Behavioral/Cognitive

Distributed Neural Representations of Phonological Features during Speech Perception

Jessica S. Arsenault and [ID]Bradley R. Buchsbaum

Rotman Research Institute at Baycrest, Toronto, Ontario M6A 2E1, Canada, and Department of Psychology, University of Toronto, Toronto, Ontario M5S 3G3, Canada

A fundamental goal of the human auditory system is to map complex acoustic signals onto stable internal representations of the basic sound patterns of speech. Phonemes and the distinctive features that they comprise constitute the basic building blocks from which higher-level linguistic representations, such as words and sentences, are formed. Although the neural structures underlying phonemic representations have been well studied, there is considerable debate regarding frontal-motor cortical contributions to speech as well as the extent of lateralization of phonological representations within auditory cortex. Here we used functional magnetic resonance imaging (fMRI) and multivoxel pattern analysis to investigate the distributed patterns of activation that are associated with the categorical and perceptual similarity structure of 16 consonant exemplars in the English language used in Miller and Nicely's (1955) classic study of acoustic confusability. Participants performed an incidental task while listening to phonemes in the MRI scanner. Neural activity in bilateral anterior superior temporal gyrus and supratemporal plane was correlated with the first two components derived from a multidimensional scaling analysis of a behaviorally derived confusability matrix. We further showed that neural representations corresponding to the categorical features of voicing, manner of articulation, and place of articulation were widely distributed throughout bilateral primary, secondary, and association areas of the superior temporal cortex, but not motor cortex. Although classification of phonological features was generally bilateral, we found that multivariate pattern information was moderately stronger in the left compared with the right hemisphere for place but not for voicing or manner of articulation.

Key words: auditory cortex; fMRI; motor cortex; multivoxel pattern analysis; phonological features; speech perception

Introduction

The phonemes that constitute the basic sound structure of spoken language can be defined according to their particular combination of phonological features, such as voicing, manner, and place of articulation. Psychophysical investigations of speech perception have attempted to discover how these distinctive features are represented in "psychological space" by examining the patterns of confusability between individual phonemes (Miller and Nicely, 1955; Shepard, 1972). Such studies have shown that the underlying psychometric representation of English consonants is imperfectly related to a taxonomic structure based only on distinctive features. The psychological representation of speech must reflect an underlying neural code, and with recent developments in multivariate approaches to functional neuroimaging (Kriegeskorte et al., 2008; Brouwer and Heeger, 2009), we can now ask how the perceptual organization of speech emerges from neural patterns of activation.

Recent work with electrode recordings from human superior temporal gyrus (STG; Mesgarani et al., 2014) has made significant strides in this endeavor, showing that areas within the left STG are sensitive to specific phonological features. There is still debate, however, regarding the nature of such representations in regions outside of the left auditory cortex. Motor theories of speech perception have ranged from the strong view that the motor system is critical for speech perception (Liberman et al., 1967; Meister et al., 2007) to more nuanced positions that the motor system supports speech perception under some circumstances (D'Ausilio et al., 2009; Du et al., 2014). Functional neuroimaging studies have shown that areas in the frontal-motor speech system are often active during speech perception (Wilson et al., 2004; but see Szenkovits et al., 2012), although we still know little about the representational nature of such activation.

Some authors have argued that, on the contrary, the auditory cortex is sufficient for speech perception under normal listening conditions (Hickok, 2009; Lotto et al., 2009). According to the asymmetric sampling in time (AST) hypothesis (Poeppel, 2003), speech is processed bilaterally in primary auditory cortex and then elaborated in either left or right nonprimary auditory areas, depending on the length of the temporal integration window. AST predicts that distinctive features that are processed on faster time scales (Fig. 1) would be left lateralized while those that are processed on slower
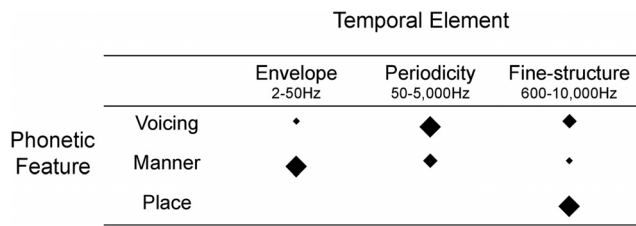
Temporal Element

Figure 1. Temporal elements of phonological features. The role of the three main temporal cues for speech—envelope, periodicity, and fine structure—in the perception of distinctive features. Each element's dominant fluctuation rate is listed in hertz and increases in frequency from left to right. The size of the diamonds indicates the extent to which a particular element contributes to a particular phonological feature, with a blank space indicating very weak or nonexistent cues. While all three temporal elements are relevant to voicing and manner of articulation, the envelope (slowest dominant fluctuation rate) is particularly informative of manner, and periodicity (intermediate dominant fluctuation rate) is particularly informative of voicing. Spectral shape, which is characterized by fine structure (fastest dominant fluctuation rate), is the primary acoustic cue for place. Adapted from Rosen (1992).

Table 1. Descriptions and selected examples of phonological features

Feature	Description	Selected categories	English examples
Voicing	Whether or not the vocal cords vibrate during speech production	Voiced Voiceless	/z/, /v/ /s/, /f/
Manner of articulation	The configuration of articulators during speech production	Fricatives Stops Nasals	/f/, /ʃ/ /p/, /k/ /m/, /n/
Place of articulation	Where the major constriction of the vocal passage occurs	Labial Dental Alveolar Palatoalveolar Velar	/b/, /m/ /ð /, /θ/ /d/, /t/ /ʃ/, /ʒ/ /k/, /g/

The selected categories were those that were investigated in the current study. Note that with the exception of voicing, the list does not include all possible categories, but rather those especially relevant for English consonants.

time scales would be right lateralized. It is unclear whether the frontal-motor system possesses the fine-scale temporal precision required for the detection of acoustic–phonetic features.

