et al., 2008; Thivard et al., 2001). Inspection of the peak activations in Table 1 shows that, even though the peak voxels occurred in left STG for both main effects, the cluster extents and statistical heights of the local peaks are largely similar across the hemispheres. Rosen et al. (in revision) found similar equivalence in the extent of activation between left and right STG for the processing of the unintelligible conditions. However, they modeled an additive interaction of the two factors, and found the peak activation in right STG showing a much greater response to the $S_{mod}A_{mod}$ condition than to the conditions with only one type of modulation ($S\emptyset A_{mod}$, and $S_{mod}A\emptyset$). An F-contrast examining the interaction of the two modulation types in the current experiment also yielded a right-hemisphere peak, but visual inspection of the plots of the contrast estimates indicated no evidence for a truly additive response anywhere in the activated regions. This may reflect differences in the design of the two experiments. Rosen et al. employed a blocked design in PET, in which listeners were exposed to around 1 minute of stimulation from a single condition during each scan. This may have allowed for a slower emergence of a greater response to the $S_{mod}A_{mod}$ condition than the immediate responses measured in the current, event-related, design. Nonetheless, the univariate analyses in both studies offer no support

for a left-hemispheric 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. Previous approaches to calculating laterality effects in functional neuroimaging data (Boemio et al., 2005; Josse et al., 2009; Obleser et al., 2008; 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 is compromised in large ROIs). The multivariate approach using linear support vector machines in the current study affords great sensitivity within large ROIs by providing sparse solutions. Moreover, 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 asymmetries, our findings stand in contrast to previous claims for subtle left-right differences in preference for temporal and spectral information (Zatorre & Belin, 2001; Boemio et al., 2005). Instead, our results suggest that when speech-derived, intelligibility-relevant modulations are considered, there is no difference in the acoustic processing of speech between the left and right hemispheres. Importantly, we were able to investigate these acoustic effects in the absence of intelligibility, so that this factor was not conflated with the acoustic manipulations (cf Obleser, Eisner & Kotz (2008), in which the parametric manipulation of spectral and temporal resolution in real sentences had concomitant effects on intelligibility). Furthermore, we observed a significant within-ROI advantage for the classification of spectral modulations compared with amplitude modulations in bilateral Heschl's gyrus and STG+MTG. This indicates that the lack of evidence for hemispheric asymmetries in acoustic processing was not due to insufficient sensitivity in our MVPA.

Several neuroimaging studies have identified peak responses to intelligible speech in left STS, either in the context of strongly lateralized activation (Scott et al., 2000; Narain et al., 2003; Eisner et al., 2010) or rather more bilaterally-distributed responses along both left and right STS (Awad et al., 2007; Davis & Johnsrude, 2003; Scott et al., 2006). 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 of these cases, there were clear perceptual differences between the intelligible and unintelligible conditions, which 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 $S_{mod}A_{mod}$ stimuli as sounding like someone speaking, but with no sense of intelligibility (Rosen et al, in revision). In the absence of clear perceptual differences between these two modulated conditions, listeners should have attended equivalently to stimuli from these two categories, in the expectation that they could be intelligible. Our univariate analysis revealed a left-dominant response to intelligible speech, in STG/STS, insula and IFG, with the implication that a left lateralized response to speech depends neither on acoustic sensitivities, nor on attentional differences. However, any left dominant effect would benefit from some elaboration by a statistical comparison across the hemispheres. 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. In contrast, we have focused on direct statistical comparisons of classification accuracy in the left and right hemispheres. We believe that the employment of multivariate approaches in our study not only supports but supplements 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 (or combinations of modulations) in the absence of intelligibility, is consistent with the previous literature.

