The Pathways for Intelligible Speech: Multivariate and Univariate Perspectives

S. Evans[1,8], J.S. Kyong[2], S. Rosen[2], N. Golestani[1,4], J.E. Warren[5], C. McGettigan[1,7], J. Mourão-Miranda[3,6], R.J.S. Wise[5] and S.K. Scott[1]

[1]Institute of Cognitive Neuroscience, [2]Department of Speech, Hearing and Phonetic Sciences, [3]Department of Computer Science, Centre for Computational Statistics and Machine Learning, University College London, London WC1E 6BT, UK [4]Department of Clinical Neuroscience, University Medical School, Geneva CH-1211, Switzerland [5]Computational, Cognitive and Clinical Neuroimaging Laboratory, Imperial College London, London W12 0NN, UK [6]Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College, London SE5 8AF, UK [7]Department of Psychology, Royal Holloway University, University of London, Egham TW20 0EX, UK and [8]MRC Cognition and Brain Sciences Unit, Cambridge CB2 7EF, UK

Address correspondence to S. Evans, Institute of Cognitive Neuroscience, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, UK. Email: samuel.evans@ucl.ac.uk

S. Evans and J.S. Kyong are joint first authors.

An anterior pathway, concerned with extracting meaning from sound, has been identified in nonhuman primates. An analogous pathway has been suggested in humans, but controversy exists concerning the degree of lateralization and the precise location where responses to intelligible speech emerge. We have demonstrated that the left anterior superior temporal sulcus (STS) responds preferentially to intelligible speech (Scott SK, Blank CC, Rosen S, Wise RJS. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 123:2400–2406.). A functional magnetic resonance imaging study in Cerebral Cortex used equivalent stimuli and univariate and multivariate analyses to argue for the greater importance of bilateral posterior when compared with the left anterior STS in responding to intelligible speech (Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G. 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. 20: 2486–2495.). Here, we also replicate our original study, demonstrating that the left anterior STS exhibits the strongest univariate response and, in decoding using the bilateral temporal cortex, contains the most informative voxels showing an increased response to intelligible speech. In contrast, in classifications using local "searchlights" and a whole brain analysis, we find greater classification accuracy in posterior rather than anterior temporal regions. Thus, we show that the precise nature of the multivariate analysis used will emphasize different response profiles associated with complex sound to speech processing.

Keywords: fMRI, intelligibility, multivariate pattern analysis, speech perception, superior temporal sulcus

Introduction

Studies in humans and nonhuman primates suggest that auditory information is processed hierarchically, with primary auditory cortex (PAC) responding in a relatively nonselective fashion, and lateral regions responding selectively to stimuli of greater complexity (Rauschecker 1998; Kaas and Hackett 2000; Wessinger et al. 2001; Davis and Johnsrude 2003). In the monkey, anatomical and functional evidence supports the notion of an anterior "what" pathway that is sensitive to different conspecific communication calls (Rauschecker 1998; Tian et al. 2001). A similar pathway has been suggested in humans, but there is disagreement concerning the extent to which it is left lateralized, and as to whether intelligible speech is processed in predominantly anterior, posterior, or both temporal fields (Scott et al. 2000; Davis and Johnsrude 2007; Hickok and Poeppel 2007; Rauschecker and Scott 2009; Peelle et al. 2010).

Many functional imaging studies have attempted to isolate neural regions that are sensitive to intelligible speech compared with those regions that respond to acoustic complexity. The selection of a suitable baseline comparison condition has proved difficult due to the inherent acoustic complexity of the speech signal; low level auditory baselines such as tones and noise bursts make it difficult to distinguish between neural responses that are specific to speech, and those that are a consequence of the perception of a complex sound. Using rotated speech (Blesser 1972), which is well matched to speech in both spectral and amplitude variations, we have shown that the left anterior superior temporal sulcus (STS) responds preferentially to intelligible speech (Scott et al. 2000; Narain et al. 2003). These previous studies employed univariate statistical analyses, which identified the regions in which there was a greater mean response to 2 different kinds of intelligible speech [clear and noise-vocoded speech (Shannon et al. 1995)] when compared with 2 kinds of unintelligible sounds (rotated speech and rotated-noise-vocoded speech).

Other researchers have suggested that the recognition of intelligible speech arises in bilateral posterior STS (Okada and Hickok 2006; Hickok and Poeppel 2007; Hickok et al. 2009; Vaden et al. 2010). A recent study in *Cerebral Cortex* (Okada et al. 2010) replicated the Scott et al. (2000) methodology with functional magnetic resonance imaging (fMRI). The univariate analysis in the study showed widespread bilateral activation to the summation of clear and noise-vocoded speech relative to their unintelligible rotated equivalents. The authors then conducted a multivariate pattern analysis (Pereira et al. 2009) within small cube-shaped regions of interest (ROIs) at specific sites in the temporal cortex. This showed that the bilateral anterior and posterior STS (in addition to the right mid-STS) contained sufficient information to separate intelligible from unintelligible sounds. Two sets of classifications were performed, classifications in which the conditions differed in intelligibility (e.g. clear vs. rotated speech and noise-vocoded speech vs. rotated-noise-vocoded speech) and those in which the conditions differed predominantly in spectral detail (clear

vs. noise-vocoded speech and rotated speech vs. rotated-noise-vocoded speech). The left posterior and right mid-STS showed the greatest classification accuracy in discriminations of intelligibility when they were expressed relative to the accuracy in discriminations of spectral detail. The left anterior STS successfully classified the contrasts of intelligibility, as well as one of the contrasts that differed in spectral detail (clear vs. noise-vocoded speech). This was interpreted as showing that the left anterior STS was unlikely to be a key region involved in resolving intelligible speech owing to its additional sensitivity in discriminating "spectral detail".

Here, we also replicate the Scott et al. (2000) study in fMRI using univariate general linear modeling and multivariate pattern analysis. We conduct additional univariate and multivariate analyses that allow a more complete description of the role of the bilateral anterior and posterior temporal cortices, and regions beyond the temporal lobe, in responding to intelligible speech.

Materials and Methods

Participants

Twelve right-handed native English speakers with no known hearing or language impairments participated in the experiment (aged 18–38, mean age 25, 3 males). All participants gave informed consent. The experiment was performed with the approval of the local ethics committee of the Hammersmith Hospital.

Stimuli

Stimuli were as described in Scott et al. (2000) and Narain et al. (2003). In brief, all stimuli were drawn from low-pass filtered (3.8 kHz) digital representations of the Bamford-Kowal-Bench sentence corpus (Bench et al. 1979). There were 4 stimuli conditions: Natural speech (clear), noise-vocoded (NV), spectrally rotated (rot), and spectrally rotated-noise-vocoded speech (rotNV).

The rotation of speech is achieved by inverting the frequency spectrum around 2 kHz using a simple modulation technique; this retains spectral and temporal complexity, but makes the speech unintelligible (Blesser 1972). It has been described previously as sounding like an alien speaking your language but with different articulators (Blesser 1972). It contains some phonetic features, for example, the presence of voicing, but these features do not generally give rise to intelligible sounds without significant training. A preprocessing filter was used to give the rotated speech approximately the same long-term average spectrum as the original, unrotated speech.

Noise-vocoding involves passing the speech signal through a filter bank (in this case 6 filters) to extract the time-varying envelopes associated with the energy in each spectral channel. Envelope detection occurred at the output of each analysis filter by half-wave rectification and low-pass filtering at 320 Hz. The extracted envelopes were then multiplied by white noise and combined after refiltering (Shannon et al. 1995). This retains the amplitude envelope cues within specified spectral bands, but removes spectral detail. With 6 bands, the speech can be understood with a small amount of training. It sounds like a harsh whisper with only a weak sense of pitch. Subjects underwent a short training, as described in Scott et al. (2000), to ensure that they understood the noise-vocoded speech. The combination of vocoding and rotation sounds like intermittent noise with weak pitch changes. It does not contain phonetic content and is not intelligible or recognizable even as "alien" speech.

The clear and noise-vocoded speech are both intelligible, but the rotated and rotated-noise-vocoded speech are not, while clear and rotated speech contain more detailed spectral information than noise-vocoded and rotated-noise-vocoded speech.