Here we investigate the relative contributions of auditory and motor regions to the distributed neural patterns supporting the perception of phonemes. We propose that the classic perceptual confusability patterns derived from psychophysical judgments should be statistically related to brain activity in regions important for phonological speech discrimination, offering a strong test for both motor and auditory theories of speech perception. We also investigate the extent to which distinctive features are coded in superior temporal and frontal cortices, and whether the patterns of lateralization in auditory cortex are consistent with the AST hypothesis.

Materials and Methods

Experimental methods

Participants. Twenty-five healthy young adults (mean age, 24.28 years; SD, 4.69 years; 14 females) were recruited from the Baycrest Hospital participant database. All were right-handed fluent English-speakers with no known neurological or psychiatric issues and no history of hearing or speech disorders. Participants gave informed written consent according to guidelines established by Baycrest's Research Ethics Board. One participant (28-year-old male) was excluded from imaging data analysis due to excessive movement (>5 mm maximum displacement from reference volume), leaving a sample of 24 participants.

Stimuli. Stimuli consisted of the following 16 phonemes followed by the vowel /a/ as in father: /p/, /t/, /k/, /f/, /θ/ ("th" as in thumb), /s/, / ʃ / ("sh" as in shoe), /b/, /d/, /g/, /v/, /ð/ ("th" as in that), /z/, / ʒ / ("s" as in measure), /m/, and /n/. The recordings were tokens from the standardized University of California, Los Angeles, version of the Nonsense Syllable Test (Dubno and Schaefer, 1992). Each consonant–vowel syllable (CV) was produced by four different speakers (two male, two female) and each speaker produced three different recordings of each sound. Thus, 12 different versions of each CV were used, ensuring that there was some acoustic variability across tokens of the same type. The sounds were played at a comfortable volume for participants through electrodynamic MR-Confon headphones. As depicted in Table 1, the stimuli spanned the major distinctive features.

Procedure. Testing occurred in a single 2 h session at Baycrest Hospital. All participants completed a practice block of 17 trials (16 CVs plus 1 silent trial) with feedback before the experiment began. Using Eprime 2.0, stimuli were presented sequentially while participants were in the MRI scanner and were first-order counterbalanced using a type 1, index 1 randomization algorithm (Aguirre, 2007) to minimize carryover effects. As seen in Figure 2, a crosshair appeared on the screen for the duration of the trial and went off for 200 ms to signify the beginning of a

new trial. Participants were instructed to fixate on the crosshair for the duration of the experiment. One CV was heard during each sound trial (1300 ms) and no sound was played during silent trials (4000 ms). Silent and sound trials were intermixed, with all 16 CVs and silent trials occurring with equal frequency (once every 17 trials). To ensure that participants focused on the sound stimuli without requiring explicit judgments about the acoustic–phonetic features of interest, participants were asked to respond to each sound trial with the gender of the speaker via button press. They were instructed to respond as quickly and accurately as possible after the onset of the speech stimulus. Half of the sounds were produced by male speakers and half by female speakers.

The short duration of trials required participants to respond very quickly; thus, even though the task itself involved making a simple decision, doing well on the task required a relatively high degree of attention. Participants were informed of their percentage of correct trials at the end of each block. Ten blocks of 204 trials each (total, 2040 trials) were presented such that the 16 CVs and silent trials were presented 120 times each throughout the experiment.

Imaging methods

MRI set-up and data acquisition. Participants were scanned with a 3.0 T Siemens Magnetom Trio MRI scanner using a 12-channel head coil system. High-resolution gradient-echo multislice T1-weighted scans (160 slices of 1 mm thickness, 19.2 × 25.6 cm field of view) coplanar with the echo-planar imaging scans (EPIs) as well as whole-brain magnetization prepared rapid gradient echo (MP-RAGE) 3-D T1-weighted scans were first acquired for anatomical localization, followed by T2*-weighted EPIs sensitive to BOLD contrast. Images were acquired using a two-shot gradient-echo EPI sequence (22.5 × 22.5 cm field of view with a 96 × 96 matrix size, resulting in an in-plane resolution of 2.35 × 2.35 mm for each of 26 3.5 mm axial slices with a 0.5 mm interslice gap; repetition time, 1.5 s; echo time, 27 ms; flip angle, 62°).

MRI preprocessing and whole-brain univariate regression analyses. Functional images were converted into NIfTI-1 (Neuroimaging Informatics Technology Initiative-1) format, motion-corrected, and realigned to the first image of the first run with AFNI's (analysis of functional neuroimages; Cox, 1996) 3dvolreg program. All image volumes were then smoothed with a 4 mm FWHM Gaussian kernel. Single-subject multiple-regression modeling was performed using the AFNI program 3dDeconvolve. Each of the 16 phonemes was modeled by convolving a hemodynamic response function (statistical parametric mapping canonical function as implemented in AFNI) with a delta function formed from the vector of speech stimulus onsets. An additional set of five nuisance regressors (a constant term plus linear, quadratic, and higher-order polynomial terms) was included for each scanning run to model low-frequency noise in the time series data. Separate regression analyses were performed for each of the 10 scanning runs, yielding a set of 16 (one per CV) β regression weights for each run. The regression models were computed separately for each run so that we would have independent samples for cross-validation in the multivariate analyses.

All statistical analyses, both voxelwise and region-of-interest (ROI) based, were first conducted on the spatially smoothed and realigned functional images in the participant's native EPI space. The MP-RAGE anatomical scan was normalized to the Montréal Neurological Institute (MNI) template using nonlinear symmetric normalization implemented in Advanced Normalization Tools (ANTs; Avants et al., 2008). An equivalent transformation was then applied to maps of univariate statistical results (see Voxelwise correlation of MDS dimensions with fMRI activation) derived from functional images using ANTs to normalize these maps to MNI space for multisubject analyses. Statistical significance at the group level was determined using Monte Carlo simulations of expected cluster sizes under the null hypothesis using the AFNI program AlphaSim. For a voxelwise threshold of $p < 0.005$, only clusters with >13 voxels were determined to be significant at the cluster-corrected ($p < 0.05$) level.