A key difference between the acoustic and intelligibility contrasts reported in this study is the enhancement of responses in insula and inferior frontal cortex to intelligible speech. Responses in the left premotor cortex (including portions of the left inferior frontal gyrus) have been previously implicated in studies of degraded speech comprehension as correlates of increased comprehension or perceptual learning (Davis & Johnsrude, 2003; Adank & Devlin, 2010, Eisner et al., 2010; Osnes et al., 2010). Eisner et al. (2010) related activation in posterior parts of the left inferior frontal gyrus to variability in working memory capacity, and suggested a working memory mediated strategy as a potential mechanism for perceptual learning of noise-vocoded speech. A recent study by Osnes et al. (2010) 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. In the current study, the peak activation to intelligible speech (when contrasted with the unintelligible $S_{mod}A_{mod}$ condition) was located in the left anterior insula. The anterior insula has previously been associated with normal speech production (Wise et al., 1999), and Dronkers (1996) reported a lesion mapping study in which she found the anterior insula to be damaged in 100% of stroke patients with a speech apraxia i.e. a deficit in the ability to coordinate movements for speech output. The activation of this site 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 inferior frontal and insular cortex. For example, vowel onsets in continuous speech are associated with the relationship between amplitude and the spectral shape of the signal (Kortekaas et al., 1996). The way in which the stimuli have been constructed for the current experiment means that, while the unintelligible $S_{mod}A_{mod}$ stimuli may sound like someone talking, the formant and envelope cues to events such as vowel onset may no longer be temporally coincident.

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 amplitude and spectral modulations in the absence of intelligibility. We show a left dominant pattern of activation for intelligible speech, in which we identify the insula as a site for strategic

articulatory processing of coherent speech stimuli for comprehension. Finally, 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 and Wise, 2004).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Adank P, Devlin JT. On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. NeuroImage. 2010; 49:1124–1132. [PubMed: 19632341]

Altmann CF, de Oliveira CG, Heinemann L, Kaiser J. Processing of spectral and amplitude envelope of animal vocalizations in the human auditory cortex. Neuropsychologia. 2010; 48:2824–2832. [PubMed: 20493891]

Awad M, Warren JE, Scott SK, Turkheimer FE, Wise RJS. A common system for the comprehension and production of narrative speech. Journal of Neuroscience. 2007; 27:11455–11464. [PubMed: 17959788]

Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A. Lateralization of speech and auditory temporal processing. Journal of Cognitive Neuroscience. 1998; 10:536–540. [PubMed: 9712682]

Bench J, Kowal A, Bamford J. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. British Journal of Audiology. 1979; 13:108–112. [PubMed: 486816]

Boemio A, Fromm S, Braun A, Poeppel D. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nature Neuroscience. 2005; 8:389–395.

Boersma, P.; Weenink, D. Praat: Doing phonetics by computer. 2008. Downloaded from: http://www.fon.hum.uva.nl/praat

Brainard DH. The psychophysics toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]

Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. Journal of Neuroscience. 2003; 23:3423–3431. [PubMed: 12716950]

Dehaene-Lambertz G, Pallier C, Serniclaes W, Sprenger-Charolles L, Jobert A, Dehaene S. Neural correlates of switching from auditory to speech perception. NeuroImage. 2005; 24:21–33. [PubMed: 15588593]

Demsar J. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research. 2006; 7:1–30.

Destrieux C, Fischl B, Dale A, Halgren E. Automatic parcellation of human cortical gyru and sulci using standard anatomical nomenclature. NeuroImage. 2010; 53:1–15. [PubMed: 20547229]

Dronkers NF. A new brain region for coordinating speech articulation. Nature. 1996; 384:159–161. [PubMed: 8906789]

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM. Improved auditory cortex imaging using clustered volume acquisitions. Human Brain Mapping. 1999; 7:89–97. [PubMed: 9950066]

Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage. 2005; 25:1325–1335. [PubMed: 15850749]

Eisner F, McGettigan C, Faulkner A, Rosen S, Scott SK. Inferior Frontal Gyrus Activation Predicts Individual Differences in Perceptual Learning of Cochlear-Implant Simulations. Journal of Neuroscience. 2010; 30:7179–7186. [PubMed: 20505085]

Fellner C, Doenitz C, Finkenzeller T, Jung EM, Rennert J, Schlaier J. 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. 2009; 43:71–82. [PubMed: 19713602]

Formisano E, De Martino F, Bonte M, Goebel R. "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. Science. 2008; 322:970–973. [PubMed: 18988858]

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A. Representation of the temporal envelope of sounds in the human brain. Journal of Neurophysiology. 2000; 84:1588–1598. [PubMed: 10980029]

Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ, Laufs H. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron. 2007; 56:1127–1134. [PubMed: 18093532]