Functional Neuroimaging

Subjects were scanned on a Philips (Philips Medical Systems, Best, The Netherlands) Intera 3.0-T MRI scanner using Nova Dual gradients, a phased-array head coil and sensitivity encoding (SENSE) with an underlying sampling factor of 2. Functional MRI images were acquired using a T_2*-weighted gradient-echo planar imaging sequence, which covered the whole brain (repetition time: 10 s, acquisition time: 2 s, echo time (TE): 30 ms, flip angle: 90°). Thirty-two axial slices with a slice thickness of 3.25 mm and interslice gap of 0.75 mm were acquired in an ascending order (resolution: $2.19 \times 2.19 \times 4.00$ mm; field of view $280 \times 224 \times 128$ mm). Quadratic shim gradients were used to correct for magnetic field inhomogeneities. T_1 images were acquired for all subjects (resolution $= 1.20 \times 0.93 \times 0.93$ mm). Participants listened to sounds within the scanner using an MR-compatible binaural headphone set (MR confon GmbH, Magdeburg, Germany). All the stimuli were presented using E-Prime software (Psychology Software Tools, Inc., Pittsburgh, PA, USA) installed on an MR interfacing integrated functional imaging system (Invivo Corporation, Orlando, FL, USA).

Data were acquired using sparse acquisition, which ensured that the stimuli were presented in silence (Hall et al. 1999). Stimuli were presented during a 7.5-s MR silent period, which was followed by a 2-s image acquisition and a 0.5-s silence. Two runs of data were acquired, with each run consisting of 24 trials of each condition presented in a pseudorandomized order (96 trials/volumes per run). A total of 192 trials/volumes were acquired for each subject. Each trial comprised three randomly selected unique sentences from one experimental condition with each sentence lasting <2 s in duration. Subjects listened passively to the sentences in the scanner and were instructed to try and understand each sentence.

Data Analysis

Univariate Analysis

Data were analyzed using Statistical Parametric Mapping (SPM8; http://www.fil.ion.ucl.ac.uk/spm/, last accessed March 25, 2013). Scans were realigned, unwarped, and spatially normalized to 2 mm³ isotropic voxels using the parameters derived from the segmentation of each participant's T_1-weighted image, and smoothed with a Gaussian kernel of 10-mm full-width at half maximum. A first-order finite impulse response (FIR) filter with a window length equal to the time taken to acquire a single volume—a box car function—was used to model the hemodynamic response. The 4 stimulus conditions (and 6 movement regressors of no interest) were entered into a general linear model at the first level. A 2×2 repeated-measures analysis of variance (ANOVA) with the factors intelligibility (+/−) and spectral detail (+/−) was conducted at the second level using the con images generated from the first level. The pairwise subtraction of all the conditions including the simple effects of (clear − rot) and (NV − rotNV) were examined using separate repeated-measures t-tests. All statistical maps were thresholded at peak level $P < 0.001$ (uncorrected) with a false discovery rate (FDR) correction of $q < 0.05$ at the cluster level. Activations were localized using SPM anatomy (Eickhoff et al. 2005). Follow-up analyses were conducted on functionally defined ROIs. Data were extracted from $7 \times 7 \times 7$ voxel cubes ($14\,\text{mm} \times 14\,\text{mm} \times 14\,\text{mm} = 2744\,\text{mm}^3$) using the Marsbar toolbox (http://marsbar.sourceforge.net/) (Brett et al. 2002). The location of these ROIs was defined by constructing ROIs around the peaks of the main positive effect of intelligibility in the anterior and posterior temporal cortex. Note that this does not constitute "double dipping" as these ROIs were used to examine between region differences—in a fully balanced design, any statistical bias will be equivalent between regions (see Kriegeskorte et al. (2009) supplementary materials and Friston et al. (2006)).

Multivariate Pattern Analysis

Pattern analysis involves using an algorithm to learn a function that distinguishes between experimental conditions using the pattern of voxel activity. Typically, data are divided into training and test sets; following a training phase, the success of the function is evaluated by assessing its ability to correctly predict the experimental conditions

associated with previously unseen brain images. The underlying assumption is that if the function successfully predicts the experimental conditions from the previously withheld images at a level greater than chance, then there is information within those images concerning the conditions. Pattern analysis considers the pattern of activation across multiple voxels, and this allows weak information at each voxel to be accumulated such that voxels that do not carry information individually can do so when jointly analyzed (Haynes and Rees 2006). Furthermore, as classification does not necessarily require spatial smoothing, it can afford a very high spatial specificity.

A linear support vector machine (SVM) is a discriminant function that attempts to fit a linear boundary separating data observed in different experimental conditions. When applied to fMRI analysis, an SVM attempts to fit a linear boundary that maximizes the distance between the most similar training examples from each condition within a multidimensional space with as many dimensions as voxels. These examples are referred to as the support vectors and are the training examples which are most difficult to separate. In the present study, they refer to the brain volumes where neural responses to intelligible and unintelligible sounds are most closely matched. The separating boundary is the direction in the data of maximum discrimination. A weight vector lies orthogonal to this boundary and is the linear combination or weighted average of the support vectors. Every voxel receives a weight, with larger weights indicating voxels that contribute more to classification. Given a positive and a negative class (+1 = intelligible speech and −1 = unintelligible sounds), a positive weight for a voxel means that the weighted average in the support vectors for that voxel was higher for listening to intelligible speech when compared with unintelligible sounds, and a negative weight means that the weighted average was lower for intelligible speech relative to unintelligible sounds (Mourao-Miranda et al. 2005). The classification prediction (whether an unseen example is classified as belonging to the intelligible or unintelligible class), is achieved by summing the activation values at each voxel multiplied by their associated weight value, and adding a bias term. If the resulting value is greater than zero, an unseen example will be classified (in the case of this experiment) as an intelligible speech trial, and if that value is less than zero it will be classed as an unintelligible trial. As the classification solution derived from SVMs is based on the whole spatial pattern, local inferences about single voxels should be interpreted only within the context of their contribution to a wider discriminating pattern.

Functional images were unwarped, realigned to the first acquired volume, and normalized, but not smoothed, using SPM8. Linear and quadratic trends were removed, and the data were z-score transformed by run within voxel (to remove amplitude differences between runs) and by trial across voxels (to remove overall amplitude differences between individual trials). This removes differences in the overall amplitude of the signal between runs and ensures that classification is achieved based on differences in the pattern of voxel responses between the conditions, rather than on an overall increase in response to one condition over another in all the voxels within a searchlight or ROI. Training/test examples were constructed from single brain volumes. Data were separated into training and test sets by run to ensure that training data did not influence testing (Kriegeskorte et al. 2009). The first classifier was trained on the first run and tested on the second, and vice versa for the second classifier. The "true" accuracy of classification was estimated by averaging the performance of the 2 classifiers for each subject.

The activation patterns submitted to the classification analyses were defined using (1) a whole-brain searchlight approach in which classification was conducted at each and every voxel in turn using the surrounding local neighborhood of voxels (Kriegeskorte et al. 2006) and (2) using an anatomical mask in which classification was conducted using all voxels within the bilateral temporal cortex (including PAC), and an additional control region within the visual cortex.

Searchlight Analyses

Searchlight analyses were conducted on the whole-brain volume with each searchlight consisting of a cube of $7 \times 7 \times 7$ voxels (14 mm × 14 mm × 14 mm = 2744 mm^3), using the searchmight toolbox (http://minerva.csbmb.princeton.edu/searchmight/, last accessed March

25, 2013) (Pereira and Botvinick 2011) and a linear SVM with the margin equal to 1/the number of voxels in each searchlight (1/343). Classifications of each intelligibility contrast were conducted: (clear vs. rot) and (NV vs. rotNV). Classification maps were generated for each subject with the value at each voxel reflecting the classification accuracy of the surrounding local neighborhood (proportion correct). Each classification accuracy value was subtracted from 0.5 (chance level) to center the values on zero, and these images were then submitted to a random effects one-sample t-test using SPM8. Classification maps were thresholded at a peak level $P < 0.001$ (uncorrected) with FDR correction at the cluster level of $q < 0.05$.

Anatomical Mask Analyses

Classifications of each intelligibility contrast were also conducted using an anatomical mask that included all the voxels in the bilateral temporal and auditory cortices and in a control region (the inferior occipital gyrus). This allowed us to understand how each voxel within the bilateral temporal lobes contributed to the classification of intelligible speech. The linear SVM from the Spider toolbox (http://www.kyb.tuebingen.mpg.de/bs/people/spider/, last accessed March 25, 2013), with the Andre optimization and a hard margin, was used to train and validate models. A large anatomical ROI was constructed that consisted of bilateral PAC, defined using the maximum probability maps from the SPM anatomy toolbox (Eickhoff et al. 2005), and the bilateral superior, middle, and inferior temporal gyri taken from the AAL ROI library. These temporal lobe ROIs had previously been defined by hand on a brain matched to the MNI/ICBM template using the definitions of Tzourio-Mazoyer et al. (2002) and are available via the Marsbar toolbox. An additional control region, the inferior occipital gyrus, was used to validate the analysis approach—this was also derived from the AAL library. For each intelligibility classification, the classification accuracy and the weight vector were extracted.