Data analysis
Behavioral data. Trials with reaction times of <400 ms were removed from behavioral analyses. Trials for which the participant did not respond within 1500 ms of stimulus onset were scored as incorrect. A 16 (syllable) × 4 (speaker) repeated-measures ANOVA was performed on accuracy scores. We conducted further ANOVAs to test for an effect of manner (stops, fricatives, and nasals) and place of articulation (labial, dental, alvolar, palatoalveolar, and velar), as well as a paired-sample *t* test to assess voicing effects (voiced vs voiceless).

Multidimensional scaling analysis. Miller and Nicely (1955) recorded phonemic identification data (how often each CV stimulus was incorrectly identified as another CV) from participants using the 16 phonemes used in the current study to establish confusability matrices at different levels of noise. They found that while the number of confusions is affected by noise, the overall pattern is not. Thus, the fact that, for example, /m/ is more confusable with /n/ than it is with /k/ is both reliable and informative—it tells us something about the way in which our brains process speech sounds. Following Shepard (1972), nonmetric multidimensional scaling [MDS; using isoMDS from the MASS (Modern Applied Statistics with S) package in the R programming language; Venables and Ripley, 2002] was applied to the phonemic confusion data taken directly from Tables 1–6 in Miller and Nicely's (1955) report (noise levels: −18, −12, −6, 0, +6, +12 dB). The acoustic confusion data were first converted to distances using a logarithmic transformation of the normalized confusion probabilities (Shepard, 1972). Nonmetric MDS was then applied to the resultant distance matrix and a two-dimensional solution was computed, resulting in a continuous and empirically derived map reflecting the relative positions of each phoneme in coordinate space. The result of this analysis closely resembles Shepard's original solution (Fig. 3; cf. Shepard, 1972, Fig. 4).

The first dimension, plotted on the *x*-axis, approximately corresponds to the feature known as voicing, with unvoiced stimuli on the left side and voiced stimuli on the right. The second dimension, plotted on the *y*-axis, is related to manner of articulation, with nasals occupying the upper half of the graph and fricatives and stops occupying the lower half.
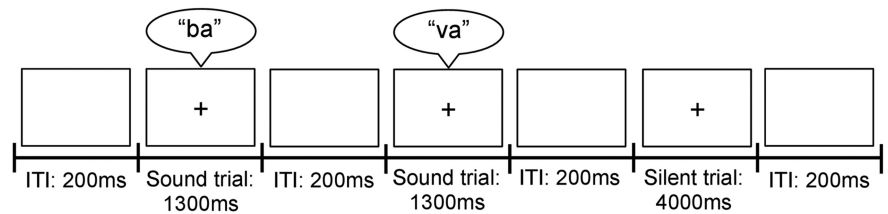

Figure 2. Task design. Two sound trials (1300 ms each) and one silent trial (4000 ms) are depicted in slides with crosshairs. Participants hear a single CV during every sound trial and must respond with the gender of the speaker. Silent trials contain only a visual crosshair, thus requiring no response and allowing the hemodynamic response function to return to homeostatic levels.
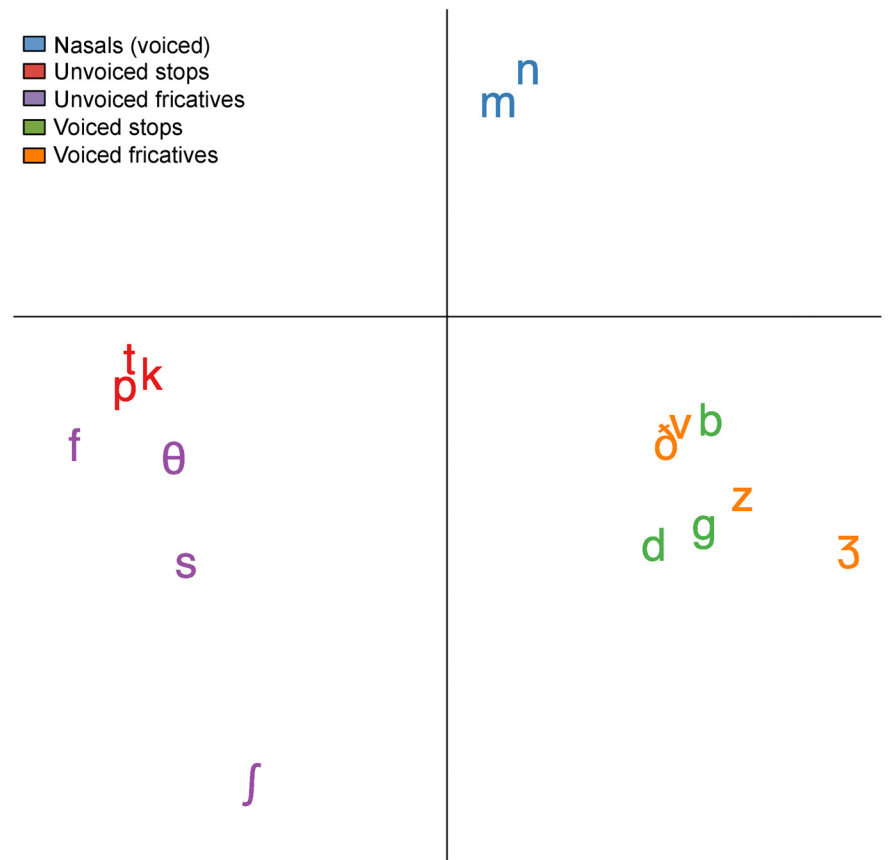

Figure 3. Multidimensional scaling of 16 consonants. Two-dimensional spatial representation of 16 phonemes based on pooled data from Miller and Nicely (1955). Voicing is represented on the *x*-axis, with unvoiced phonemes occupying the left side and voiced phonemes occupying the right side. Manner of articulation is approximately represented on the *y*-axis, with nasals occupying the top half of the figure and fricatives and stops occupying the bottom. A clear division is observed between the unvoiced stops and fricatives, while overlap exists between voiced stops and fricatives. /θ/, "th" in thumb; /ʃ/, "sh" in shoe; /ð/, "th" in that; /ʒ/, "s" in measure.

