Cerebral Cortex October 2010;20:2486–2495
doi:10.1093/cercor/bhp318
Advance Access publication January 25, 2010

Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech

Kayoko Okada[1], Feng Rong[1], Jon Venezia[1], William Matchin[1], I-Hui Hsieh[2], Kourosh Saberi[1], John T. Serences[3] and Gregory Hickok[1]

[1]Center for Cognitive Neuroscience and Department of Cognitive Sciences, University of California, Irvine, CA 92697, USA, [2]Institute of Cognitive Neuroscience, National Central University, Jhongli City, 32001 Taiwan and [3]Department of Psychology, University of California, San Diego, La Jolla, CA 92093, USA

Address correspondence to Gregory Hickok. Email: greg.hickok@gmail.com.

Hierarchical organization of human auditory cortex has been inferred from functional imaging observations that core regions respond to simple stimuli (tones) whereas downstream regions are selectively responsive to more complex stimuli (band-pass noise, speech). It is assumed that core regions code low-level features, which are combined at higher levels in the auditory system to yield more abstract neural codes. However, this hypothesis has not been critically evaluated in the auditory domain. We assessed sensitivity to acoustic variation within intelligible versus unintelligible speech using functional magnetic resonance imaging and a multivariate pattern analysis. Core auditory regions on the dorsal plane of the superior temporal gyrus exhibited high levels of sensitivity to acoustic features, whereas downstream auditory regions in both anterior superior temporal sulcus and posterior superior temporal sulcus (pSTS) bilaterally showed greater sensitivity to whether speech was intelligible or not and less sensitivity to acoustic variation (acoustic invariance). Acoustic invariance was most pronounced in more pSTS regions of both hemispheres, which we argue support phonological level representations. This finding provides direct evidence for a hierarchical organization of human auditory cortex and clarifies the cortical pathways supporting the processing of intelligible speech.

Keywords: auditory cortex, fMRI, Heschl's gyrus, hierarchical organization, intelligible speech, language, multivariate pattern classification, speech, superior temporal sulcus

Introduction

Hierarchical organization appears to be a fundamental feature of cortical sensory systems. For example, in the primate cortical visual system, early processing stages in primary visual cortex V1 respond to simple visual features such as local contours within a relatively small receptive field, whereas downstream areas code for progressively more complex features such that neurons in inferotemporal cortex respond to complex objects within a relatively large receptive field (Logothetis and Sheinberg 1996; Tanaka 1996; Rolls 2000; Rousselet et al. 2004). A similar, although less thoroughly investigated, hierarchical organization has been reported in the primate auditory system (Kaas and Hackett 2000). For example, the auditory core region receives thalamic inputs from the ventral division of the medial geniculate (MGv), whereas the belt regions receive inputs from the dorsal division (MGd) and the auditory core region (see Kaas and Hackett 2000 for a review). Physiologically, cells in the auditory core region of macaque monkeys respond well and with short latencies to pure tones and have narrow frequency response curves, whereas in the belt regions, cells respond with longer latencies, are less finely tuned, and can be more responsive to spectrally complex stimuli (e.g., band-pass noise; Rauschecker et al. 1995; Recanzone, Guard, and Phan 2000; Recanzone, Guard, Phan, et al. 2000; Rauschecker and Tian 2004; Kajikawa et al. 2005; Kusmierek and Rauschecker 2009).

Functional imaging studies of human auditory cortex have suggested an analogous hierarchical organization: relatively simple stimuli such as pure tones are sufficient to drive activity in the auditory core (roughly, Heschl's gyrus [HG]), whereas more complex stimuli such as band-pass noise or speech are needed to produce maximal activation in surrounding auditory-responsive cortical regions (assumed to correspond to belt and parabelt fields; Binder et al. 2000; Scott et al. 2000; Wessinger et al. 2001; Hickok and Poeppel 2007). Moreover, some have argued that the observation of acoustic invariance in belt and parabelt fields, particularly in response to speech, implicates these regions in the maintenance of high-level perceptual representations (Rauschecker and Scott 2009). For example, one positron emission tomography study compared the response elicited by acoustically different examples of intelligible speech relative to acoustically "matched" unintelligible control stimuli in an attempt to isolate auditory responsive fields that responded equally well to high-level features of speech stimuli despite acoustic variance (Scott et al. 2000). A region in the left anterior superior temporal sulcus (aSTS) was identified with these properties (a subsequent functional magnetic resonance imaging [fMRI] study also found a posterior superior temporal sulcus (pSTS) activation [Narain et al. 2003]). However, these studies failed to demonstrate that auditory core regions were sensitive to the acoustic features in the stimuli as no activation differences were reported in core auditory areas between acoustically different intelligible speech stimuli. This is important because in order to support a claim of acoustic invariance in presumed downstream processing regions, one has to first demonstrate acoustic "variance" in the neural response in upstream processing levels. The failure to identify sensitivity to acoustic manipulations in auditory core regions in previous studies likely resulted from the use of a standard analysis approach in which activation from relatively large (>1 cm) swaths of neural tissue are averaged across subjects. This method may obscure distinct patterns of functional activity within a given region and may even have led to the appearance of acoustic invariance despite underlying sensitivity to acoustic features on a finer-grained level of analysis.

The present fMRI study assessed cortical response patterns to intelligible and unintelligible speech stimuli with varying acoustic features using a multivariate pattern analysis (MVPA) that is sensitive to the pattern of activity within a region of interest (ROI) in individual subjects rather than to the average amplitude of the response within a region across subjects (Haxby et al. 2001; Kamitani and Tong 2005; Norman et al. 2006). We used 2 intelligible but acoustically different types of

stimuli (clear speech and noise-vocoded speech) and 2 unintelligible types of stimuli (spectrally rotated versions of the clear and vocoded speech; Fig. 1, see supplemental materials for audio samples). Using a standard group-based subtraction analysis, we expected to replicate previous findings: no difference between stimulus types in core auditory regions and an "intelligibility effect" (intelligible > unintelligible) in left STS. Using MVPA, we expected 1) core auditory areas to exhibit different patterns of activation to each of the 4 stimulus types reflecting sensitivity to low-level acoustic differences and 2) downstream regions in the left STS to be invariant in response to acoustic differences between types of intelligible speech.

Materials and Methods

Subjects

Twenty (6 females) right-handed native English speakers between 18 and 47 years of age participated in the study. All volunteers had normal or corrected-to-normal vision, no known history of neurological disease, and no other contraindications for MRI. Informed consent was obtained from each participant in accordance with UCI Institutional Review Board guidelines.

Stimuli and Procedure

Participants were presented with 4 different types of auditory stimuli previously used to identify speech selective regions (Scott et al. 2000). The 4 stimuli were 1) clear speech sentences, 2) noise-vocoded speech (NV), 3) spectrally rotated speech (rot), 4) rotated noise-vocoded speech (rotNV). The first 2 types are intelligible to the perceiver, whereas the last 2 are unintelligible without extensive training. The clear sentences (Sp) were short sentences taken from BKB sentence list (Bench et al. 1979). To create intelligible speech that is acoustically dissimilar to clear speech, these sentences were passed through a channel vocoder to create noise-vocoded speech as described by Shannon et al. (1995). Noise-vocoded speech sounds like a harsh whisper and is intelligible to the perceiver with some training but lacks pitch saliency associated with speech. Unintelligible speech conditions were created by spectrally rotating speech around 2 kHz as described by Blesser (1972). Rotation of the signal preserves spectrotemporal complexity of speech and is acoustically similar to speech but renders it unintelligible without extensive training. After clear speech was rotated to create rotated clear speech, these were noise vocoded to create rotated noise-vocoded speech that is acoustically similar to noise-vocoded speech and is unintelligible to the perceiver.