Results

Univariate Analysis

Okada et al. presented the whole-brain univariate analysis for the main positive effect of intelligibility, that is, the average of the response to clear and NV relative to the average of rot and rotNV. Here, we conduct a full factorial analysis to examine the neural response associated with the main positive effect of intelligibility and spectral detail, and their interaction. An idealized intelligibility response would be described by a region that responded equivalently and positively to the 2 intelligible conditions (clear and NV), and equivalently (and negatively) to the 2 unintelligible conditions (rot and rotNV). An interaction between these factors might identify regions showing a differential response to the 2 simple intelligibility effects: (clear − rot) and (NV − rotNV).

Main Positive Effect of Spectral Detail: (Clear + rot) − (NV + rotNV)

The main effect of spectral detail gave rise to bilateral clusters of activation focused within the PAC [12.3% in the left and 15.1% of the cluster in the right were located in PAC region TE 1.0 and 1.1 (Morosan et al. 2001)] and along the length of the superior temporal gyrus (STG; Fig. 1). Peak level activations were found in PAC and the STG bilaterally. This robustly bilateral activation contrasts with Scott et al. (2000), which evidenced a solely right lateralized response to the equivalent contrast. The increased power of this study, in which there were many more repetitions of the stimuli and measurements of the neural response, is likely to explain this difference.
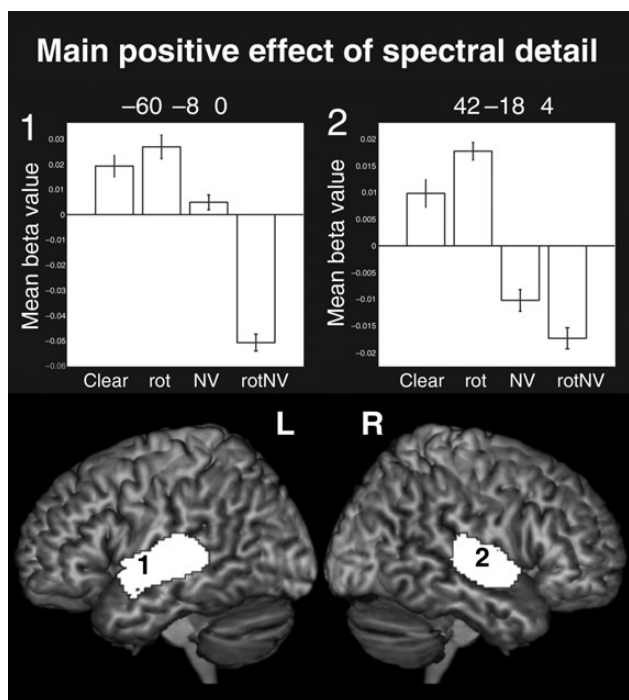
Main positive effect of spectral detail

−60 −8 0   42 −18 4

Figure 1. Surface renderings of the main positive effect of spectral detail, peak level uncorrected $P < 0.001$, FDR cluster corrected at $q < 0.05$. Plots show the mean beta value with error bars representing the standard error of the mean corrected for repeated-measures comparisons. Note that since there was no implicit baseline, the plots are mean centered to the overall level of activation to all conditions.

Main Positive Effect of Intelligibility:
(clear + NV) − (rot + rotNV)

The main effect of intelligibility was associated with clusters of activity that spread along the full length of the superior and middle temporal gyrus in the left, and the mid to anterior superior/middle temporal gyrus in the right hemisphere (Fig. 2A and Table 1). 0.6% of the clusters fell in TE 1.2 in the left hemisphere and 0.2% fell in TE 1.2 in the right hemisphere. Peak level activations were found in the bilateral STS, left fusiform gyrus, left parahippocampal gyrus, and the left hippocampus. The pattern of activation observed for this contrast is very similar to that observed by Okada et al. Response plots are shown from peak activations in the left anterior and posterior STS and the right anterior STS (no peaks were found in the right posterior STS). Plots from the left posterior STS suggested that there was a larger relative positive difference between NV and rotNV, when compared between clear versus rot (Fig. 2A, plot 1). Plots from bilateral anterior regions showed a seemingly more equivalent response between those contrasts (Fig. 2A, plots 2 and 3).

Interaction (f-test)

The interaction was associated with activation that spread predominantly across the mid to posterior superior and middle temporal gyri bilaterally (Fig. 3) with peak level activations found in the bilateral STG and STS. We explored these interactions by examining the simple effects while inclusively masking for the interaction at the same threshold. This highlighted 2 particularly interesting interaction patterns. One pattern was characterized by increased responses to rotated speech compared with all the other conditions—this was

particularly true of the response in the right planum temporale (e.g. Fig. 3, plot 3). The second pattern was characterized by an increased response to clear, rot, and NV relative to rotNV, in the absence of evidence of a difference in the response between clear, rot, and NV (see the response in the left posterior STS, Fig. 3, plot 2).

Simple Intelligibility Effects and Their Conjunction

Consistent with the identification of a significant interaction, the 2 simple intelligibility effects: (clear − rot) and (NV − rotNV) gave rise to very different statistical maps (Fig. 4A,B). The contrast of (clear − rot) solely activated the left anterior STS. This was in contrast to (NV − rotNV), which gave rise to broad activation extending along the length of the STG and STS, middle temporal gyri, and extending into the inferior temporal gyri bilaterally. 1.7% and 4.1% of the cluster extended into TE 1.0 and 1.2 in the left hemisphere, respectively, and 9.2% and 8.4% into TE 1.0 and 1.2, respectively, in the right hemisphere. Peak level activations were found in the bilateral STG/STS.

A conjunction analysis was carried out to isolate activations common to the 2 simple intelligibility effects. This statistical analysis reflects the original concept of the Scott et al. (2000) design, which used more than one intelligibility subtraction in an attempt to isolate a more acoustically invariant intelligibility response. It has been noted that there has been confusion in the past concerning the interpretation of conjunction analyses (Nichols et al. 2005). For the sake of clarity, there are 2 commonly used conjunction analyses: The global null conjunction and the more recently introduced conjunction null analysis. Narain et al. (2003) conducted the global null conjunction of the 2 intelligibility contrasts. They found activation in the left anterior and posterior STS for this analysis. However, as they used the global null conjunction, it is only possible to draw the inference that there was an effect in one or more of the intelligibility contrasts, rather than necessarily in both of them. Here, unlike Narain et al., we conduct the conjunction null rather than the global null conjunction. Statistical maps of the conjunction null show voxels that survive the specified threshold across all the individual subtractions that make up the conjunction, this allows a stronger inference to be made that there is an intelligibility effect in all the intelligibility contrasts considered. The statistical map resulting from the conjunction null was identical to the simple effect of (clear − rot)—that is, the activation generated by (NV − rotNV) encompassed the activation of (clear − rot), but not vice versa (Table 1). Note that the conjunction null map is not shown in the figure, as it was identical to (clear − rot). Our results therefore extend those of Narain et al. in showing that the left anterior STS was significantly activated by both individual intelligibility contrasts at a corrected threshold.