Greenwood DD. A cochlear frequency-position function for several species – 29 years later. Journal of the Acoustical Society of America. 1990; 87:2592–2605. [PubMed: 2373794]

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. "Sparse" temporal sampling in auditory fMRI. Human Brain Mapping. 1999; 3:213–223. [PubMed: 10194620]

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ. Spectral and temporal processing in human auditory cortex. Cerebral Cortex. 2002; 12:140–149. [PubMed: 11739262]

Hart HC, Palmer AR, Hall DA. Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. Cerebral Cortex. 2003; 13:773–781. [PubMed: 12816893]

Hickok G, Poeppel D. Opinion - The cortical organization of speech processing. Nature Reviews Neuroscience. 2007; 8:393–402.

Hickok G, Okada K, Serences JT. Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing. Journal of Neurophysiology. 2009; 101:2725–2732. [PubMed: 19225172]

Josse G, Kherif F, Flandin G, Seghier ML, Price CJ. Predicting Language Lateralization from Gray Matter. Journal of Neuroscience. 2009; 29:13516–13523. [PubMed: 19864564]

Kortekaas RW, Hermes DJ, Meyer GF. Vowel-onset detection by vowel-strength measurement, cochlear-nucleus simulation, and multilayer perceptrons. Journal of the Acoustical Society of America. 1996; 99:1185–1199. [PubMed: 8609302]

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. Nature Neuroscience. 2009; 12:535–540.

Langers DRM, Backes WH, van Dijk P. Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. NeuroImage. 2003; 20:265–275. [PubMed: 14527587]

Lisker L. "Voicing" in English: A catalogue of acoustic features signaling /b/versus /p/ in trochees. Language and Speech. 1986; 29:3–11. [PubMed: 3657346]

Morillon B, Lehongre K, Frackowiak RSJ, Ducorps A, Kleinschmidt A, Poeppel D, Giraud AL. Neurophysiological origin of human brain asymmetry for speech and language. Proceedings of the National Academy of Sciences USA. 2010; 107:18688–18693.

Narain C, Scott SK, Wise RJS, Rosen S, Leff A, Iversen SD, Matthews PM. Defining a left-lateralized response specific to intelligible speech using fMRI. Cerebral Cortex. 2003; 13:1362–1368. [PubMed: 14615301]

Obleser J, Eisner F, Kotz SA. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. Journal of Neuroscience. 2008; 28:8116–8123. [PubMed: 18685036]

Okada K, Rong F, Venezia J, Matchin W, Hsieh I, Saberi K, Serences JT, Hickok G. Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. Cerebral Cortex. 2010; 20:2486–2495. [PubMed: 20100898]

Osnes B, Hugdahl K, Specht K. Effective connectivity analysis demonstrates involvement of premotor cortex during speech perception. NeuroImage. 2011; 54:2437–2445. [PubMed: 20932914]

Parikh PT, Sandhu GS, Blackham KA, Coffey MD, Hsu D, Liu K, Jesberger J, Griswold M, Sunshine JL. 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. 2011; 32:365–373. [PubMed: 21163877]

Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Communication. 2003; 41:245–255.

Remez RE, Rubin PE, Pisoni DB, Carrell TD. Speech-perception without traditional speech cues. Science. 1981; 212:947–950. [PubMed: 7233191]

Rosen S, Wise RJS, Chadha S, Scott SK. Sense, nonsense and modulations: The left hemisphere dominance for speech perception is not based on sensitivity to specific acoustic features. PloS ONE. (under review).

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. [PubMed: 11099443]

Scott SK, Wise RJ. The functional neuroanatomy of prelexical processing in speech perception. Cognition. 2004; 92:13–45. [PubMed: 15037125]

Scott SK, Rosen S, Lang H, Wise RJS. Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. Journal of the Acoustical Society of America. 2006; 120:1075–1083. [PubMed: 16938993]

Schönwiesner M, Rubsamen R, von Cramon DY. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. European Journal of Neuroscience. 2005; 22:1521–1528. [PubMed: 16190905]

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270:303–304. [PubMed: 7569981]

Thivard L, Belin P, Zilbovicius M, Poline JB, Samson Y. A cortical region sensitive to auditory spectral motion. NeuroReport. 2000; 11:2969–2972. [PubMed: 11006976]