Seventy sentences were digitally recorded at a sampling rate of 44.1 kHz. Noise-vocoded (NV) speech was created by band-pass filtering each sentence into 6 bands from 70 to 4000 Hz using a 512-point finite impulse response (Hamming windowed) digital filter. The width of each band was selected to approximate equal and constant distances along the basilar membrane using equation (1) from (Greenwood 1990) with cutoff frequencies at 70, 213, 446, 882, 1431, 2413, and 4000 Hz. The temporal envelope of each band was then extracted using the Hilbert transform, lowpass filtered, and multiplied by Gaussian noise

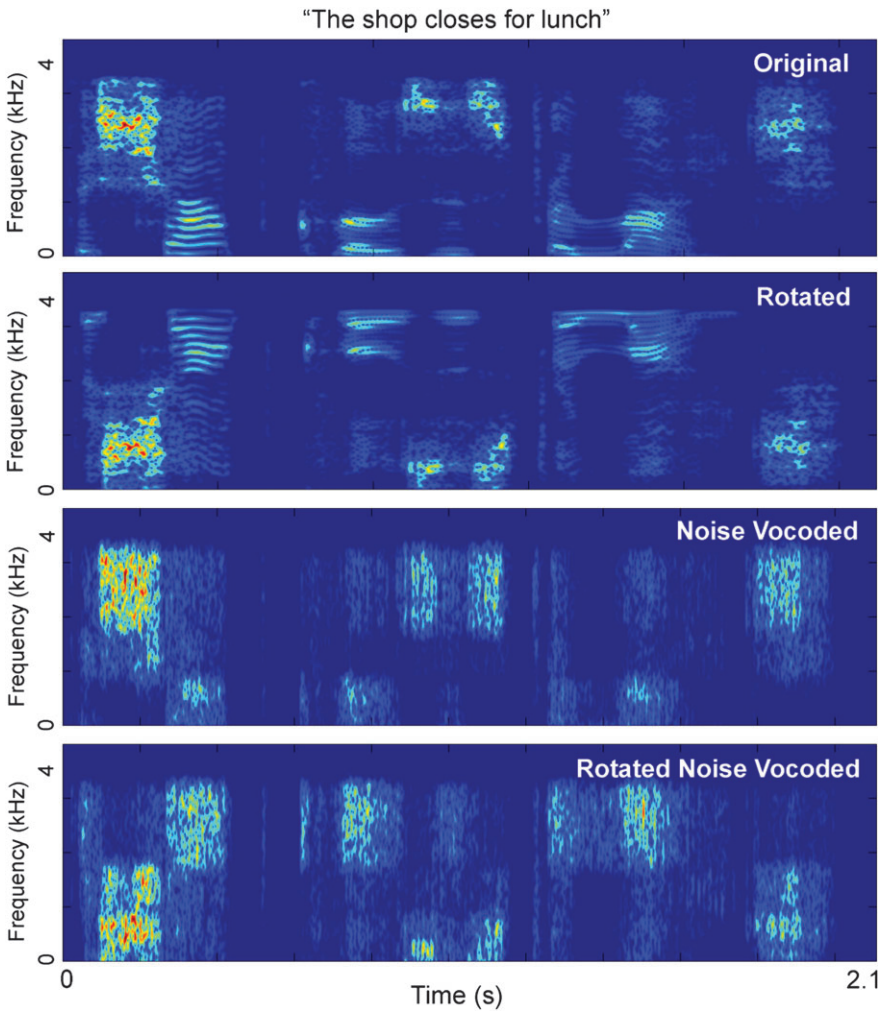

Figure 1. Spectrograms for a sample set of stimuli based on the sentence, *the shop closes for lunch*.

with the same bandwidth as each original filtered band. The resulting waveforms were summed across bands to generate the 6-channel noise-vocoded speech. All stimuli were then lowpass filtered at 3.8 kHz prior to spectral rotation (Scott et al. 2000) and lowpass filtered again at 3.8 kHz after rotation to maintain bandwidth symmetry about the 2-kHz rotation frequency. Stimuli were spectrally rotated by multiplying each waveform with a 4-kHz pure tone in the time domain, which results in convolution of their spectra in the frequency domain and, thus, spectral rotation around the 2-kHz axis. Following (Blesser 1972), we rescaled the original rotated speech spectrum using the weighting function where f is frequency in Hz, making the long-term amplitude spectra of the rotated sentences similar to those of the original sentences (W linearly decreases from +20 to −20 dB as frequency increases from 0 to 4 kHz). All stimuli were normalized to equal root-mean-square amplitude.

A single trial was 13.5 s in length created by concatenating 6 sentences of a single type (e.g., all clear, all rot, all NV, or all rotNV) that comprised 12 s of each trial followed by 1.5 s of silence during which subjects responded. The experiment started with a short practice session, and participants were exposed to 2 trials from each condition. On each trial, participants were asked to indicate with a button press if the sentences they heard were intelligible or unintelligible. A total of 10 sessions followed the practice scan. Four trials of each type were randomly presented in each session along with 4 rest trials (scanner noise) for a total of 40 trials per condition in the experiment. Following the 10 speech sessions, the study ended with an A1 localizer scan which consisted of 10 cycles of amplitude-modulated broadband noise (8 Hz with a modulation depth of 70%) alternating with rest (scanner noise) in 13.5-s intervals. All stimuli were presented over MR compatible headset and stimulus delivery, and timing were controlled using Cogent software (http://www.vislab.ucl.ac.uk/cogent_2000.php) implemented in Matlab 6 (Mathworks, Inc., Natick, MA).

Prior to scanning, subjects were exposed to each type of stimulus, and they were pretrained on the NV sentences. Participants listened to NV sentences one at a time and were asked to repeat the sentence. If the subject could not repeat the sentences, he/she was played the corresponding clear sentence. This procedure was repeated until participants could correctly repeat 15 consecutive sentences in a row. None of the sentences used in the pretraining was used in the experiment.

Scanning Parameters

MR images were obtained in a Philips Achieva 3T (Philips Medical Systems, Andover, MA) fitted with an 8-channel RF receiver head coil, at the John Tu and Thomas Yuen Center for Functional Onco-Imaging facility at the University of California, Irvine. We collected a total of 1144 echo planar imaging (EPI) volumes over 11 sessions using Fast Echo EPI (sense reduction factor = 2.4, matrix = 112 × 112 mm, time repetition [TR] = 2.7 s, time echo [TE] = 25 ms, size = 1.95 × 1.95 × 2 mm, flip angle = 70, number of slices = 47). After the functional scans, a high-resolution anatomical image was acquired with an magnetization prepared rapid acquisition gradient echo pulse sequence in axial plane (matrix = 256 × 256 mm, TR = 8 ms, TE = 3.7 ms, size = 1 × 1 × 1 mm).

Data Analysis

Preprocessing of the data and ROI identification were performed using AFNI software (http://afni.nimh.nih.gov/afni). In each session, the first 2 volumes and last volume of the series were discarded. Motion correction was performed by creating a mean image from all the volumes in the experiment and then realigning all volumes to that mean image using a 6-parameter rigid-body model (Cox and Jesmanowicz 1999). Images were then smoothed with an isotropic 6-mm full-width half-maximum (FWHM) Gaussian kernel (smoothed data were used for group analysis only). The anatomical image for each subject was coregistered to his/her mean EPI image.