Voxelwise correlation of MDS dimensions with fMRI activation. To examine the relation between the two MDS-derived dimensions based on perceptual distance (Fig. 3, *x*-axis and *y*-axis) and brain activation, we computed separate Spearman rank correlations between each of the first two dimensions and the corresponding β estimates, averaged over runs, for the 16 CVs. This produced rank correlations for each MDS dimension for all voxels in the brain and all participants in the group. To test for reliable effects at the group level, these correlation images were *z*-transformed, spatially normalized to MNI space, and submitted to a voxelwise one-sample *t* test.

Multivariate analyses. Although perceptual confusability offers a means of investigating the representational basis of acoustic–phonetic speech perception, MDS coordinate maps derived from such data are dominated by only two of the three major distinctive features, namely, voicing and manner of articulation. However, because place of articula-

tion (or something akin to it) must be represented at some level in the brain— otherwise sounds that share voicing and manner but differ in place (such as "tan" and "can") would be indistinguishable from one another—perceptual confusability does not fully reflect the information contained in the underlying neural code. Moreover, from a theoretical and empirical standpoint, place of articulation has played a prominent role in studies of acoustic–phonetic speech perception (see Discussion). Furthermore, place of articulation is associated with rapid fluctuations in spectrotemporal fine structure and, unlike manner and voicing features, is not easily recognized on the basis of lower-frequency envelope information (Shannon et al., 1995). Thus, according to AST and other theories of speech perception that posit left hemisphere specialization for rapid temporal processing, place of articulation should be more robustly represented in the left auditory cortex than in the right. We therefore investigated the distribution of patterns of activity associated with the major distinctive features of voicing, manner, and place of articulation directly, rather than through the lens of perceptual confusability.

To do so, we used a class of techniques known as multivoxel pattern analysis (MVPA). The basic premise of MVPA is to identify reliable patterns of activation rather than relying on subtraction logic that is typical of univariate approaches. The main benefit of MVPA is that it takes advantage of systematic variance distributed across voxels rather than being based only on the average activity within a single voxel. Thus, we can train pattern classifiers to discriminate spatial patterns associated with one feature (i.e., voiced) from those associated with another feature (i.e., unvoiced) and then test these classifiers on independent test trials.

To identify reliable patterns of distributed brain activity associated with each feature, a series of MVPA analyses was carried out. We used Freesurfer's (Dale et al., 1999) automatic anatomical labeling ("aparc 2009"; Destrieux et al., 2010) algorithm to define a set of 148 cortical and subcortical ROIs. These ROIs were defined using each participant's high-resolution anatomical scan and therefore group analyses could be performed without applying any spatial normalization. We used anatomically defined ROIs rather than a moving "searchlight" (Kriegeskorte et al., 2006) procedure because we wished to preserve borders between spatially adjacent regions along the sylvian fissure (e.g., ventral frontal cortex and superior temporal cortex). This was especially important in light of one the major aims of this study, namely, to quantify the contribution of temporal and prefrontal regions—portions of which are adjacent in volumetric space—to the neural representations underlying phoneme perception.

For all MVPA analyses we used shrinkage discriminant analysis (SDA) as implemented in the R package "sda" (http://cran.rproject.org/web/packages/sda/). SDA is a form of linear discriminant analysis that estimates shrinkage parameters for the variance–covariance matrix of the data, making it suitable for high-dimensional classification problems. It has the advantage of these shrinkage parameters being estimated analytically from the data, obviating a doubly nested cross-validation scheme (Ahdesmäki and Strimmer, 2010). To evaluate classifier performance, we used 10-fold cross-validation where each fold of data consisted of the β regression weights of nine of the 10 runs, with one run held out for testing. Thus, during cross-validation, observations drawn from the same scanning run were never part of both the training and test datasets. MVPA analyses were performed within each anatomical ROI, or in groups of ROIs (see below), yielding regional estimates of classifier performance. The SDA classifier produces both a categorical prediction (i.e., the label of the test case) as well as a continuous probabilistic output (the posterior probability that the test case is of label x). The continuous outputs were used to compute area-under-the-curve (AUC) metrics and the categorical predictions were used to compute classifier accuracy.

To test for statistical significance at the group level, classifier performance for each ROI was evaluated with a one-sample t test where the null hypothesis assumed a theoretical chance AUC of 0.5. To validate this assumption, we performed permutation analyses in which we randomly shuffled the labels in the training set for each cross-validation fold and repeated the process 100 times for each subject, ROI, and phonological feature category. The grand mean, over subjects and ROIs, of these permutation-based AUC values was 0.50000219, which by the exact binomial test was not significantly different from the theoretical chance level of 0.5. We therefore concluded that for null hypothesis testing we could assume that an AUC of 0.5 is the expected value when the classifier is performing at chance.

Because classification analyses were conducted in anatomically defined ROIs specific to each subject, no spatial normalization was applied to the subject-specific classifier performance scores. To display statistics at the group level, the statistic of interest was projected on the parcellated (aparc 2009; Destrieux et al., 2010) cortical flat map associated with the Freesurfer average template ("fsaverage").

Exploratory MVPA of individual feature subcategories. To investigate the relative contributions of each brain area to the neural pattern of activity associated with distinctive features, separate classifiers were trained to detect the subcategories of voicing (voiced and voiceless), manner of articulation (nasals, stops, and fricatives), and place of articulation (labial, dental, alveolar, palatoalveolar, and velar). Because the full phonological feature matrix is not orthogonal for the set of 16 CVs—and thus feature dimensions are confounded across phonemes—we used a subsampling approach to ensure that classification for one feature category (e.g., nasal) could not be driven by a correlated feature category (e.g., voiced). To achieve this, we trained classifiers to discriminate each feature category from a subset of phonemes matched across the other feature dimensions. For example, a classifier was trained to discriminate labials from the subset of nonlabials matched on voicing and manner of articulation. Thus, nine classifiers (voiced, stop, fricative, nasal, labial, dental, alveolar, palatoalveolar, velar) were trained for each ROI. Note that because voicing consists of only two categories—voiced and voiceless—training two classifiers would be redundant as one is essentially the inverse of the other. We computed both classification and AUC as performance metrics. However, we preferred the latter metric for group statistics due to its increased sensitivity (Bradley, 1997). To test for significant group classifier performance using AUC, we used a one-sample t test where the null hypothesis (chance performance) assumed an AUC of 0.5, which is the expected value for the AUC when there is no relationship between the continuous classifier output and the category labels of the test cases. We then counted the number of significant feature subcategories ($p < 0.05$, uncorrected) for each ROI. Regions with ≥ 3 significant features ($p < 0.0083$, by the binomial probability distribution) were deemed significant in this exploratory analysis.