Wise RJS, Greene J, Buchel C, Scott SK. Brain regions involved in articulation. Lancet. 1999; 353:1057–1061. [PubMed: 10199354]

Zatorre RJ, Belin P. Spectral and temporal processing in human auditory cortex. Cerebral Cortex. 2001; 11:946–953. [PubMed: 11549617]
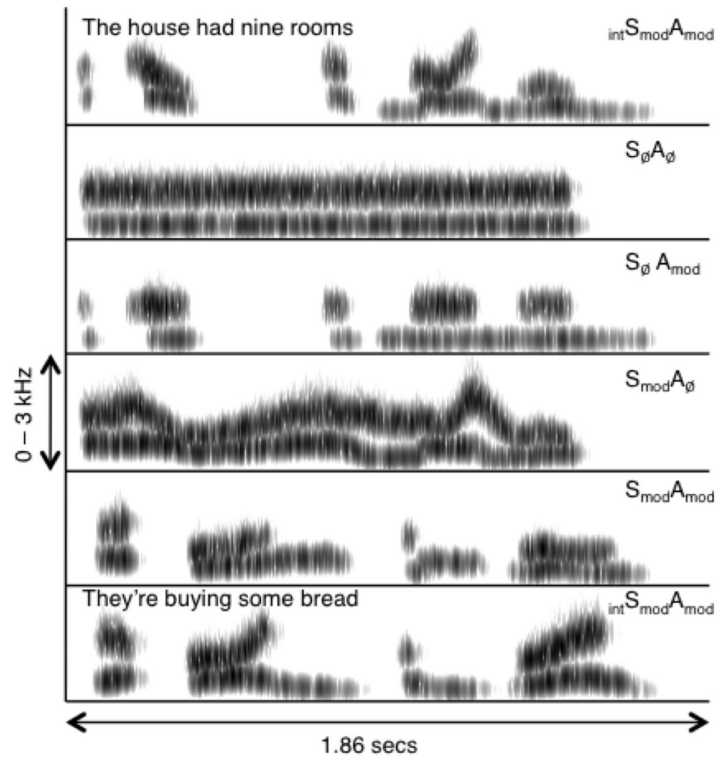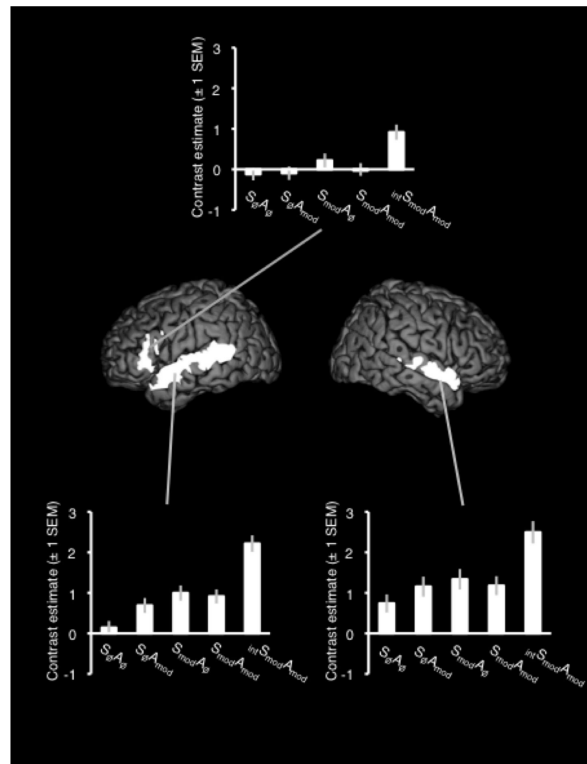
Figure 1.
Example spectrograms from the 5 auditory conditions used in the experiment. The unintelligible condition SmodAmod in the example was constructed using spectral modulations from 'The house had nine rooms' and the amplitude envelope from 'They're buying some bread'. Darker shading indicates portions of greater intensity in the signal.
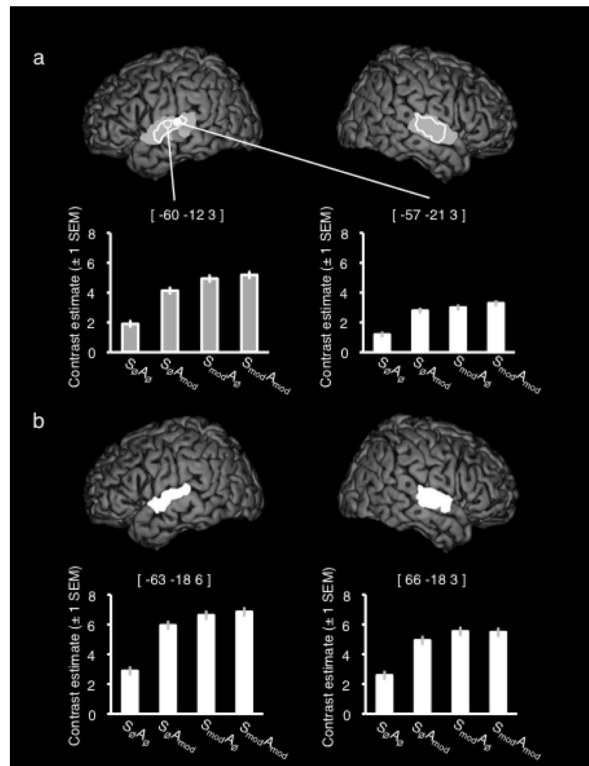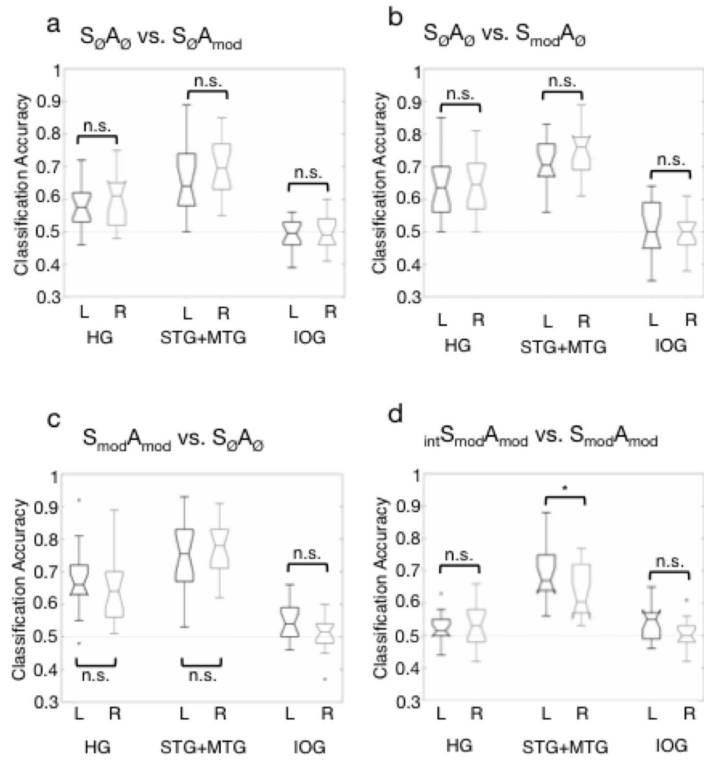
Figure 2.