Group Analysis

To investigate regions sensitive to intelligibility and acoustic similarity as demonstrated by blood oxygen level–dependent (BOLD) amplitude differences, and to compare our findings with previous results, group analysis was performed in addition to pattern analyses. After single-subject analysis, functional maps for each subject were transformed into standardized space to facilitate group analysis. The images were spatially normalized to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain (http://www.bic.mni.mcgill.ca/brainweb/) and resampled into 2-mm^3 voxels using nonlinear basis functions. Second-level analysis was performed on the linear contrasts of the parameter estimates from each participant, treating participants as a random effect, and voxelwise t-tests were performed. Statistical threshold in the group analysis was set at $P < 0.05$ using the false discovery rate correction.

Multivariate Pattern Analysis

Statistical analyses were performed in several parts. The first stage involved single-subject analysis and identifying ROIs in primary auditory cortex and several sites in STS bilaterally. Voxels from these ROIs (unsmoothed data) were used in the second part of the analysis which used MVPA implemented in Matlab. Group-level analysis was also performed to facilitate comparison of our results with previous neuroimaging studies.

ROI Identification

Regression analysis was performed, and ROIs were determined for each individual subject. The BOLD response was modeled by regressors created by convolving the predictor variables representing the time course of stimulus presentation with a standard hemodynamic response function (Boynton et al. 1996). For the speech sessions, 4 such regressors were used in estimation of the model corresponding to our 4 conditions: clear rot, NV, and rotNV. An additional 6 regressors corresponding to movement parameters determined during realignment stage of the process were entered into the model. For the A1 localizer session, one regressor associated with presentation of noise was entered into the model along with 6 motion regressors. An F statistic was calculated for each voxel, and statistical parametric maps (SPMs) were created for each subject.

In individual subjects, ROIs in auditory cortex were determined by taking the voxels significantly activated in the noise > rest contrast ($P < 0.001$) in the A1 localizer session (a more liberal threshold was used because this localizer was run in only one scan—note that because this was an independent localizer, a more liberal threshold cannot systematically bias the results). One subject did not yield significant activation in auditory cortex and was omitted from MVPA analysis (both hemispheres, $N = 18$). To identify ROIs along STS, we first identified regions sensitive to intelligible speech defined by the contrast of clear > rot ($P < 0.0001$). We chose this contrast because the amplitude response curves for these 2 conditions in A1 were virtually identical and therefore one might suppose were better matched acoustically. For the speech sessions, ROIs were determined using only the odd-numbered sessions, and MVPA was then performed on voxels extracted from the even-numbered sessions. Using this split-plot approach ensures that voxel selection procedure and subsequent analyses are independent. ROI identification in STS was achieved both functionally and anatomically. Anatomically, we partitioned STS into aSTS, middle superior temporal sulcus (mSTS), and pSTS sectors defined relative to HG as visualized on each subject's own brain: aSTS was defined as anterior to the anterior most extent of HG, mSTS was defined as regions that fell between the anterior and posterior most extent of HG, and pSTS was defined as regions that fell posterior to the posterior most extent of HG. We then identified peaks of activity along the STS that fell within these anatomically defined sectors in both the left and right hemispheres. To count as a peak in a given sector, there had to be a local maximum within that sector; that is, if an activation within the boundaries of say mSTS was simply on the shoulder of a peak in say aSTS, it was not counted as a peak in mSTS but rather as a peak in aSTS. Using this method, a majority of subjects had peaks in the left and right aSTS (19/19), left and right pSTS (18/19 in the left; 14/19 in the right), and in the right mSTS (18/19). Left hemisphere mSTS peaks were found in less than half (9/19) of the subjects and, therefore, were not analyzed further. Around each identified peak voxel in each sector in each subject, we formed

a 7 × 7 × 7 voxel cube in unsmoothed data from which ROI data were extracted for MVPA analysis (results did not change qualitatively when a 5 × 5 × 5 was used, and as most activated voxels were captured by the 7^3 cube, larger sized ROIs would not contribute useful information). Only significantly activated voxels (clear > rot) within the ROI were extracted; thus, the maximum number of voxels in a given ROI for a given subject was 7^3 voxels. The average ROI size was 140 voxels (range 95–168). Again, the data that served to define ROI peaks came from odd-numbered sessions and extracted ROI data for MVPA came from (independent) even-numbered sessions.

Pattern Classification

MVPA was implemented in 7 ROIs identified in individual subjects to explore the spatial distribution of activation to stimuli that vary in terms of intelligibility and acoustic similarity. All analyses described below were performed on even-numbered sessions (runs) only, that is, runs that were independent from the ROI selection runs. MVPA was achieved using a support vector machine (SVM) (MATLAB Bioinformatics Toolbox v3.1, The MathWorks, Inc., Natick, MA) as a pattern classification method. The logic behind this approach is that if an SVM is able to successfully classify one condition from another based on the pattern of response in an ROI, then the ROI must contain information that distinguishes the 2 conditions. In each ROI, 4 different pairwise classifications were performed: (i) clear versus rot, (ii) NV versus rotNV, (iii) clear versus NV, and (iv) rot versus rotNV. Note that (i) and (ii) involve classification of intelligible versus unintelligible speech and therefore should be distinguishable in brain regions that are sensitive to this distinction, whereas (iii) and (iv) involve classification "within" intelligible or unintelligible speech and therefore should not be discriminable in brain regions that are acoustically invariant; this is particularly true for classification (iii) that involves 2 intelligible but acoustically different stimuli.

Preprocessing procedures of the signals before applying SVM include normalization and averaging. First, we normalized the motion-corrected and spatially aligned fMRI time series in each recording session (run) by calculating voxel-based z scores. Second, the normalized data were averaged across the volumes within each trial. In addition, to ensure that overall amplitude differences between the conditions were not contributing to significant classification, the mean activation level across the voxels within each trial was removed prior to classification. We then performed SVM on the preprocessed data set using a leave-one-out cross validation approach (Vapnik 1995). In each iteration, we used data from all but one even session to train the SVM classifier and then used the SVM to classify the data from the remaining session. The SVM-estimated condition labels for the testing data set were then compared with the real labels to compute a classification accuracy score. Classification accuracy for each subject was derived by averaging the accuracy scores across all leave-one-out sessions, and an overall accuracy score was computed by averaging across subjects for each pairwise classification.

We then statistically evaluated the classification accuracy scores using nonparametric bootstrap methods (Lunneborg 2000). Similar classification procedures were repeated 10 000 times for each pairwise classification within each individual data set, the only difference from above method is that the condition labels in training data set for each leave-one-out session were randomly reshuffled per repetition. Therefore, we obtained a random distribution of the bootstrap classification accuracy scores that ranged from 0 to 1 for each subject and pairwise classification, where the ideal mean of this distribution is at the accuracy value of 0.5. We then tested the null hypotheses that the original classification accuracy score equals to the mean of the distribution by computing a one-tailed accumulated percentile of the original classification accuracy score in the distribution. If the accumulated $P > 0.95$, then we rejected the null hypotheses and concluded that for this subject, signal from the corresponding ROI can classify the 2 tested experimental conditions. Furthermore, a bootstrap-T approach was used to assess the significance of the classification accuracy on the group level. For each repetition of the bootstrap, a t-test of the accuracy scores across all subjects against the ideal accuracy score (0.5 in our case) was performed. The t-score from the original classification procedures across the subjects was then statistically tested against the mean value of the distributed bootstrap t-scores. Same as in the within-subject approach, an accumulated $P > 0.95$ guarantees rejection of the null hypotheses and our conclusion that for this pairwise classification, accuracy score from the corresponding ROI is significantly greater than chance.