To directly compare the intelligibility response across regions, we extracted data around the peaks identified by the main effect of intelligibility. These were located in left anterior [−58 2 −16], right anterior [56 2 −18], and left posterior STS [−62 −34 0] (Fig. 5E). Note that we did not extract data from the right posterior STS (shown in gray) as activation did not spread to that region. A $3 \times 2 \times 2$ repeated-measures ANOVA with factors: Location (left anterior, right anterior, and left posterior), intelligibility (+/−), and spectral detail (+/−) was conducted. This showed there to be main effects of intelligibility ($F_{1,11} = 116.648$, $P < 0.001$) reflecting an increased response to intelligible speech
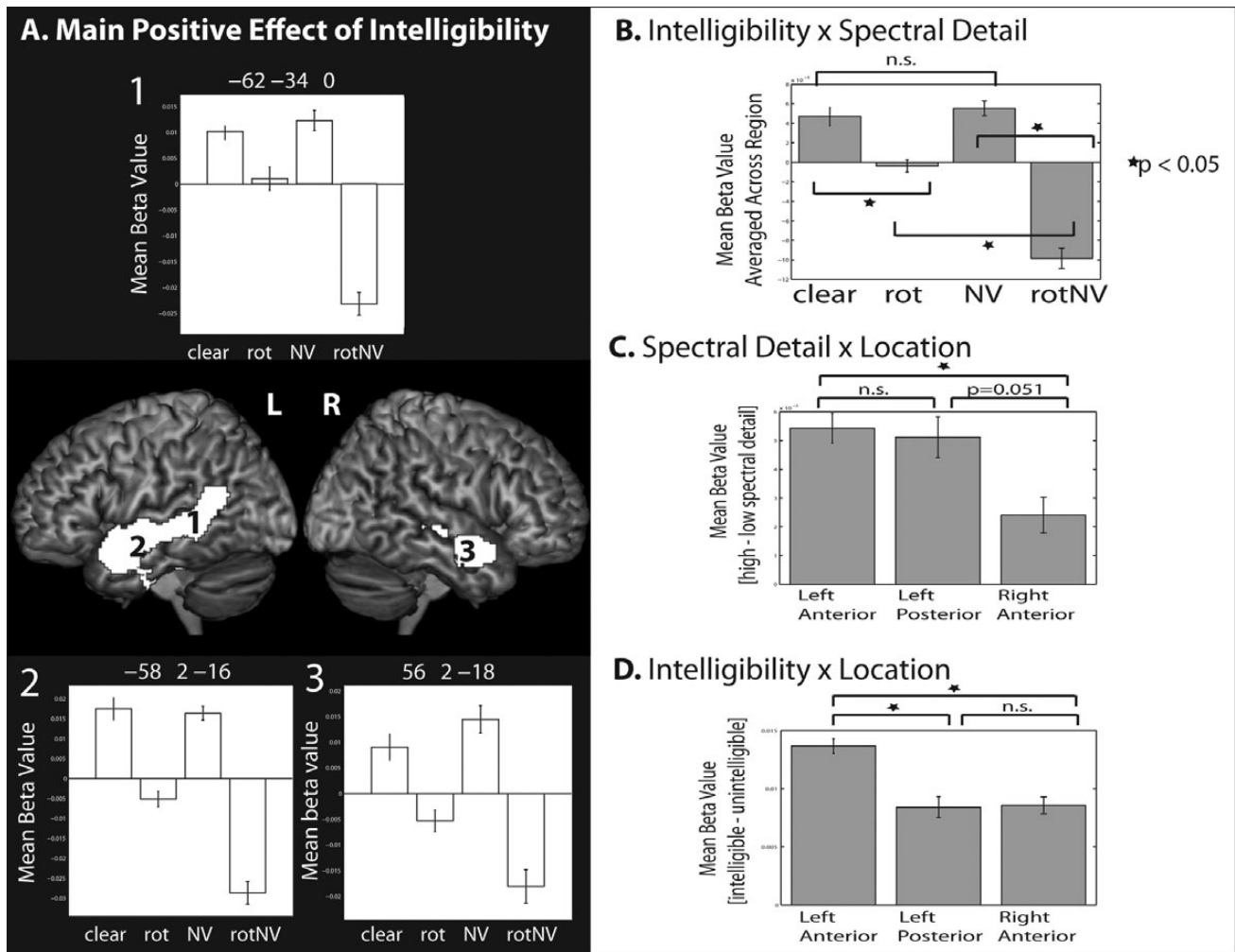
Figure 2. (A) Surface renderings of the main positive effect of intelligibility, peak level uncorrected $P < 0.001$, FDR cluster corrected at $q < 0.05$. (B) Data averaged across the region for each condition, illustrating the intelligibility × spectral detail interaction. (C) Data showing (high spectral detail − low spectral detail conditions) by location, illustrating the spectral detail × location interaction. (D) Data showing (intelligible − unintelligible conditions) by location, illustrating the intelligibility × location interaction. All plots show the mean beta value and error bars representing the standard error of the mean corrected for repeated-measures comparisons.

Table 1

Peak level activations for the main positive effect of intelligibility and the conjunction null of the intelligibility contrasts at peak level $P < 0.001$ uncorrected, FDR cluster corrected at $q < 0.05$

Location	MNI			Extent	Z
	x	y	z		
Main effect of intelligibility					
Left anterior STS	−58	2	−16	2770	7.16
Left posterior STS	−62	−34	0		6.28
Left posterior STS	−52	−50	18		4.49
Right anterior STS	56	2	−18	401	5.43
Right anterior STS	52	12	−18		5.22
Right mid-STS	50	−22	−6	183	4.09
Left fusiform gyrus	−24	−30	−24	162	3.73
Left parahippocampal gyrus	−20	−24	−20		3.71
Left hippocampus	−24	−12	−18		3.67
Conjunction null: (clear − rot) ∩ (NV − rotNV)					
Left anterior STS	−58	4	−20	256	4.48
Left anterior STS	−60	−2	−12		4.07

MNI, Montreal Neurological Institute

when compared with unintelligible sounds, and a main effect of spectral detail ($F_{1,11} = 13.731$, $P = 0.003$) reflecting an increased response to conditions containing spectral detail compared with those without. There was no main effect of location

($F_{2,22} = 0.750$, $P = 0.484$). The interactions of spectral detail × intelligibility ($F_{1,11} = 36.767$, $P < 0.001$), location × spectral detail ($F_{2,22} = 4.727$, $P = 0.020$), and location × intelligibility ($F_{2,22} = 10.372$, $P = 0.001$) were significant. The 3-way interaction was not significant, showing that there was no evidence of a difference in the interaction between intelligibility and spectral detail, as expressed across the regions ($F_{2,22} = 2.284$, $P = 0.126$). We decomposed the 2-way interactions by examining simple effect contrasts. In the case of the interaction between intelligibility and spectral detail, there was a significant difference between clear and rot ($t_{(11)} = 4.592$, $P = 0.001$), NV and rotNV ($t_{(11)} = 10.818$, $P < 0.001$), and rot and rotNV ($t_{(11)} = 6.484$, $P < 0.001$) in the absence of a difference between clear and NV ($t_{(11)} = 0.591$, $P = 0.566$; Fig. 2B). The interaction was driven by a relative deactivation to rotNV; in other words, there was no evidence that the intelligible conditions (NV and clear) differed from one another, but the unintelligible conditions (rot and rotNV) differed from one another dependent on the level of spectral detail. The interaction between location and spectral detail was driven by a larger response to conditions containing spectral detail, when compared with those without, in the left anterior STS when compared with the right anterior STS
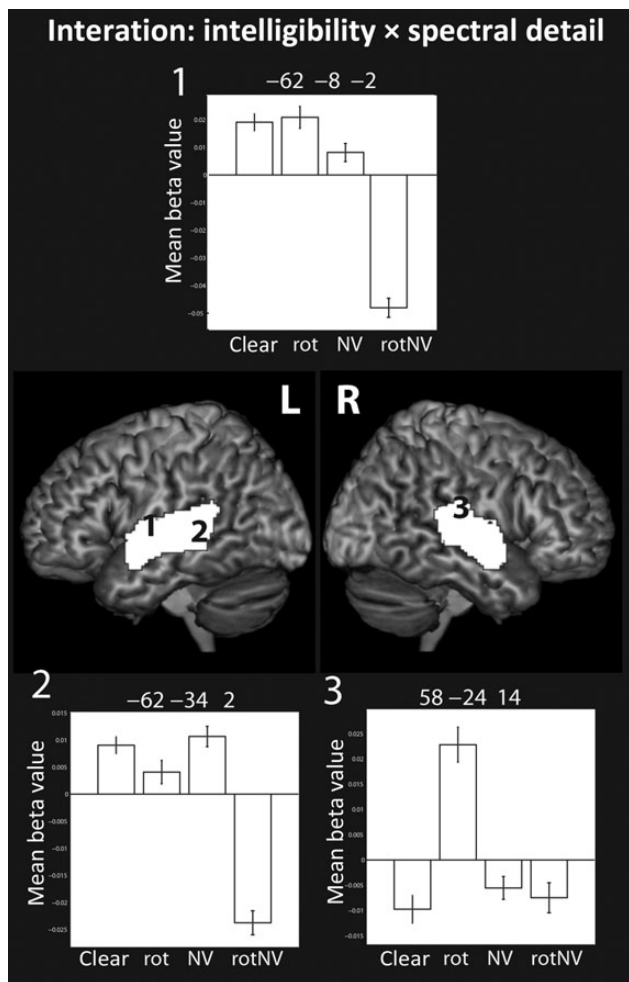
Figure 3. Surface renderings of the interaction between spectral detail and intelligibility, peak level uncorrected $P < 0.001$, FDR cluster corrected at $q < 0.05$. Plots show the mean beta value with error bars representing the standard error of the mean corrected for repeated-measures comparisons.

Figure 4. Surface renderings of the simple effects of (*A*) (clear − rot), (*B*) (NV − rotNV), peak level uncorrected $P < 0.001$, FDR cluster corrected at $q < 0.05$. Note that the conjunction null (not shown) is identical to the response to (clear − rot). Plot shows the mean beta value with error bars representing the standard error of the mean corrected for repeated-measures comparisons.