MVPA of distinctive features in speech-sensitive cortex. The overall performance for a feature (voicing, manner, and place of articulation) was then computed as a weighted average of the individual category classifiers, where the weights were determined by the frequency of each category. For example, the performance score for manner of articulation was calculated by averaging classifier performance for stops ($N = 6$), fricatives ($N = 8$), and nasals ($N = 2$) with the following weights: 0.375, 0.5, and 0.125. Statistical significance at the group level was evaluated with a one-sample t test corrected for multiple comparisons using a false discovery rate of $q = 0.05$.

In this analysis we limited the ROI search space to regions known to be sensitive to tasks involving the production and perception of speech. We used Neurosynth (http://neurosynth.org/; Yarkoni et al., 2011) to create a meta-analytic mask using the search term "speech" (http://neurosynth.org/features/speech/). This resulted in a coordinate-based activation mask constructed from 424 studies and encompassing the language-related areas in the temporal and frontal lobes. We intersected this meta-analytic mask with the Freesurfer aparc 2009 ROI mask as defined in MNI space. If any of the intersected ROIs had ≥ 10 voxels, we included that ROI in our search space. To ensure hemispheric symmetry, if a left hemisphere ROI was included so too was its right hemisphere homolog. This resulted in an ROI mask consisting of 38 ROIs (19 left hemisphere and 19 right hemisphere), which are shown in Figure 4.

Hemispheric differences in the superior temporal lobe. To test for lateralization effects in MVPA measures within broadly defined auditory processing structures of the superior temporal lobe, paired-samples t tests were performed on a subset of left versus right auditory cortex ROIs for classification performance of voicing, manner, and place. Auditory cortex was defined as any region in the temporal lobe extending inferiorly to the middle temporal gyrus, medially to the posterior insula, anteriorly to the planum polare, and posteriorly to the lateral fissure. According to this

definition, the following ROIs were included: transverse temporal gyrus [Heschl's gyrus (HG)], transverse temporal sulcus, STG, middle temporal gyrus, planum polare, planum temporale (PT), lateral fissure, insula, and superior temporal sulcus (Fig. 4, 1–9). Based on the AST hypothesis (Poeppel, 2003), our prediction was that features that are processed on faster time scales would be left lateralized while those that are processed on slower time scales would be right lateralized. As reviewed in Rosen (1992), the three main temporal cues for speech are the envelope, periodicity, and fine structure, each of which is based on varying fluctuation rates and have a differential impact on the perception of phonological features (Fig. 1). Thus, because voicing and manner of articulation are based on all three temporal cues, it was predicted that these features would be represented bilaterally in the superior temporal cortices, while place would be left lateralized due to its requirement for rapid temporal processing.

Deriving neural confusability matrices through classification of individual phonemes. In a final analysis, we examined whether individual phonemes could be reliably classified from patterns of activity across the left and right superior temporal lobe and whether the pattern of confusions made by the classifier is correlated with the original perceptual confusability matrices reported in Tables 1–6 of Miller and Nicely's (1955) classic study. Using the combined ROI sets defining the superior temporal lobe described in the previous section, we trained three classifiers (one for each hemisphere and one for both hemispheres combined) to distinguish among each of the 16 CVs, and recorded classification accuracy (chance performance, 0.0625; or 1 of 16) and the full matrix of phoneme confusions produced by the classifier predictions. The raw confusion counts were then normalized by the row totals (i.e., the number of times the classifier predicted each phoneme), yielding conditional probabilities for each ordered phoneme pair. Confusion probabilities between the same pairs (e.g., m → n and n → m) were then averaged to generate a symmetric probability matrix. This matrix was converted to a distance matrix using Shepard's log transformation (see Multidimensional scaling analysis) and then rank correlated with the distance matrix shown in Figure 2. To evaluate statistical significance for phoneme classification across the group, we computed three (left, right, both hemispheres) one-sample *t* tests against a null hypothesis of chance accuracy (0.0625).

Figure 4. ROIs in speech-relevant areas of the temporal and frontal lobes. An ROI mask was created by intersecting a Neurosynth automated meta-analysis (search term: "speech"; Yarkoni et al., 2011) and the 148 Freesurfer ROIs (aparc 2009 atlas; Destrieux et al., 2010). The final set of 38 ROIs (19 left, 19 right) spanned cortical areas in the temporal and frontal lobes important for the perception and production of speech.

(1) Planum polare
(2) Inferior insula
(3) Heschl's gyrus
(4) Transverse temporal sulcus
(5) Superior temporal gyrus
(6) Superior temporal sulcus
(7) Middle temporal gyrus
(8) Planum temporale
(9) Posterior lateral fissure
(10) Inferior supramarginal gyrus
(11) Subcentral gyrus/sulcus
(12) Postcentral gyrus
(13) Central sulcus
(14) Precentral gyrus
(15) Inferior precentral sulcus
(16) Inferior opercular gyrus
(17) Anterior lateral fissure
(18) Inferior triangularis
(19) Frontal superior gyrus

Figure 5. Brain activity related to MDS dimensions. Activation maps of MDS results rendered onto MNI template showing clusters of >13 voxels at *p* < 0.005. The figures on the top display clusters correlating with positive values on Dimension 1 (approximately corresponding to the continuum of voicing information), with activation in bilateral anterior STG; the bottom images display clusters correlating with positive values on Dimension 2 (approximately corresponding to manner of articulation), with activation in bilateral HG and right PT. Both sagittal images represent right hemisphere templates.