Results

Subjects judged both clear speech and noise-vocoded speech as intelligible and accurately judged both rotated speech and rotated noise-vocoded speech as unintelligible with greater than 98% accuracy, indicating that subjects perceived the stimuli veridically while in the scanner.

Standard Analysis

In the standard subtraction group-based analysis, the contrast between intelligible and unintelligible speech (clear + NV) – (rot + rotNV) replicated and extended previous findings (Fig. 2 and Table 1). Consistent with previous reports, activation was largely in the lateral superior temporal cortex with no significant voxels on the supratemporal plane in the vicinity of HG or surrounding areas, indicating that early auditory cortical fields respond to all stimulus types equally. However, unlike previous studies, which report a predominantly left anterior temporal focus of activation, we found robust activation bilaterally in aSTS/superior temporal gyrus (STG) as well as posterior portions of STS/STG, also bilaterally (Fig. 2). Additional smaller foci of activation were found in the inferior temporal gyrus (right), fusiform gyrus (bilateral), parahippocampal gyrus (left), inferior and middle frontal gyri (left), and cerebellum (right). The failure of previous studies to find bilateral activation may be because those studies were underpowered (N range = 7–11 subjects). We also characterized group-level activity associated with our auditory localizer scan (8-Hz amplitude modulated noise compared with background scanner noise) that revealed activation on the supratemporal plane including HG and immediately surrounding tissue (Fig. 3*A*). This activation likely includes the auditory core as well as immediately surrounding belt regions.

Multivariate Pattern Analysis

For MVPA, ROIs were identified in individual subject data. Two ROIs, one in each hemisphere, were identified in and around HG on the supratemporal plane using the auditory localizer scan (see Experimental Procedures). ROIs in the STS were

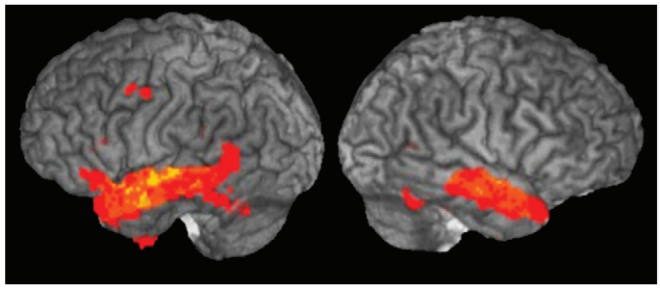

Figure 2. Group results from a standard BOLD amplitude subtraction analysis (P = 0.05, false discovery rate corrected) projected onto a surface-rendered template brain showing regions that respond more to intelligible speech (clear + NV) than unintelligible speech (rot + rotNV).

identified using the contrast, clear–rot (a standard BOLD subtraction analysis of clear vs. rot conditions resulted in a virtually identical activation map compared with the intelligible–unintelligible contrast, see Figure S1). This contrast was chosen for ROI identification because it produced the most similar average activation amplitude in the HG ROI, providing prima facie evidence for these 2 conditions being the most closely matched acoustically as has been claimed previously (Scott et al. 2000; Fig. 3*B*). Note that this contrast should identify regions that are most selective for speech intelligibility (because clear and rotated speech differ in this respect) and least affected by acoustic features (because clear and rotated speech produced identical average activation in core auditory areas and are argued to be well-matched acoustically). That is, this contrast will identify "candidate" acoustic invariant regions that we can further assess using a more fine-grained analysis (again, ROI identification and MVPA used independent data). Using this clear–rot contrast, 2 ROIs were identified in the left hemisphere, aSTS and pSTS, and 3 ROIs were indentified in the right hemisphere, aSTS, mSTS, and pSTS (see Experimental Procedures for details on anatomical criterion for defining ROIs) in a majority of subjects. The locations of the peak voxel in each subject's STS ROIs are plotted on a standardized brain image in Figure 4, the Talairach coordinates of these ROIs are provided in Table 2, and the average time course in these STS ROIs across subjects is presented in Figure S2. The pattern of activity in each ROI was then assessed in its ability to classify 4 different pairs of speech conditions: (i) clear versus rot, (ii) NV versus rotNV, (iii) clear versus NV noise-vocoded speech (NV), and (iv) rot versus rotNV. We will refer to these as "classification contrasts." Note that (i) and (ii) involve classification of intelligible versus unintelligible speech (intelligibility contrasts) and therefore

Table 1

Talairach coordinates of the peak voxels in activated cluster (thresholded, minimum 5 voxels; group analysis: intelligible > unintelligible, false discovery rate = 0.05)

Region	Approximate Brodmann area	Peak *x*	Peak *y*	Peak *z*
Left hemisphere				
Middle temporal gyrus	BA 21	−62	0	−6
Fusiform gyrus	BA 37	−54	−56	−18
Fusiform gyrus/parahippocampal gyrus		−38	−38	−22
Inferior parietal lobe	BA 40	−58	−34	24
Parahippocampal gyrus		−10	−34	−4
Inferior frontal gyrus	BA 45	−54	22	20
Middle frontal gyrus	BA 6	−50	6	50
Right hemisphere				
Middle temporal gyrus	BA 21	64	−6	−8
Medial temporal lobe	BA 38	22	10	−28
Cerebellum		22	−74	−42
Fusiform gyrus	BA 37	56	−52	−18
Inferior temporal gyrus	BA 20	42	−2	−40
Cerebellum		34	−36	−26
Cerebellum		26	−26	−24