($t_{(11)} = 3.359$, $P = 0.006$) and the difference between the left posterior when compared with the right anterior STS was marginally significant ($t_{(11)} = 2.192$, $P = 0.051$), in the absence of a difference between the left anterior and left posterior STS ($t_{(11)} = 0.289$, $P = 0.778$; Fig. 2*C*). The interaction between location and intelligibility was driven by a larger response to intelligible speech, when compared with unintelligible sounds, in the left anterior STS as contrasted with both the left posterior STS ($t_{(11)} = 3.831$, $P = 0.003$) and the right anterior STS ($t_{(11)} = 5.118$, $P < 0.001$), in the absence of a difference between the left posterior and right anterior STS ($t_{(11)} = 0.106$, $P = 0.918$; Fig. 2*D*). This demonstrated that the left anterior STS exhibited the strongest univariate intelligibility effect.

Multivariate Pattern Analysis

Searchlight Analyses

Okada et al. conducted pattern classifications within single small cubes of data in the left and right anterior and posterior STS (and the right mid-STS), defined in each subject by identifying local maxima within these regions in the univariate contrast of (clear − rot). We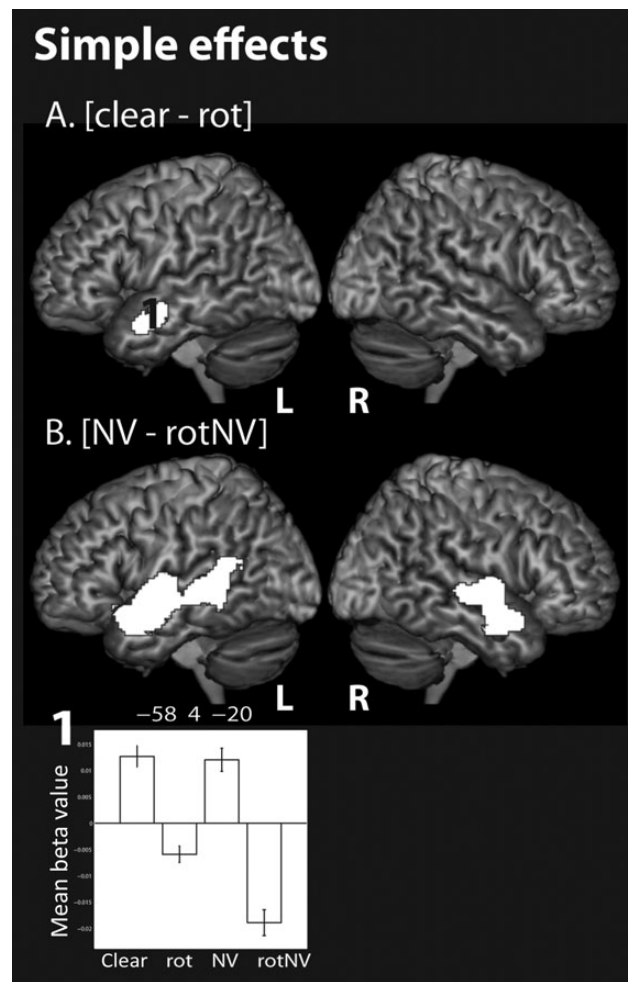 extend their analysis by conducting a searchlight analysis in which we examine local multivariate information across the whole-brain volume. This has the advantage of allowing us to more fully probe the response within the temporal lobes rather than just within a small number of selective sites, as well as allowing us to examine the multivariate response of frontal and parietal cortices, regions that have previously been shown to be associated with intelligibility responses (Davis and Johnsrude 2003; Obleser et al. 2007; Abrams et al. 2012).

Random effects one-sample *t*-tests were conducted on searchlight accuracy maps generated from the group of subjects. In the case of classifications of clear versus rotated speech (Fig. 5*A*), above-chance classification was identified within a fronto-temporo-parietal network, including the bilateral inferior frontal gyri (pars opercularis and triangularis), left angular gyrus, and both anterior and posterior superior/middle temporal gyri. Small clusters were also identified in the right middle frontal gyrus, supplementary motor area, and bilateral insulae as well as the left thalamus and caudate nucleus. In the left hemisphere, 7.1% of the temporal lobe cluster fell within the PAC, and this was the case for 12.5% of the temporal lobe cluster in the right hemisphere.
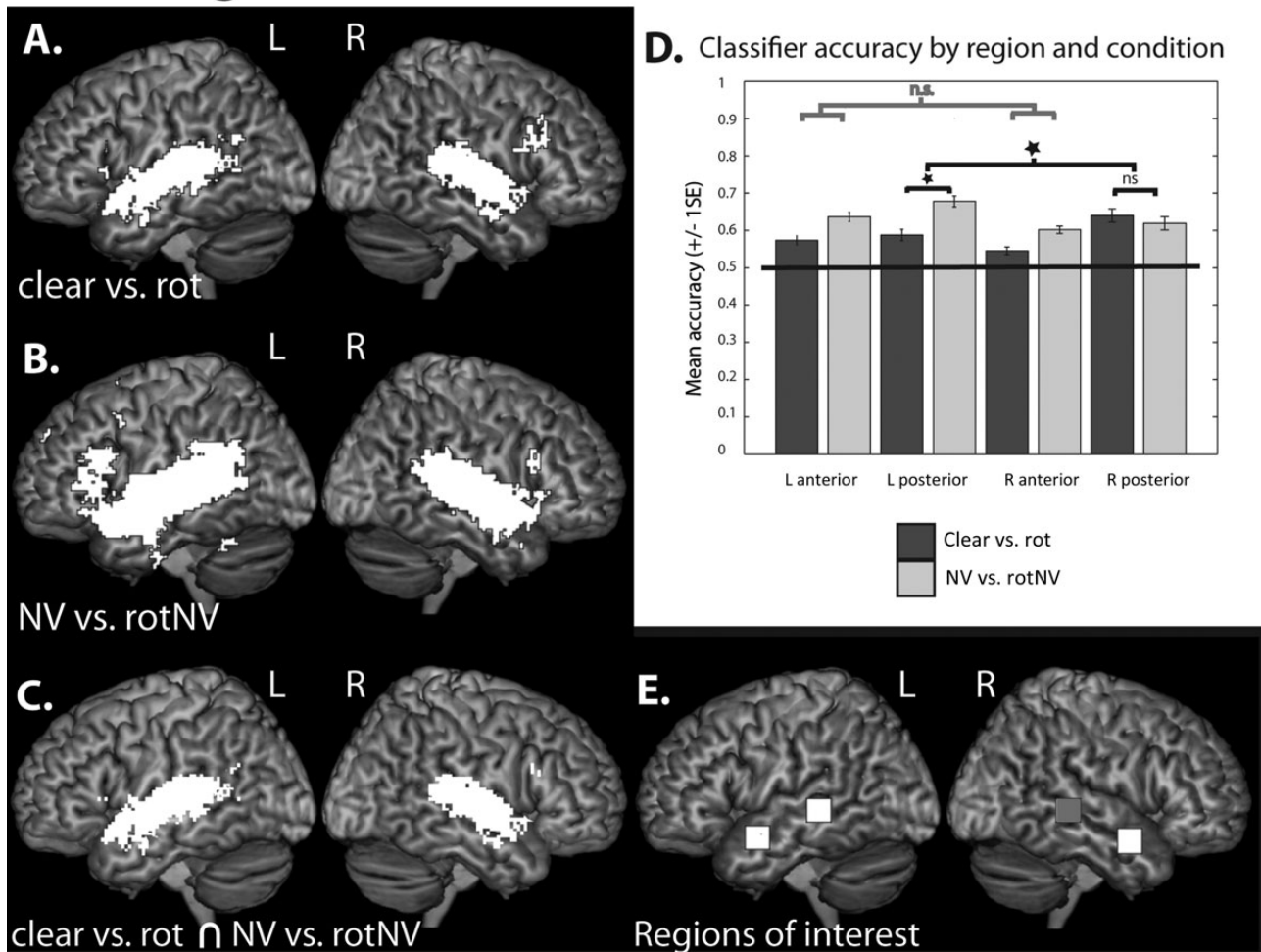
Searchlight classifications

Figure 5. Whole-brain searchlight classifications of (*A*) (clear vs. rot), (*B*) (NV vs. rotNV), peak level $P < 0.001$ uncorrected, FDR cluster corrected at $q < 0.05$. (*C*) Voxels commonly implicated in classifications of (clear vs. rot) and (NV vs. rotNV). (*D*) Plots of classifier accuracy (proportion correct) extracted for (clear vs. rot) and (NV vs. rotNV) within the left anterior, left posterior, right anterior, and right posterior STS. Plots show mean accuracy with error bars representing the standard error of the mean corrected for repeated-measures comparisons. (*E*) Regions of interest, $7 \times 7 \times 7$ voxel cubes in the left anterior [−58 2 −16], left posterior [−62 −34 0], right anterior [56 2 −18], and right posterior STS (gray) [62 −34 0] (MNI coordinates).