(a) Activation peaks and extents for the main effect of spectral modulation (Grey) and amplitude modulation (White). Plots show the contrast estimates (± 1 SEM) for each condition taken from the peak voxel in each contrast. (b) Activation extent for the interaction of spectral and amplitude modulation. Plots show the contrast estimates (± 1 SEM) for each condition taken from the peak voxel in each hemisphere. 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 Montreal Neurological Institute (MNI) stereotactic space.
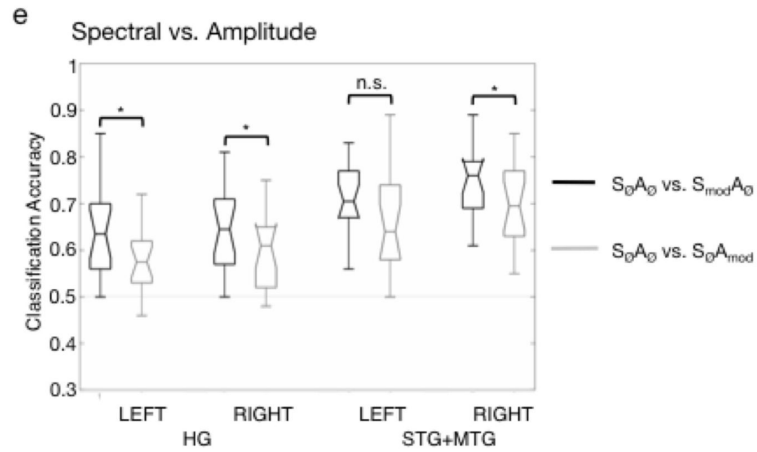
Figure 3.
Activation in the contrast of intSmodAmod > SmodAmod. Plots show the contrast estimates (± 1 SEM) for each condition taken from the peak voxel in each cluster. The image is shown at a corrected (family-wise error) height threshold of p < .05 and a cluster extent threshold of 40 voxels. Coordinates are given in Montreal Neurological Institute (MNI) stereotactic space.