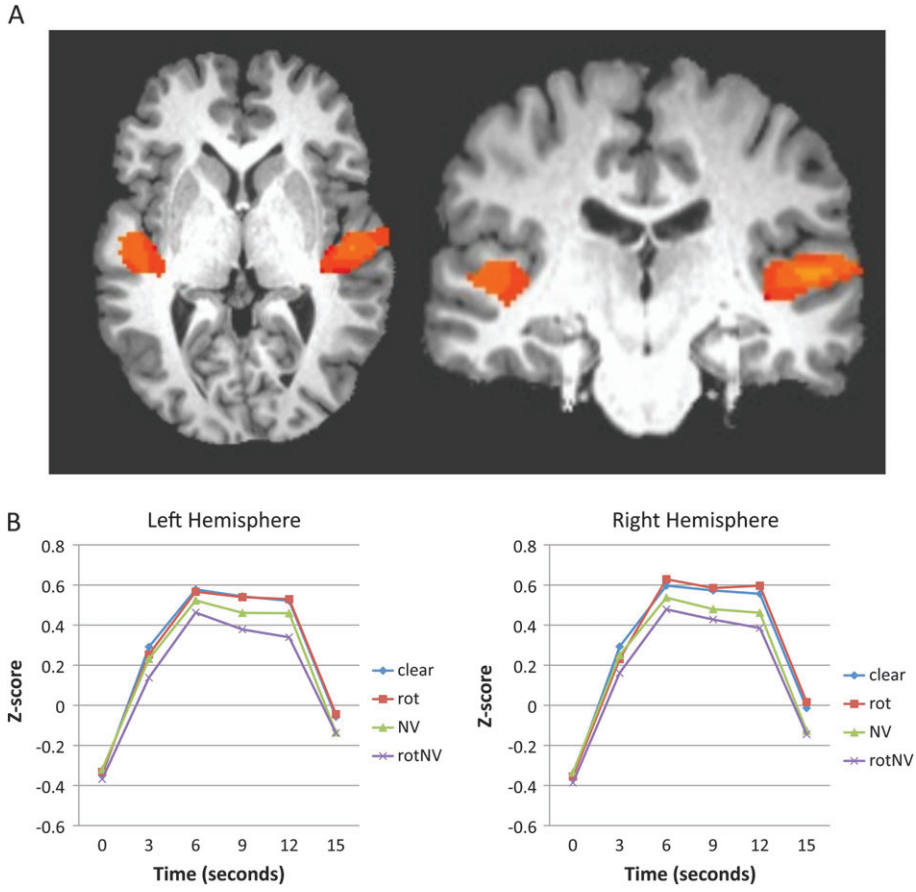

Figure 3. Activation of auditory core regions (HG) to 8-Hz amplitude-modulated wideband noise versus rest (scanner noise), $P < 0.001$, uncorrected. (*A*) Axial and coronal slices showing activation in HG and immediately surrounding tissue. Images are in radiological convention (left = right). (*B*) Average signal time course for each speech condition for the HG ROI in each hemisphere. Note the virtually identical average response to clear and rot speech.

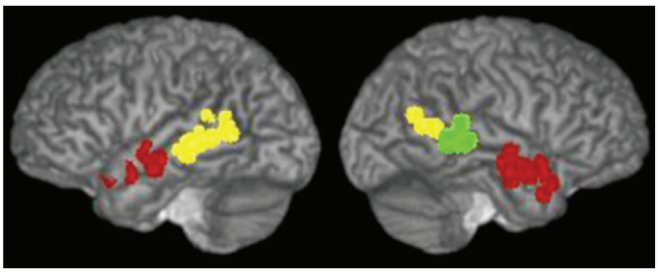

Figure 4. Location of the peak voxel in STS ROIs for each subject projected onto a surface-rendered template brain. ROIs were identified using the contrast, clear minus rot with a threshold of $P < 0.0001$, uncorrected. Red, aSTS; green, mSTS; yellow, pSTS. Note the nonoverlapping distributions of these ROIs across subjects.

Table 2
Talairach coordinates [*x*, *y*, *z*] for the peak voxel in each STS ROI in each subject

	Left hemisphere		Right hemisphere		
	aSTS	pSTS	aSTS	mSTS	pSTS
S1	[−66 6 −18]	[−70 −40 4]	[62 0 −10]	[58 −18 −2]	[44 −46 6]
S2	[−54 18 −22]	[−60 −50 16]	[64 −6 −8]	[66 −24 6]	[62 −56 12]
S3	[−64 −2 −4]	[−70 −30 0]	[56 4 −16]	[48 −16 −12]	
S4	[−62 6 −10]	[−52 −42 6]	[62 −4 −12]	[64 −18 −2]	[54 −36 8]
S5	[−50 22 −20]	[−66 −52 12]	[50 16 −20]	[56 −30 2]	[50 −60 18]
S6	[−66 2 −6]	[−66 −36 8]	[66 −6 −2]	[70 −22 −2]	[48 −40 2]
S7	[−64 10 −14]		[62 10 −10]	[66 −24 2]	[60 −30 4]
S8	[−60 4 −14]	[−58 −28 2]	[58 −10 −8]	[68 −22 −2]	[62 −50 12]
S9	[−74 −6 −10]	[−72 −44 2]	[60 6 −12]	[50 −22 −10]	[48 −42 2]
S10	[−58 14 −20]	[−68 −30 4]	[58 −2 −10]		
S11	[−60 4 −14]	[−64 −28 −4]	[54 16 −22]	[50 −22 −6]	
S12	[−66 −4 2]	[−64 −40 4]	[60 −6 −8]	[68 −26 −2]	[46 −42 6]
S13	[−60 −10 −12]	[−66 −48 8]	[52 −6 −18]	[50 −26 0]	[64 −48 10]
S14	[−72 4 −10]	[−66 −24 −2]	[72 −8 −16]		
S15	[−60 6 −10]	[−66 −40 18]	[62 −6 −4]	[66 −26 0]	[58 −34 14]
S16	[−56 −12 −6]	[−70 −40 2]	[62 10 −8]	[54 −26 −2]	
S17	[−60 14 −18]	[−70 −28 4]	[50 12 −26]	[50 −6 −16]	[70 −32 2]
S18	[−60 6 −18]	[−54 −34 2]	[60 14 −16]	[70 −8 −8]	[52 −36 2]
S19	[−60 −8 −8]	[−60 −52 8]	[62 −8 −10]	[52 −24 −6]	[64 −40 4]

should be discriminable in brain regions that are sensitive to this distinction, whereas (iii) and (iv) involve classification within intelligible/unintelligible speech conditions (acoustic contrasts) and therefore should not be discriminable in brain regions that are acoustically invariant.

The results of the classification analyses for each contrast in each ROI are presented in Figure 5. Several results are noteworthy. First, the activation pattern within the HG ROI in each hemisphere was successful at classifying each of the 4 contrasts. Thus, despite the statistically identical responses in this region to each of the 4 conditions in the BOLD amplitude signal (in particular in the clear vs. rot conditions, see Fig. 3*B*), the HG ROI is nonetheless highly sensitive to the acoustic differences between these classes of stimuli, as we expected. Second, although all STS ROIs successfully classified the intelligibility contrasts (left 2 bars in each graph), they showed varying degrees of classification ability in the 2 acoustic contrasts, that is, classification contrasts that differ acoustically but not in terms of intelligibility (right 2 bars in each graph). In the left hemisphere, the aSTS ROI successfully classified the 2 intelligible conditions (clear vs. NV; i.e., acoustic invariance did not hold in this ROI for this contrast) but did not classify the 2 unintelligible conditions (rot vs. rotNV). The pSTS ROI classified neither of the acoustic contrasts (i.e., acoustic invariance held in this ROI). In other words, the aSTS exhibited some degree of sensitivity to acoustic differences between intelligible speech conditions, whereas the pSTS was only successful at classifying intelligible from unintelligible conditions and was not sensitive to acoustic differences. A different pattern was found in the right hemisphere, with aSTS failing to classify the acoustic contrasts and the mSTS and pSTS classifying all contrasts significantly, similar to the HG ROI, but with less impressive classification accuracy values for the acoustic contrasts.

Two things are apparent from inspection of the classification accuracy graphs. One is that several of the classification accuracy values hover around the statistical threshold for significance. Another is that the ROIs not only differ in their ability to significantly classify the various contrasts but also in the magnitude of the effects. For example, consider the patterns in the right hemisphere ROIs. Looking only at the pattern of significant classifications, the HG ROI is identical to mSTS and different from aSTS. However, when one takes the magnitude of classification accuracy into consideration, it appears that mSTS is showing the largest "intelligibility effect" with a greater difference between the intelligibility contrasts (left 2 bars) compared with the acoustic contrasts (right 2 bars) and appears to have a very different pattern compared with the HG ROI.