In the case of classifications of (NV vs. rotNV), above-chance classification was also identified in a bilateral fronto-temporo-parietal network, including the inferior frontal gyri (pars triangularis, pars opercularis, and pars orbitalis), inferior parietal cortices, and both anterior and posterior superior/middle temporal cortices (Fig. 5*B*). Additional small clusters were found in the bilateral cerebellum, left inferior temporal gyrus, precentral gyrus, supplementary motor area, precuneus, and superior and medial frontal gyrus. 3.3% of the left hemisphere cluster fell within TE 1.0, and 6.4% and 4.1% of the right hemisphere cluster fell within TE 1.0 and TE 1.2, respectively.

There were a large number of voxels from clusters that showed above-chance classification for both (clear vs. rot) and (NV vs. rotNV). These voxels were found predominately in bilateral anterior and posterior temporal cortices, but small numbers of voxels were also identified in the bilateral inferior frontal gyri (pars opercularis and triangularis) and the left angular gyrus (Fig. 5*C*).

The classification accuracies from the searchlight maps from each subject were extracted from single voxels (the

classification scores at these locations were reflective of the accuracy of the surrounding $7 \times 7 \times 7$ voxels) at the same locations in the left anterior, right anterior, and left posterior STS that were examined in the earlier post hoc univariate analyses (Fig. 5*E*). A right posterior STS ROI was constructed by taking the homotopic equivalent of the left posterior STS region (shown in gray). A $2 \times 2 \times 2$ repeated-measures ANOVA with factors: Hemisphere (left and right), position (anterior and posterior), and contrast (clear vs. rot and NV vs. rotNV) was conducted (Fig. 5*D*). There was a significant main effect of position ($F_{1,11} = 8.020$, $P = 0.016$) with posterior regions ($M = 0.63$, proportion correct) showing higher accuracy than anterior regions ($M = 0.59$) and a main effect of contrast ($F_{1,11} = 13.991$, $P = 0.003$) with classifications of NV versus rotNV ($M = 0.63$) higher than clear versus rot ($M = 0.59$), but no main effect of hemisphere ($F_{1,11} = 0.802$, $P = 0.390$). There were no significant 2-way interactions: Hemisphere × position ($F_{1,11} = 0.524$, $P = 0.484$), hemisphere × contrast ($F_{1,11} = 2.471$, $P = 0.144$), and position × contrast ($F_{1,11} = 0.711$, $P = 0.417$). There was a significant 3-way interaction of hemisphere ×

position × contrast ($F_{1,11} = 11.849$, $P = 0.006$). Examining the simple interaction effects we identified that the simple hemisphere × contrast interaction was significant in the posterior ($F_{1,11} = 6.721$, $P = 0.025$) but not in the anterior regions ($F_{1,11} = 0.033$, $P = 0.858$). We further examined the second-order simple effects in the posterior regions; this showed that accuracy in the left posterior STS was significantly higher for NV versus rotNV when compared with clear versus rot ($P = 0.011$), while there was no significant difference between the contrasts in the right posterior STS ($P = 0.567$). To summarize classification accuracies were higher in posterior when compared with the anterior temporal cortex and were higher for classifications of (NV vs. rotNV) when compared with (clear vs. rot), this was likely to be driven by the high classifications of (NV vs. rotNV) in the left posterior STS.

Anatomical Mask Analysis

Okada et al. conducted classification analyses within small local neighborhoods of voxels within the temporal cortex. Here, we extend this analysis by conducting classifications using the entirety of the bilateral temporal cortex, including PAC. In doing so, we can gain insights into how multivariate information is integrated across the temporal cortex in the decoding of intelligible speech.

Classification was conducted using an ROI consisting of the entire bilateral temporal cortex, including PAC and a control region, the inferior occipital gyrus. Cross validation using the run structure demonstrated that this ROI correctly classified 0.74 (proportion correct) of volumes of (clear vs. rot), and 0.79 of volumes of (NV vs. rotNV) correctly (chance level 0.50; Fig. 6*A*i,*B*i); and in both cases, classification was significantly greater than chance (clear vs. rot: $t_{(11)} = 5.961$, $P < 0.001$; NV vs. rotNV: $t_{(11)} = 9.949$, $P < 0.001$) when Bonferroni correcting for 2 tests ($P = 0.025$). There was no evidence that the inferior occipital gyrus performed at a level greater than chance (clear vs. rot: $P = 0.812$; NV vs. rotNV: $P = 0.967$). Having demonstrated above-chance classification using the bilateral temporal cortex in both intelligibility classifications, we extracted the weight vector from a classifier trained using both runs of data for each intelligibility classification.

For classifications of (clear vs. rot) and (NV vs. rotNV), the weight vector values were multiplied by the averaged response to the intelligible trials in each instance (i.e. the average of the clear trials was multiplied by the weight vector from the classification of (clear vs. rot) and the average of the NV trials was multiplied by the weight vector from the classification of (NV vs. rotNV). This allowed us to understand how each voxel directly contributed to the classifiers' prediction in favor of an intelligible trial being classified as such for each intelligibility classification. Voxels with resulting positive values contribute to increasing the likelihood that a trial will be classified as an intelligible trial. To understand how these values varied, each value was expressed as a percentage of the largest overall positive value and plotted against that value's ranked magnitude expressed as a percentage of the total number of values. This plot demonstrated that the values decayed rapidly in relation to the largest positive value, with values ranking in the top 15% accounting for the majority of the range (Fig. 6*A*ii,*B*ii). Note that all the values in the top 15% were positive for every subject and for both classifications. As all the values were positive, this could be achieved by 1 of 2 mechanisms, either (1) the average response pattern for the intelligible condition at a particular

voxel was positive and the weight vector value at that location was also positive or (2) the average response pattern at a particular voxel was negative and the weight vector value was also negative. Note from the above, we can deduce that if a voxel received a positive weight then it follows that the response at that voxel for the averaged intelligible trial was also positive (i.e. there was a relative increase in signal for intelligible when compared with unintelligible trials), and if a voxel received a negative weight then the response at that voxel for the averaged intelligible trial was also negative (i.e. there was a relative increase in signal for unintelligible when compared with intelligible trials). In order to differentiate these effects at each voxel, we separated positively and negatively weighted voxels when generating binary maps of the voxels that contributed most to classification, for voxels in the top 5%, 10%, and 15% of values. These were then summed in order to establish the confluence across subjects and visualized for different percentage bandings.

These maps show that, in the case of both intelligibility classifications, the voxels contributing most to the classification of intelligible speech, in which voxel activation was higher for intelligible speech when compared with unintelligible sounds, were found in the left anterior temporal lobe (Fig. 6*C*,*D*). In contrast, voxels lateral and posterior to PAC in the STG and planum temporale (and to a lesser extent in the posterior STS) contributed to classification, but as a result of a relative increase to unintelligible sounds when compared with intelligible speech. We then identified positive and negative weights featuring in the top 15% of weights common to more than 4 participants and to both intelligibility contrasts (Fig. 6*E*). This confirmed the previous observation that the left anterior temporal lobe was associated with informative voxels in which activation was higher for intelligible speech when compared with unintelligible sounds, with the center of mass for these weights located at [−56 −4 −12]. In contrast, regions closer to PAC in the STG and planum temporale were also informative, but reflected greater activation to unintelligible sounds when compared with intelligible speech, with the center of mass of voxels in the left hemisphere found at [−46 −32 13] and in the right hemisphere at [54 −27 13].