Figure 4.
Box plots of group classification performance on (a) SØAØ vs. SØAmod, (b) SØAØ vs. SmodA0, (c) SØAØ vs. SmodAmod (d) Spectral vs Amplitude classification (i.e. (SØAØ vs. SmodAØ) versus (SØAØ vs. SØAmod) in each of the anatomically-defined regions-of-interest (ROIs). Annotations indicate the result of pairwise comparisons (paired, two-sided Wilcoxon signed rank tests) of performance in left and right ROIs (a-c, e) and within-ROI (d). * = significant at a corrected level of $p < .017$ (a-c, e) or $p < .013$ (d), HG = Heschl's gyrus, STG+MTG = combined superior temporal gyrus and middle temporal gyrus, IOG = inferior occipital gyrus, L = left hemisphere, R = right hemisphere.

Figure 5.
Box plots of group classification performance on (a) SØAØ vs. SØAmod, (b) SØAØ vs. SmodAØ, (c) SØAØ vs. SmodAmod (d) Spectral vs Amplitude classification (i.e. (SØAØ vs. SmodAØ) versus (SØAØ vs. SØAmod) in each of the anatomically-defined regions-of-interest (ROIs). Annotations indicate the result of pairwise comparisons (paired, two-sided Wilcoxon signed rank tests) of performance in left and right ROIs (a-c, e) and within-ROI (d). * = significant at a corrected level of $p < .017$ (a-c, e) or $p < .013$ (d), HG = Heschl's gyrus, STG+MTG = combined superior temporal gyrus and middle temporal gyrus, IOG = inferior occipital gyrus, L = left hemisphere, R = right hemisphere.

Table 1

Peak and sub-peak (if more than 8mm apart) activations from the contrasts of acoustic effects in the univariate analysis.

Contrast	No of voxels	Region	Coordinates			F	z
			x	y	z		
Main Effect of Spectral Modulation	369	Left superior temporal gyrus	−60	−12	3	240.02	>8
	290	Right superior temporal gyrus	−48	−36	15	33.87	5.15
			63	−12	3	142.98	Inf
Main Effect of Amplitude Modulation	121	Left superior temporal gyrus	−57	−21	3	105.37	>8
	80	Right superior temporal gyrus	−63	−15	3	101.73	>8
			66	−15	3	68.40	6.86
			63	−24	9	40.11	5.54
Interaction of Spectral and Amplitude Modulation	135	Right superior temporal gyrus	66	−18	3	76.38	7.15
	169	Left superior temporal gyrus	−63	−18	6	71.72	6.99
			−54	−6	−3	52.38	6.18

Table 2

Peak and sub-peak (if more than 8mm apart) activations from the contrast of intelligibility in the univariate analysis.

Contrast	No of voxels	Region	Coordinates			T	z
			x	y	z		
Intelligible > Unintelligible	614	Left superior temporal sulcus	−57	−6	−9	10.99	>8
			−57	−24	0	8.98	7.62
			−51	−48	12	8.15	7.08
			−54	−33	3	8.08	7.03
	273	Left inferior frontal gyrus	−30	27	0	8.48	7.30
			−42	27	−3	7.90	6.91
			−48	15	21	7.47	6.61
	272	Right superior temporal sulcus	57	3	−12	8.26	7.15
			51	−24	0	7.12	6.36
			60	−15	0	6.78	6.11