To avoid drawing conclusions based solely on thresholds for statistical significance, and also to capture the patterns found in the magnitude of classification accuracies across ROIs, we implemented a second criterion for assessing functional hierarchies in auditory cortex. Using all 4 classification contrasts, we calculated an "acoustic invariance index" for each ROI in each subject. The reasoning behind this score is that a region that is coding a higher level feature of a stimulus should exhibit maximal sensitivity to that feature and minimal sensitivity to lower level features. In the present context, such an effect would manifest as a large intelligibility effect (a large classification accuracy difference for intelligible versus unintelligible conditions) and a minimal acoustic effect (small classification accuracy differences for acoustic classification contrasts). We quantified this first by taking the sum of the 2 intelligibility classification contrasts (clear vs. rot + NV vs. rotNV) and subtracting the 2 acoustic classification contrasts (clear vs. NV + rot vs. rotNV). This is a metric of the size of the effect of intelligibility. We then corrected this value with the size of the acoustic effect by subtracting the sum of the absolute values of the acoustic effects (|clear vs. NV – NV vs. rotNV| + |clear vs. rot – NV vs. rotNV|). Thus,

$$\text{acoustic invariance index} = (IC_1 + IC_2) - (AC_1 + AC_2) - |(IC_1 - IC_2)| + |(AC_1 - AC_2)|, \quad (1)$$

where IC is intelligibility contrast and AC is acoustic contrast. In the limit, this index can range from -1 to 1 where acoustically invariant responses are indicated by positive values and greater sensitivity to acoustic features relative to the intelligibility manipulation is indicated by negative values. Thus, this index is a measure of the size of the intelligibility effect relative to the size of the acoustic effect.

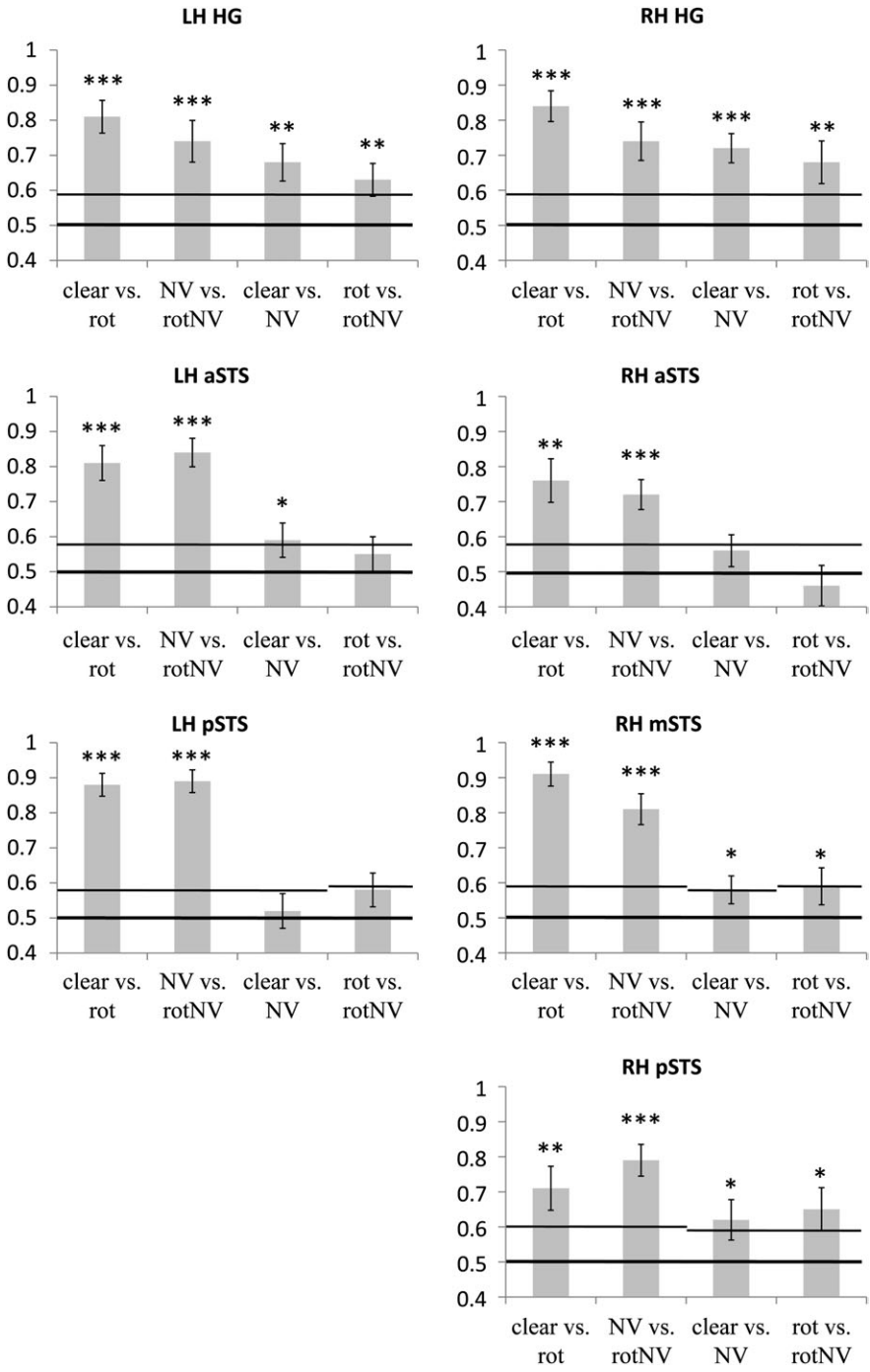

Figure 5. Classification accuracy (proportion correct) for the 4 classification contrasts in the 3 left hemisphere ROIs and 4 right hemisphere ROIs. The left 2 bars in each graph are "intelligibility" contrasts (intelligible vs. unintelligible) and the right 2 bars in each graph are acoustic contrasts (acoustically different intelligible vs. intelligible and acoustically different unintelligible vs. unintelligible). Thick black horizontal line indicates chance (0.5) and thin black line marks the upper bound of the 95% confidence interval determined via a bootstrapping procedure. As the bootstrapping procedure was calculated separately for each classification contrast, the 95% confidence interval boundary can vary from one condition to the next. LH, left hemisphere; RH, right hemisphere; clear, clear speech; rot, rotated speech; NV, noise-vocoded speech; rotNV, rotated noise-vocoded speech. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

Figure 6 shows the mean acoustic invariance values for each ROI in the left and right hemispheres. In the left hemisphere, a clear hierarchical progression is evident with the HG ROI showing the least acoustic invariance and pSTS with the most; aSTS falls in between. A similar pattern holds in the right hemisphere except the top of the hierarchy is the mSTS rather than the pSTS, the latter behaving much more like the HG ROI. Wilcoxon matched paired tests (2 tailed) revealed that the acoustic invariance index was significantly higher for pSTS compared with the HG ROI in the left hemisphere ($Z = 2.87$, $P = 0.004$) and for mSTS compared with the HG ROI in the right hemisphere ($Z = 2.04$, $P = 0.04$). No other comparisons reached

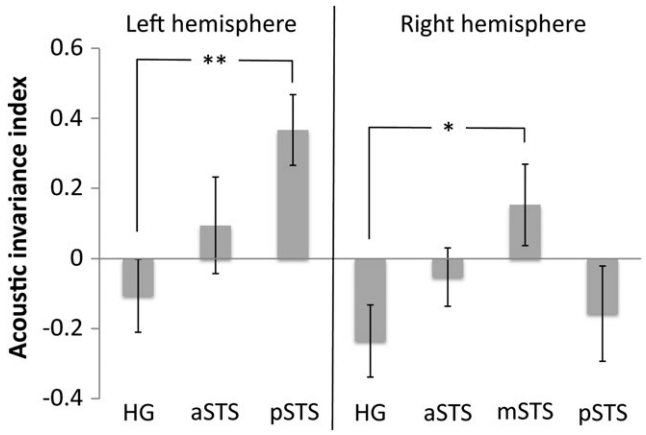

Figure 6. Acoustic invariance across the 7 ROIs as measured using an acoustic invariance index (see text). This index effectively ranges from −1 to 1 where positive values indicate higher degrees of acoustic invariance and negative values indicate lower degrees of acoustic invariance (i.e., more acoustic sensitivity). $*P < 0.05$, $**P < 0.01$, 2 tailed.

significance, although all *P* values were less than 0.2 (2 tailed) suggestive of the hierarchical trend seen in the pattern of means across ROIs (see Table 3 for complete Wilcoxon test results).