Discussion

In this study, we replicated and extended the findings of Okada et al. (2010). Similar to Okada et al., we found bilateral activation spreading along much of the STS for the main positive effect of intelligibility. Extending their findings, we demonstrated that there was a significant interaction between intelligibility and spectral detail in the bilateral mid-posterior superior temporal cortex. The interaction at a location in the left posterior STS was driven by a reduced relative response to rotated-noise-vocoded speech, in the absence of observed differences between clear, rotated, and noise-vocoded speech. Consistent with the demonstration of a significant interaction, the simple intelligibility effects of (clear − rot) and (NV − rotNV) gave rise to very different statistical maps and, indeed, only the left anterior STS was activated by the conjunction null of the 2 intelligibility contrasts. Follow-up analyses demonstrated that the intelligibility response in the left anterior STS was reliably stronger than in other regions. Further to the univariate analysis, we conducted multivariate analyses examining the contribution of local pattern information to
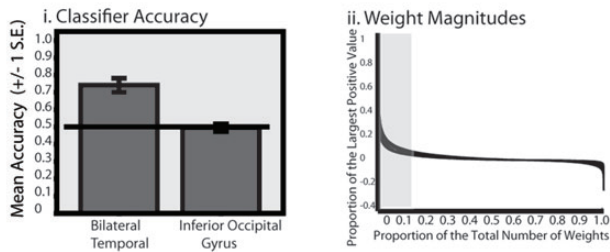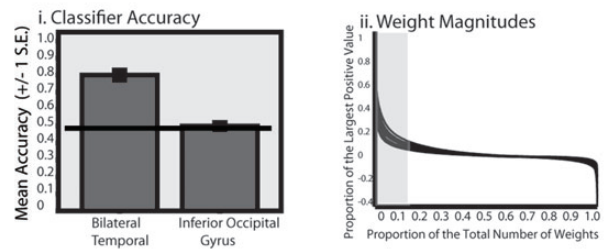
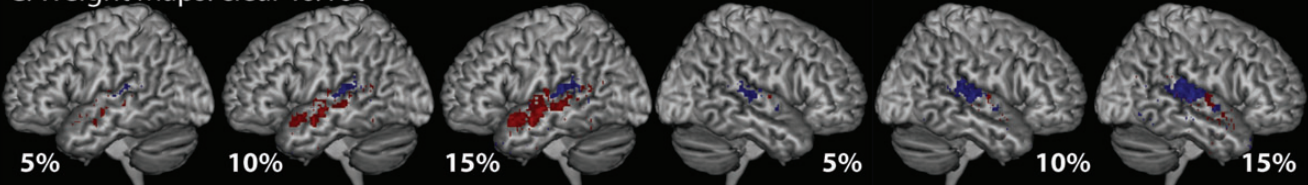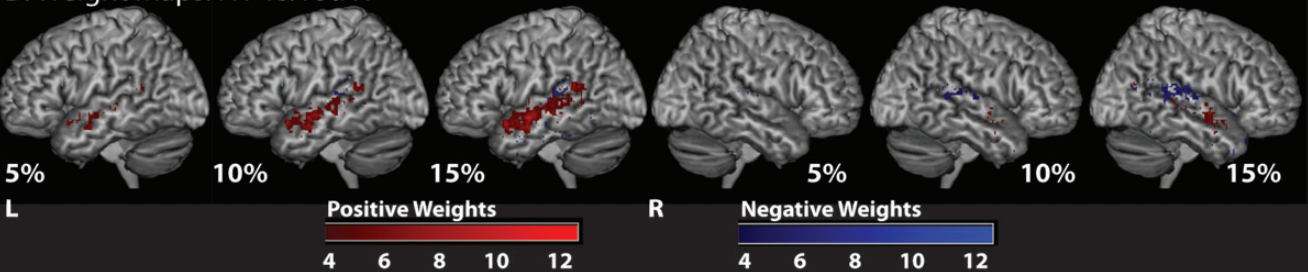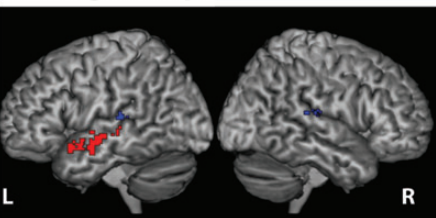**Figure 6.** Classifier accuracy and weight maps from classifications using the entirety of the bilateral temporal cortex (including PAC) and the inferior occipital gyrus (used as a control region). (*A*i) Classifier accuracy (proportion correct) for (clear vs. rot). (*A*ii) Weight magnitude values of weights from (clear vs. rot). (*B*i) Classifier accuracy (proportion correct) for (NV vs. rotNV). (*B*ii) Weight magnitude values of weights from (NV vs. rotNV). (*C*) The most discriminative 5%, 10%, and 15% of classifier weights from a classifier trained to discriminate (clear vs. rot). Color gradient indicates the degree of concordance across subjects. (*D*) The most discriminative 5%, 10%, and 15% of classifier weights from a classifier trained to discriminate (NV vs. rotNV). Color gradient indicates the degree of concordance across subjects. (*E*) Weights featuring in the top 15% of weight values common to both (clear vs. rot) and (NV vs. rotNV) and more than 4 subjects.

classifications of intelligibility. Here, we replicated, using a whole-brain searchlight analysis, the findings of Okada et al. in showing above-chance classification within the bilateral anterior and posterior STS for each of the intelligibility classifications, and showed that there was more information in the bilateral posterior when compared with the bilateral anterior STS, albeit this was driven mainly by the high classification of (NV vs. rotNV) in the left posterior STS. We conducted an additional analysis in which classification was conducted using the whole of bilateral temporal cortex, rather than using small local neighborhoods of voxels. Using classifier weights from this analysis, we showed that voxels contributing most to classification and exhibiting a relative increase in response to intelligible speech when compared with unintelligible

sounds were found predominantly in the left anterior STS. In contrast, bilateral STG and the planum temporale contributed to classification, but as a result of a relative increase in activation to unintelligible sounds.

The most robust univariate intelligibility effects were located in the left anterior STS consistent with our original Scott et al. (2000) study. Note that we assume that intelligibility includes all stages of comprehension over and above early acoustic processing, and as such the intelligibility responses that we see likely reflect multiple processes including acoustic–phonetic, semantic, and syntactic processing, and associated representations. Okada et al. (2010) did not test for interactions between intelligibility and spectral detail, present the conjunction null of the simple effects, and conduct

univariate ROI analyses to differentiate the response within the anterior and posterior regions or plot data from peak level voxels. As a consequence, it is not possible to ascertain whether the univariate effects demonstrated in their study were also stronger in the left anterior STS or, indeed whether there were any regions that showed a differential magnitude of response to the 2 intelligibility contrasts. We have conducted a number of previous studies in which we have compared speech with complex acoustic baselines. In these studies, while activation in the left anterior STS has been consistently observed, responses in other temporal regions have been identified much less consistently (Scott et al. 2000, 2006; Narain et al. 2003; Spitsyna et al. 2006; Awad et al. 2007; Friederici et al. 2010). This may suggest that the univariate amplitude of response is lower, or that intersubject variability is higher in temporal regions outside of the left anterior STS. The classifier weight maps from the multivariate analyses provide converging evidence that increases in activation in anterior regions are associated with responses to intelligible speech. These findings are consistent with work, showing that the left anterior temporal cortex responds to specifies-specific vocalizations in the monkey (Poremba et al. 2004) and that in humans "phonemic maps", voice responses, and the "semantic hub" reside in the anterior temporal cortex (von Kriegstein et al. 2003; Obleser et al. 2006; Patterson et al. 2007; Leaver and Rauschecker 2010). The weight maps also provide support for the notion of a hierarchical speech processing system (Davis and Johnsrude 2003) in which regions closest to PAC engage in predominantly acoustic processing and are driven robustly by complex nonspeech sounds, and those further from primary auditory regions, in the STS, exhibit a preferential response to linguistically relevant stimuli.

A region of the left posterior STS was implicated in the interaction between intelligibility and spectral detail at the whole-brain level. This region was close, but slightly superior, to the center of the ROI in the left posterior STS examined in the post hoc univariate analyses (which had been defined based on the peak for the main effect of intelligibility). The peak implicated in the interaction showed no evidence of responding differentially to clear, rotated, and noise-vocoded speech. Note that the response profile at this location (Fig. 3, plot 2) is almost identical to the profile at a similar location in the original Scott et al. (2000) study (Fig. 2, plot 2, p 2403). It may be that this region of the left posterior STS engages in acoustic phonetic processing that supports the resolution of intelligible percepts, explaining its elevated response to the unintelligible rotated speech condition as well as to the intelligible stimuli. Rotated speech contains phonetic features such as the absence/presence of voicing, but these features do not generally give rise to intelligible sounds; rotated-noise-vocoded speech in contrast does not contain any recognizable phonetic features explaining its relative deactivation. A number of regions of the temporal cortex, especially within the right planum temporale, responded more strongly to rotated speech than to any other condition. Rotated speech retains much of the spectro-temporal structure of speech, including formants and a quasi-harmonic structure, making it arguably the most appropriate nonspeech baseline used to date. It does however differ from speech in a number of ways. For example, the rotation of fricatives results in broadband energy in the low frequencies, a feature not ordinarily characteristic of speech, and while rotation maintains the equal

spacing of the harmonics, it changes their absolute frequencies, giving rise to a slightly unusual pitch percept. These factors may contribute to why rotated speech drives some auditory regions more strongly than speech. We have recently developed alternative speech/nonspeech analogs by synthesizing, and further noise-vocoding, sinewaves tracking the formants of speech, and then combining the amplitude tracks from one sentence with the frequency tracks from another to generate unintelligible equivalents. These unintelligible sentences are arguably more closely matched acoustically and perceptually to their intelligible counterparts than rotated speech is to clear speech. In studies using these stimuli, we have found robust activation in the anterior and posterior temporal cortices when responses to the (mostly) intelligible condition were compared with the unintelligible condition (McGettigan and Evans et al. 2012; Rosen et al. 2011). However, as these studies did not directly compare the response of anterior with posterior regions, it is difficult to ascertain whether the response was stronger in anterior regions. As the intelligible condition was also degraded, these findings may be more informative in understanding the neural systems underlying effortful speech comprehension than speech comprehension more generally.