Results
Behavioral performance
The average accuracy for the gender judgment task was 97% (SD, 2.7; range, 87–100%). A 16 (syllable) × 4 (speaker) repeated-measures ANOVA showed no significant main effects or interactions for accuracy data (*p* > 0.05). Further ANOVAs showed no significant effects on accuracy scores for manner ($F_{(2)} = 0.269$, $p = 0.766$) and place of articulation ($F_{(4)} = 2.185$, $p = 0.076$); a paired-samples *t* test revealed no significant difference between voiced and voiceless stimuli ($t_{(24)} = 0.327$, $p = 0.746$). Thus, participants were highly accurate for the behavioral task and
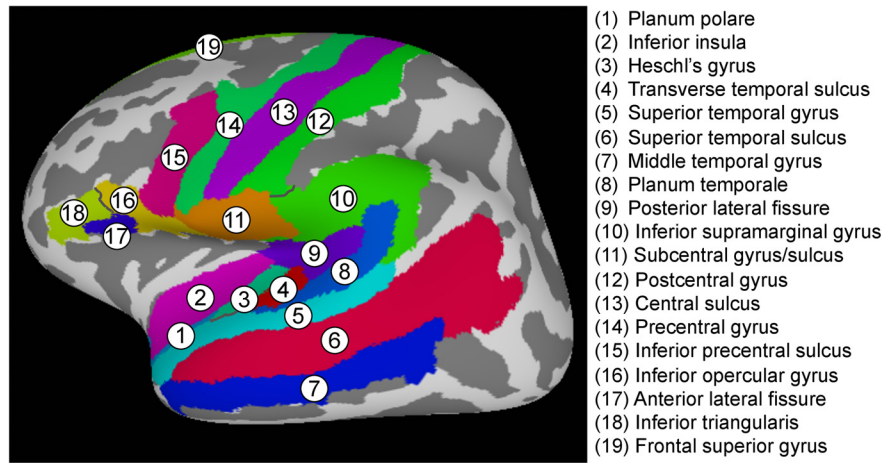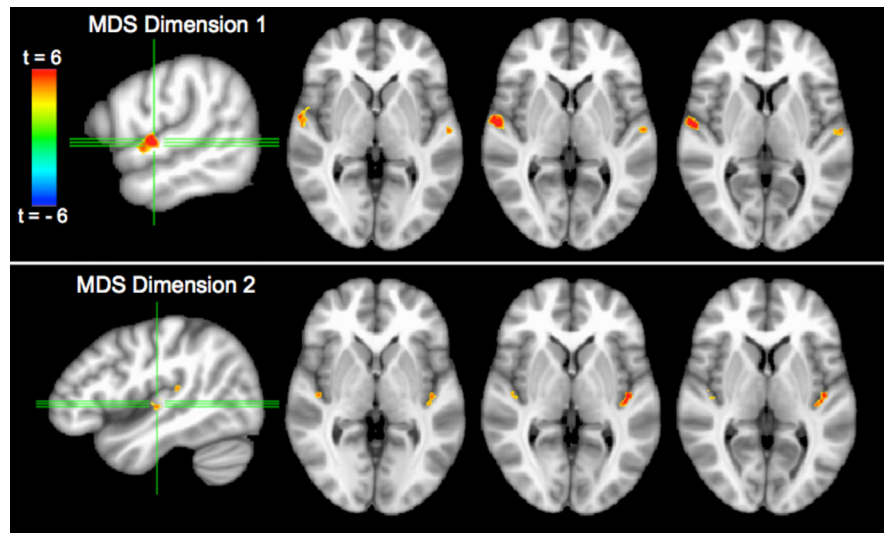
showed approximately equivalent performance across speakers, syllables, and phonological features.

Correlation of MDS dimensions with fMRI activation
As can be seen in Figure 5, brain activity that correlated with the first MDS dimension was found in the midanterior portion of the STG, bilaterally. Positive *t* statistics correspond to activation positively correlated with the MDS coordinates, and are indicated in warm colors. Thus, because voiced phonemes have positive values on the first dimension, activation in the midanterior STG essentially increased as a function of voicing. The second MDS dimension, which distinguished among syllables based on manner of articulation, was significantly correlated with clusters located anterior and posterior to HG. Here, warm-colored clusters approximately indicate areas where activity was significantly greater for nasals than for fricatives or stops. No significant effects

Table 2. Peak voxels correlated with MDS dimensions

Region	X	Y	Z	Cluster size	t
Dimension 1					
Right anterior STG	60	−9	4	44	9.32
Left anterior STG	61	12	7	31	5.24
Dimension 2					
Left anterior HG	43	21	4	18	6.83
Right PT	57	−27	16	15	4.65
Right anterior HG	48	−15	1	14	4.69

MNI coordinates for peak voxels of increased activation during speech perception that correlate with Dimension 1 and 2.
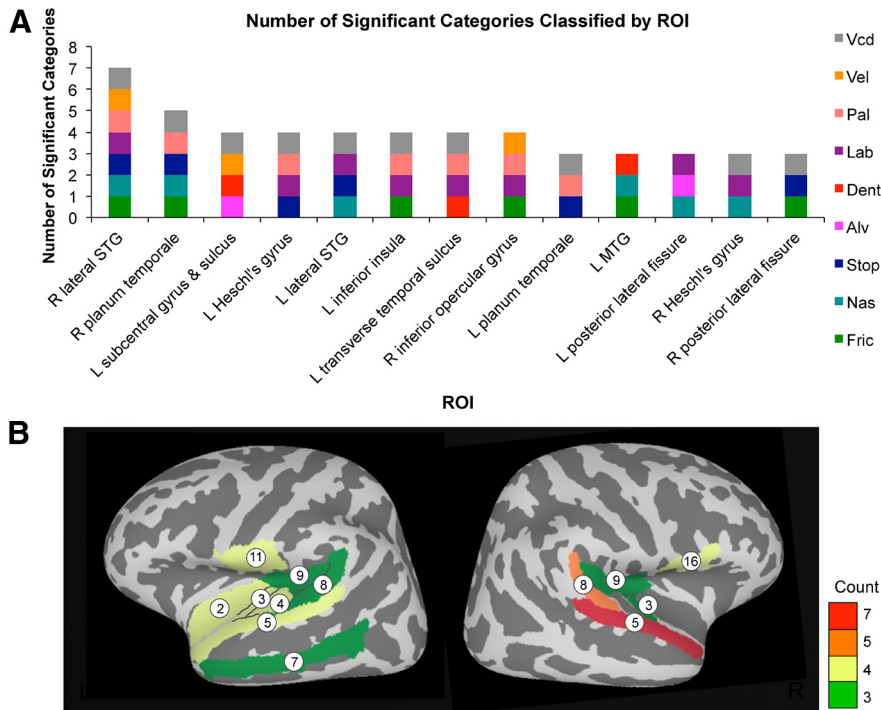