The results of the classification accuracy thresholding analysis (Fig. 5) and the acoustic invariance index analysis (Fig. 6) converge in the left hemisphere but lead to somewhat different conclusions for the right hemisphere ROIs. On the thresholding analysis, the right aSTS appears to be the ROI that is acoustically invariant, whereas in the acoustic invariance index analysis, the mSTS is the most acoustically invariant ROI. The reason for this discrepancy is that the acoustic invariance index analysis avoids thresholding artifacts and takes into consideration the magnitude of the intelligibility effect relative to the magnitude of the acoustic effect. We therefore place more weight on the acoustic invariance index analysis.

Discussion

The experiment reported here provides direct evidence for a hierarchical organization of human auditory cortex by demonstrating an increasing degree of acoustic invariance in the response to speech stimuli across several cortical regions. Multivariate pattern analysis showed that core auditory regions on the dorsal plane of the STG are sensitive to acoustic variation both within and across speech intelligibility categories, whereas downstream regions in the STS are both more sensitive to the higher order intelligibility features of speech stimuli and less sensitive to acoustic variation within intelligibility categories. Specifically, we found that auditory core regions were sensitive to (i.e., successfully classified) acoustic differences between clear speech, rotated speech, noise-vocoded speech, and rotated noise-vocoded speech. Regions in the anterior and posterior sectors of the STS bilaterally were sensitive to the differences between intelligible and unintelligible speech stimuli but less sensitive to acoustic differences within intelligibility categories. The pSTS in the left hemisphere and the mSTS in right hemisphere (see below) showed the highest degree of acoustic invariance, whereas the aSTS showed a degree of sensitivity to acoustic variation that was intermediate between core auditory areas (HG) and the pSTS/mSTS.

Table 3
Results of the Wilcoxon matched paired tests for the acoustic invariance index scores across 3 left hemisphere and 3 right hemisphere ROIs

	N	Z	*P*
Left hemisphere			
HG vs. aSTS	17	1.45	0.15
HG vs. pSTS	16	2.87	**0.004**
aSTS vs. pSTS	18	1.42	0.16
Right hemisphere			
HG vs. aSTS	17	1.31	0.19
HG vs. mSTS	15	2.04	**0.04**
aSTS vs. mSTS	17	1.41	0.16

Note: The right hemisphere pSTS ROI included only 14 subjects and so was excluded from this test.

The locations of our ROIs in the STS were defined anatomically relative to HG. An activation focus anterior to the anterior-most extent of HG was defined as aSTS, a focus posterior to the posterior-most extent of HG was defined as pSTS, and the region falling in between the anterior–posterior extent of HG was defined as mSTS. In the left hemisphere, consistent activation peaks were found only in aSTS and pSTS, whereas in the right hemisphere, consistent activation peaks were found in all 3 sectors of the STS. Functionally, however, the left pSTS and mSTS appear to be homologous and on the broader anatomical scale of the entire temporal lobe, the left pSTS ROIs and the right mSTS ROIs appear to be largely overlapping in the posterior half of the STS (Fig. 4). The most posterior activation in the right hemisphere exhibited acoustic sensitivity similar to that found in HG (Figs 5 and 6). This was an unexpected finding and one that requires further investigation to understand.

The hierarchical progression from HG to aSTS to pSTS in terms of acoustic invariance does not necessarily mean that cortical auditory pathways follow a serial route through these regions. For example, it is conceivable that anterior and posterior regions represent parallel auditory pathways. Such an organization has been proposed recently by Rauschecker and Scott (2009) although the present data are incompatible with the functions they ascribe to the 2 streams. These authors propose that an anterior processing stream supports recognition of auditory objects such as speech and a posterior stream supports spatial hearing and sensory–motor integration. The present finding that pSTS regions (bilaterally) exhibit the greatest degree of acoustic invariance when processing speech suggest instead that pSTS regions are critically involved in the recognition of auditory objects such as speech (a more dorsal region on the planum temporale supports sensory-motor integration; Hickok and Poeppel 2000, 2004, 2007; Hickok et al. 2003, 2009). A vast amount of lesion data also support this view (Bates et al. 2003; Dronkers et al. 2004; Hickok and Poeppel 2007; Dronkers and Baldo 2009), as does functional imaging evidence implicating the posterior half of the STS bilaterally in phonological level processing (Liebenthal et al. 2005; Okada and Hickok 2006; Hickok and Poeppel 2007; Vaden et al. 2010). The involvement of the pSTS in phonological processing likely explains the trend in the left pSTS ROI toward significant classification of rotated versus rotated noise-vocoded conditions in the present study. Although rotated speech is unintelligible, some amount of phonemic information is recoverable (Blesser 1972; Scott et al. 2000); this is not true of rotated noise-vocoded speech. The classification contrast

between these conditions, thus, compares a stimulus with partial phonemic information (rot) against a condition with effectively none (rotNV), which may have been the basis for the trend toward classification in the pSTS ROI. This is consistent with Scott et al. (2000) who report a posterior activation for the conditions that have some degree of phonemic information (clear, NV, rot) relative to the condition that does not (rotNV). Given this collection of observations, we suggest that the relative acoustic invariance of the pSTS revealed in the present study reflects this region's role in processing phonological level information.

It is unclear what functions might be supported by anterior temporal regions. Several studies have implicated the anterior temporal lobe in sentence-level processing (including syntactic functions; Mazoyer et al. 1993; Humphries et al. 2001, 2005, 2006; Vandenberghe et al. 2002; Friederici et al. 2003; Crinion and Price 2005) that certainly could drive the intelligibility effect, but the present finding that the aSTS is sensitive to acoustic features suggests, in addition, a lower level acoustic function. One candidate acoustic function comes from functional imaging studies that report that manipulation of prosodic/intonation information in speech stimuli modulates activity in the anterior superior temporal lobe (Buchanan et al. 2000; Humphries et al. 2005) and from a transcranial magnetic stimulation study showing that stimulation of the anterior temporal region (in the right hemisphere at least) can interfere with the perception of emotional prosody in speech (Hoekert et al. 2008). It is conceivable that acoustic differences in the prosodic cues present in clear, rotated, and noise-vocoded speech contributed to the acoustic effects in aSTS. Additional research is needed to fully understand the basis for the patterns of activation in the aSTS.

Previous studies using speech stimuli similar to those employed in the present experiment report left-dominant activation for intelligible versus unintelligible speech that is restricted, in some studies, to anterior temporal regions (Scott et al. 2000, 2006; Narain et al. 2003). In contrast, the present study found robust bilateral activation in the same contrast that included both anterior and posterior temporal regions. The discrepancy is likely due to previous studies being underpowered in terms of the number of subjects that were studied (Scott et al. 2000, $N = 8$; Narain et al. 2003, $N = 11$; Scott et al. 2006, $N = 7$). The finding that intelligible speech, relative to unintelligible speech, activates a large bilateral network along the length of the STS is an important result because the previous left-dominant and exclusively anterior activation (in some studies) have been used as critical evidence for a left-dominant, anterior pathway for speech recognition (Rauschecker and Scott 2009). The current study indicates that there is no longer an empirical basis for this claim and supports a different view, namely, that speech recognition is supported by a bilateral network that critically includes the pSTS (Hickok and Poeppel 2007).