The fact that there were some regions of the temporal cortex that responded more strongly to rotated than to clear speech highlights the need to account for how the relative magnitude of activation to each condition contributes to classification accuracy. It would be possible for a region to exhibit a high level of classification solely because of a relative increase in signal to the unintelligible conditions, which would be hard to reconcile with the suggestion that an area coded a response to fully resolved intelligible speech percepts. To address this, we conducted searchlight analyses in which the mean level of activation had been removed for each trial—this prevented classification from being achieved as a result of a relative increase in signal across all the voxels within a searchlight to one condition over another. By conducting a searchlight analysis, rather than classification within single small ROIs in the temporal cortex as was conducted by Okada et al., we were able to more fully map the response of the temporal cortex and regions beyond. We chose not to adopt the approach of Okada et al. in conducting classifications of spectral detail and creating an "acoustic invariance index" (by comparing the relative accuracy of intelligibility and spectral detail classifications) because 6 band noise-vocoded and clear speech differ in both spectral detail and intelligibility, thus confounding the index. Indeed, one might imagine that differences in intelligibility between the clear and noise-vocoded speech would have been further exacerbated by continuous data acquisition in which sounds were played over the competing noise of the scanner (in contrast to the sparse acquisition conducted in this study).

Our results indicated that there was significant local information in the anterior and posterior STS bilaterally for both intelligibility classifications (unlike in the whole-brain univariate analysis in which this was the case only in anterior regions), and that posterior regions contained relatively more information than anterior regions. The identification of greater information in posterior regions seemingly contradicts the finding that classifier weights in anterior areas were associated with increases in activity for intelligible speech and the observation of stronger univariate intelligibility effects in the left

anterior STS. This may, however, reflect the fact that there are multiple ways that information, capable of discriminating responses to intelligible speech from unintelligible sounds, can be extracted from the fMRI signal. It remains to be seen which of the analysis methods best imitates the neural processes involved in encoding intelligible speech, or indeed whether our findings reflect the fact that intelligibility responses are encoded using multiple complementary coding systems.

The searchlight analysis also identified significant information within the inferior frontal and inferior parietal cortices. These findings are in keeping with a recent searchlight analysis conducted by Abrams et al. (2012) in which the inferior frontal and parietal cortices were shown to discriminate clear from rotated speech in the absence of a similar effect in those regions in an accompanying whole-brain univariate analysis. Indeed, while intelligibility responses have been shown in these regions previously (Davis and Johnsrude 2003; Obleser et al. 2007), they are arguably less consistently identified than in temporal lobe regions (Abrams et al. 2012). Our results, like those of Abrams et al. (2012), suggest that univariate analyses underestimate the extent of regions involved in responding to intelligibility. Indeed, these findings suggest that local information resides in a network of regions including anterior and posterior temporal, inferior parietal, and inferior frontal cortices. This is in accord with the suggestion that the comprehension of speech recruits multiple streams of processing radiating anteriorly and posteriorly from primary auditory regions (Davis and Johnsrude 2007; Peelle et al. 2010).

To conclude, we identified that the most robust univariate intelligibility effects were found in the left anterior STS, consistent with our previous findings. When multivariate classifications were conducted in which information could be integrated across the anterior and posterior temporal cortices, increases in activation in anterior regions were shown to be maximally discriminative in classifying intelligible speech, again indicating the relative importance of the left anterior STS. However, when classification was conducted within small local neighborhoods in which the mean activation to each condition was removed, we found greater information in the posterior when compared with anterior regions, and identified a much wider intelligibility network that included the inferior parietal and frontal cortices. These results are consistent with the suggestion that the comprehension of spoken sentences engages multiple streams of processing, including an anterior stream that shows evidence of being most strongly engaged when data are analyzed using univariate methods, and a posterior stream or streams that can be identified more readily (at a whole-brain level) with multivariate pattern-based methods. Hence, as has been suggested recently by Abrams et al. (2012), inconsistencies in the literature may be reconciled by the use of more sensitive multivariate methods that allow the identification of a wider intelligibility network. The exact contribution of each component of this network is unknown, and we hope that future research will focus on attempting to understand how these components and their interaction contribute to resolving intelligibility.

Funding

Notes

References

Abrams DA, Ryali S, Chen T, Balaban E, Levitin DJ, Menon V. 2012. Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's Areas. Cereb Cortex. (In press).

Awad M, Warren JE, Scott SK, Turkheimer FE, Wise RJS. 2007. A common system for the comprehension and production of narrative speech. J Neurosci. 27:11455–11464.

Bench J, Kowal A, Bamford J. 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Br J Audiol. 13:108–112.

Blesser B. 1972. Speech perception under conditions of spectral transformation. 1. Phonetic characteristics. J Speech Hearing Res. 15:5–41.

Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain, 2-6 June 2002, Sendai, Japan. Available on CD-ROM in NeuroImage, Vol. 16, No. 2.

Davis MH, Johnsrude IS. 2007. Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hearing Res. 229:132–147.

Davis MH, Johnsrude IS. 2003. Hierarchical processing in spoken language comprehension. J Neurosci. 23:3423–3431.

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

Friederici AD, Kotz SA, Scott SK, Obleser J. 2010. Disentangling syntax and intelligibility in auditory language comprehension. Hum Brain Mapp. 31:448–457.

Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN. 2006. A critique of functional localisers. Neuroimage. 30:1077–1087.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. 1999. "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp. 7:213–223.

Haynes JD, Rees G. 2006. Decoding mental states from brain activity in humans. Nat Rev Neurosci. 7:523–534.

Hickok G. 2009. The functional neuroanatomy of language. Phys Life Rev. 6:121–143.

Hickok G, Poeppel D. 2007. Opinion—the cortical organization of speech processing. Nat Rev Neurosci. 8:393–402.

Kaas JH, Hackett TA. 2000. Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci USA. 97:11793–11799.

Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. Proc Natl Acad Sci USA. 103:3863–3868.

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

Leaver AM, Rauschecker JP. 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci. 30:7604–7612.

McGettigan C, Evans S, Rosen S, Agnew ZK, Shah P, Scott SK. 2012. An application of univariate and multivariate approaches in fmri to quantifying the hemispheric lateralization of acoustic and linguistic processes. J Cogn Neurosci. 24:636–652.

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage. 13:684–701.

Mourao-Miranda J, Bokde ALW, Born C, Hampel H, Stetter M. 2005. Classifying brain states and determining the discriminating

activation patterns: support vector machine on functional MRI data. Neuroimage. 28:980–995.

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

Nichols T, Brett M, Andersson J, Wager T, Poline JB. 2005. Valid conjunction inference with the minimum statistic. Neuroimage. 25:653–660.

Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. 2006. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp. 27:562–571.

Obleser J, Wise RJS, Dresner MA, Scott SK. 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. J Neurosci. 27:2283–2289.

Okada K, Hickok G. 2006. Identification of lexical-phonological networks in the superior temporal sulcus using functional magnetic resonance imaging. Neuroreport. 17:1293–1296.

Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G. 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb Cortex. 20:2486–2495.

Patterson K, Nestor PJ, Rogers TT. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat Rev Neurosci. 8:976–987.

Peelle JE, Johnsrude I, Davis MH. 2010. Hierarchical processing for speech in human auditory cortex and beyond. Front Hum Neurosci. 4:1–3.

Pereira F, Botvinick M. 2011. Information mapping with pattern classifiers: a comparative study. Neuroimage. 56:476–496.

Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage. 45:S199–S209.

Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M. 2004. Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature. 427:448–451.

Rauschecker JP. 1998. Cortical processing of complex sounds. Curr Opin Neurobiol. 8:516–521.

Rauschecker JP, Scott SK. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 12:718–724.

Rosen S, Wise RJS, Chadha S, Conway EJ, Scott SK. 2011. Hemispheric asymmetries in speech perception: sense, nonsense and modulations. PLoS One. 6:e24672.

Scott SK, Blank CC, Rosen S, Wise RJS. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 123:2400–2406.

Scott SK, Rosen S, Lang H, Wise RJS. 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech—a positron emission tomography study. J Acoust Soc Am. 120:1075–1083.

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

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJS. 2006. Converging language streams in the human temporal lobe. J Neurosci. 26:7328–7336.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. 2001. Functional specialization in rhesus monkey auditory cortex. Science. 292:290–293.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15:273–289.

Vaden KI, Muftuler LT, Hickok G. 2010. Phonological repetition-suppression in bilateral superior temporal sulci. Neuroimage. 49:1018–1023.

von Kriegstein K, Eger E, Kleinschmidt A, Giraud AL. 2003. Modulation of neural responses to speech by directing attention to voices or verbal content. Cogn Brain Res. 17:48–55.

Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP. 2001. Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J Cogn Neurosci. 13:1–7.