Figure 6. ROIs with significant classification accuracy in ≥3 categories. *A*, The number of categories for which each ROI obtained significant (equivalent to a *p* value < 0.0083) classification accuracy is displayed on the *y*-axis, with ROIs listed across the *x*-axis. Gray represents voicing information, warm colors indicate place, and cool colors indicate manner of articulation. *B*, ROIs associated with significant classification in ≥3 categories projected onto an inflated surface. Colors represent number of phonological categories significantly classified per region (note that the scale is not continuous). Vcd, Voiced; Vel, velar; Pal, palatoalveolar; Lab, labial; Dent, dental; Alv, alveolar; Nas, nasal; Fric, fricative; 2, inferior insula; 3, HG; 4, transverse temporal sulcus; 5, STG; 7, middle temporal gyrus; 8, PT; 9, posterior lateral fissure; 11, subcentral gyrus/sulcus; 16, inferior opercular gyrus.

were found outside of the auditory cortex. The peak coordinates for all significant clusters are listed in Table 2.

Multivariate results
MVPA of individual feature subcategories
To investigate distributed patterns of activity associated with major distinctive features, rather than perceptual confusability per se, MVPA was used to classify groups of phonemes based on the phonological category to which they belonged. To get a general overview of areas sensitive to abstract phonological features, we tallied for each ROI the number of individual subcategories (e.g., labial, fricative, nasal, etc.) that were reliably classified at *p* < 0.05 at the group level. ROIs with significant classification accuracy for ≥3 categories (corresponding to an uncorrected *p* value < 0.0083 by the binomial distribution) are listed in Figure 6A. Figure 6B displays the significant ROIs projected onto an inflated surface.

The right STG was associated with the highest number of categories for which significant classification was observed, with

seven of the nine categories being represented. The vast majority of ROIs with significant classification accuracy in ≥3 categories were contained within the temporal lobes, with the exception of the left subcentral gyrus (Brodmann area 43 at the ventral postcentral gyrus; Fig. 6B, 11) and right inferior frontal gyrus (IFG) opercularis (Brodmann area 44; Fig. 6B, 16), which were each associated with significant classification of four feature categories, three of which were from the place of articulation feature.

MVPA of distinctive features
ROIs within our speech-sensitive mask with significant (false discovery rate, 0.05) classification accuracy for distinctive features (voicing, manner, place) are displayed in Figure 7. As stated in the Materials and Methods, this analysis is based on the frequency-weighted means of the performance measures derived from the individual category classifier.

In this analysis, with the exception of the subcentral gyrus/sulcus (Brodmann area 43), only areas in the superior temporal lobe show significant classification accuracy to one or more of the three distinctive features. Specifically, information regarding voicing and place is prominent throughout the left STG, with voicing information extending anteriorly along medial auditory cortex and posteriorly into PT, and place information throughout auditory cortex and perisylvian areas. All three features overlap in the left HG. In the right hemisphere, voicing and manner are more prominent in the PT and HG, with manner extending into posterior lateral fissure. Overlap of all three features can be seen in right STG. Note the absence of significant feature classification in the right IFG, as was seen in the analysis of individual categories from the previous section. This area showed high but subthreshold *t* values for place of articulation in the weighted feature analysis ($t_{(23)} = 2.16$, $p = 0.04$), so the discrepancy is more apparent than real.

Hemispheric differences in the superior temporal lobe
A subset of nine ROIs (left and right; Fig. 4, 1–9) was used to further investigate the role of hemispheric lateralization specific to the bilateral superior temporal cortex. To test the prediction of the AST model that place of articulation should be left biased, we computed a one-tailed *t* test comparing classification performance across hemispheres. A significant one-tailed effect was found for place ($t_{(23)} = 1.87$, $p = 0.038$). A comparison of the hemispheric differences for the other features was not significant, however (voicing, $t_{(23)} = -0.726$, $p = 0.4$; manner, $t_{(23)} = -0.31$, $p = 0.6$). To further investigate which ROIs were driving the laterality effect for place of articulation, one-tailed *t* tests were performed on each ROI listed above. In seven of the nine ROIs, greater classification accuracy was observed on the left versus right hemisphere, although the difference was statistically reliable only in the posterior temporal lobe along the lateral fissure ($t_{(23)} = 1.874$, $p = 0.037$). The difference scores for the mean AUC values for each ROI are displayed in Figure 8.

MVPA classification of individual consonants

The restricted set of nine superior temporal cortex ROIs described above was combined as a single superior temporal lobe mask to classify individual phonemes in each participant. Three classifiers were trained: one for the left hemisphere, one for the right hemisphere, and one that pooled ROIs across hemispheres. Classification accuracy (chance accuracy, 0.0625) for the combined set of left and right hemisphere superior temporal lobe ROIs was reliable across participants (mean accuracy, 0.088; $t_{(23)} = 7.18$, $p < 0.0001$).