Finally, the use of spectrally rotated stimuli as an acoustic control for speech in neuroimaging studies has gained popularity in recent years on the basis of claims that it provides an ideal control for acoustic complexity (Scott et al. 2000). This claim appears to be substantiated by the virtually identical average amplitude of the fMRI BOLD response to clear and rotated speech in the auditory core regions (Fig. 3*B*). However, clear and rotated speech proved to be highly discriminable (accuracy > 80%) using multivariate classification analysis in the same set of auditory core region voxels, showing that clear and rotated speech are far from acoustically comparable. More generally, this result undermines the logic of many studies that attempt to isolate auditory speech processing networks by comparing the hemodynamic response to speech relative to nonspeech acoustic controls. It has been suggested that such designs may not be sufficiently sensitive to the fine-grained organization of speech networks within the typically broad ROIs that comprise the theoretical focus of most functional imaging studies (Okada and Hickok 2006; Hickok and Poeppel 2007). The present finding validates this concern by showing that even in core auditory areas clear speech produces a very different pattern of response than acoustic "control" stimuli, even when the average response to these 2 types of stimuli are identical.

In sum, this report demonstrates the hierarchical organization of human auditory cortex by showing that auditory core regions are sensitive to acoustic features in speech and speech-like stimuli, whereas downstream regions in the STS are more sensitive to higher level features such as intelligibility and less sensitive to acoustic variation. The pSTS shows the highest degree of acoustic invariance likely reflecting its role in phonological processing in speech recognition.

Supplementary Material

Figures S1 and S2 and other supplementary materials can be found at http://www.cercor.oxfordjournals.org/.

Funding

National Institutes of Health (DC03681).

Notes

Conflict of Interest: None declared.

References

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF. 2003. Voxel-based lesion-symptom mapping. Nat Neurosci. 6:448–450.

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

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET. 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex. 10:512–528.

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

Boynton GM, Engel SA, Glover GH, Heeger DJ. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. J Neurosci. 16:4207–4221.

Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K, Jancke L. 2000. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Brain Res Cogn Brain Res. 9:227–238.

Cox RW, Jesmanowicz A. 1999. Real-time 3D image registration for functional MRI. Magn Reson Med. 42:1014–1018.

Crinion J, Price CJ. 2005. Right anterior superior temporal activation predicts auditory sentence comprehension following aphasic stroke. Brain. 128:2858–2871.

Dronkers N, Baldo J. 2009. Language: aphasia. In: Squire LR, editor. Encyclopedia of neuroscience. Oxford: Academic Press. p. 343–348.

Dronkers NF, Wilkins DP, Van Valin RD, Jr., Redfern BB, Jaeger JJ. 2004. Lesion analysis of the brain areas involved in language comprehension. Cognition. 92(1-2):145–177.

Friederici AD, Ruschemeyer SA, Hahne A, Fiebach CJ. 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb Cortex. 13:170–177.
Greenwood DD. 1990. A cochlear frequency-position function for several species—29 years later. J Acoust Soc Am. 87:2592–2605.
Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 293:2425–2430.
Hickok G, Buchsbaum B, Humphries C, Muftuler T. 2003. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci. 15:673–682.
Hickok G, Okada K, Serences JT. 2009. Area Spt in the human planum temporale supports sensory-motor integration for speech processing. J Neurophysiol. 101:2725–2732.
Hickok G, Poeppel D. 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci. 4:131–138.
Hickok G, Poeppel D. 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition. 92:67–99.
Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nat Rev Neurosci. 8:393–402.
Hoekert M, Bais L, Kahn RS, Aleman A. 2008. Time course of the involvement of the right anterior superior temporal gyrus and the right fronto-parietal operculum in emotional prosody perception. PLoS One. 3:e2244.
Humphries C, Binder JR, Medler DA, Liebenthal E. 2006. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. J Cogn Neurosci. 18:665–679.
Humphries C, Love T, Swinney D, Hickok G. 2005. Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Hum Brain Mapp. 26:128–138.
Humphries C, Willard K, Buchsbaum B, Hickok G. 2001. Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. Neuroreport. 12:1749–1752.
Kaas JH, Hackett TA. 2000. Subdivisions of auditory cortex and processing streams in primates. Proc Natl Acad Sci USA. 97: 11793–11799.
Kajikawa Y, de La Mothe L, Blumell S, Hackett TA. 2005. A comparison of neuron response properties in areas A1 and CM of the marmoset monkey auditory cortex: tones and broadband noise. J Neurophysiol. 93:22–34.
Kamitani Y, Tong F. 2005. Decoding the visual and subjective contents of the human brain. Nat Neurosci. 8:679–685.
Kusmierek P, Rauschecker JP. 2009. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. J Neurophysiol. 102:1606–1622.
Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. 2005. Neural substrates of phonemic perception. Cereb Cortex. 15:1621–1631.
Logothetis NK, Sheinberg DL. 1996. Visual object recognition. Annu Rev Neurosci. 19:577–621.
Lunneborg CE. 2000. Data analysis by resampling: concepts and applications. Pacific Grove (CA): Duxbury Press.
Mazoyer BM, Tzourio N, Frak V, Syrota A, Murayama N, Levrier O, Salamon G, Dehaene S, Cohen L, Mehler J. 1993. The cortical representation of speech. J Cogn Neurosci. 5:467–479.
Narain C, Scott SK, Wise RJ, 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.
Norman KA, Polyn SM, Detre GJ, Haxby JV. 2006. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 10:424–430.
Okada K, Hickok G. 2006. Identification of lexical-phonological networks in the superior temporal sulcus using fMRI. Neuroreport. 17:1293–1296.
Rauschecker JP, Scott SK. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 12:718–724.
Rauschecker JP, Tian B. 2004. Processing of band-passed noise in the lateral auditory belt cortex of the rhesus monkey. J Neurophysiol. 91:2578–2589.
Rauschecker JP, Tian B, Hauser M. 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. Science. 268:111–114.
Recanzone GH, Guard DC, Phan ML. 2000. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol. 83:2315–2331.
Recanzone GH, Guard DC, Phan ML, Su TK. 2000. Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. J Neurophysiol. 83:2723–2739.
Rolls ET. 2000. Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron. 27:205–218.
Rousselet GA, Thorpe SJ, Fabre-Thorpe M. 2004. How parallel is visual processing in the ventral pathway? Trends Cogn Sci. 8:363–370.
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 RJ. 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 F-G, Kamath V, Wygonski J, Ekelid M. 1995. Speech recognition with primarily temporal cues. Science. 270:303–304.
Tanaka K. 1996. Inferotemporal cortex and object vision. Annu Rev Neurosci. 19:109–139.
Vaden KI, Jr., Muftuler LT, Hickok G. 2010. Phonological repetition-suppression in bilateral superior temporal sulci. Neuroimage. 49:1018–1023.
Vandenberghe R, Nobre AC, Price CJ. 2002. The response of left temporal cortex to sentences. J Cogn Neurosci. 14:550–560.
Vapnik V. 1995. The nature of statistical learning theory. New York: Springer-Verlag.
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.