Figure 7. Brain activity related to phonological features. Overlapping information of voicing (green), place (red), and manner (blue) in ROIs within the speech mask projected onto inflated surface. Notably, all three phonological features are represented in right STG as well as left HG (yellow). 2, Inferior insula; 3, HG; 4, transverse temporal sulcus; 5, STG; 8, PT; 9, posterior lateral fissure; 11, subcentral gyrus/sulcus.

Classification accuracy for both the left and right hemisphere ROI sets was also reliable (left hemisphere mean accuracy, 0.084; $t_{(23)} = 4.6$, $p < 0.0001$; right hemisphere mean accuracy, 0.076; $t_{(23)} = 2.91$, $p < 0.008$). Although the left hemisphere had a marginally higher rate of classification, the difference was not significant (left–right, $t_{(23)} = 1.2$, $p = 0.23$), indicating that phoneme-specific information was contained in neural activity patterns to an approximately equal extent in both the left and right superior temporal cortices during speech perception.

Comparison of neural and behavioral phonemic confusion matrices

To evaluate whether the neural phonemic confusion matrix derived from the MVPA classification analysis of individual syllables in the superior temporal lobe described in the previous section resembles the perceptual confusion matrix from Miller and Nicely (1955; for MDS representation of these data, see Fig. 3), we tested whether the two matrices share significant variance by correlating their respective interitem distances. Thus, a Spearman rank correlation coefficient was computed between lower diagonal elements of the two distance-transformed confusion matrices, yielding a correlation of 0.44. Because the set of interitem distances are not independent, statistical significance was assessed using the permutation-based Mantel test (Abdi, 2010). Using 1000 permutations to derive a null distribution of the rank correlation between the perceptual distance matrices, the correlation of 0.44 was determined to be significant ($p = 0.0009$). This analysis confirms a substantial shared structure in the confusability of phonemes in neural and psychological space.

Discussion

The objective of the current study was to investigate both local and distributed neural codes associated with phonological feature representations during speech perception. We began with the premise that perceptual confusability—the behavioral tendency to confuse one phoneme with another—should be evident as patterns of neural similarity in brain structures critical for acoustic–phonetic perception (Shepard and Chipman, 1970; Kriegeskorte et al., 2008). The foregoing logic was used first to evaluate the resurgent idea that the motor system directly contributes to acoustic–phonetic speech perception (Meister et al., 2007; D'Ausilio et al., 2009). We showed on the contrary that the orthogonal axes derived from an MDS analysis of Miller and Nicely's (1955) phonemic confusability data were significantly correlated with regions in the bilateral auditory cortices, but not in motor/premotor cortices. Specifically, the first MDS dimension, which separates voiced from unvoiced consonants, was positively related to bilateral activity in the midanterior portion of the STG, a region that has previously been implicated in acoustic–
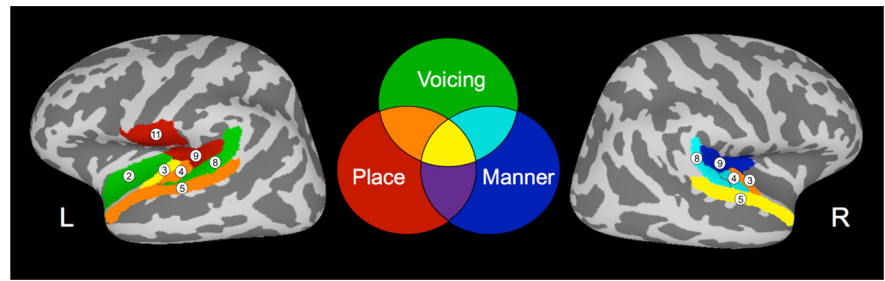
phonetic perception (Formisano et al., 2008; Moerel et al., 2012). The second MDS dimension, which separates nasals from stops and fricatives, was significantly correlated with voxels in the bilateral supratemporal plane both anterior and posterior to HG. In short, this univariate correlational analysis relating measures derived from a classic behavioral dataset published six decades ago offers the first evidence to date that the underlying dimensions describing the patterns of phonemic confusability of consonants are captured by local variation in BOLD activity in the bilateral auditory—but not frontal-motor—cortices.

Although perceptual confusability captures aspects of the psychological representation of speech stimuli, it is weakly related to place of articulation, the feature that has played a central role in neuroscience studies of acoustic–phonetic speech perception. In the context of assessing motor contributions to speech perception, investigations using fMRI (Lee, et al., 2012; Chevillet et al., 2013) and transcranial magnetic stimulation (TMS; Meister et al., 2007; D'Ausilio et al., 2009) have usually focused on place of articulation variables because of the generally straightforward hypothesis that, for example, labials and alveolars should preferentially activate lip and tongue areas of somatotopically defined motor cortex. Moreover, place of articulation requires rapid temporal processing and is theoretically relevant to the AST model of speech perception, which predicts that phonological features that require the detection of rapid spectrotemporal fluctuations should show a left hemisphere bias in the auditory cortex. Thus, we used the distinctive features of voicing, manner, and place of articulation for the MVPA analyses, rather than relying on perceptual confusability patterns.

Using a regional MVPA approach, we found significant classification of distinctive features throughout the left and right auditory cortices. Voicing was the most robustly classified, appearing as a significant feature in the majority of the auditory cortex ROIs in both hemispheres. Manner and place were also reliably classified in several auditory cortical zones. While there have been few investigations of the neural basis of manner of articulation, our results confirm research using electrocorticography showing that distributed areas of the left STG selectively respond to stops, fricatives, and nasals (Mesgarani et al., 2014). Previous research on place of articulation has broadly implicated STG (Pulvermüller et al., 2006; Steinschneider et al., 2011; Chevillet et al., 2013), as well as some areas of the motor cortex (Pulvermüller et al., 2006; D'Ausilio et al., 2009; Möttönen et al., 2013). Recently, Kilian-Hütten and colleagues (2011) used a pattern classifier to distinguish perceptual differences between syllables that vary along the place of articulation dimension (labial vs alveolar). The authors found discriminative voxels along left HG and sulcus, which is consistent with the pattern that we observed.
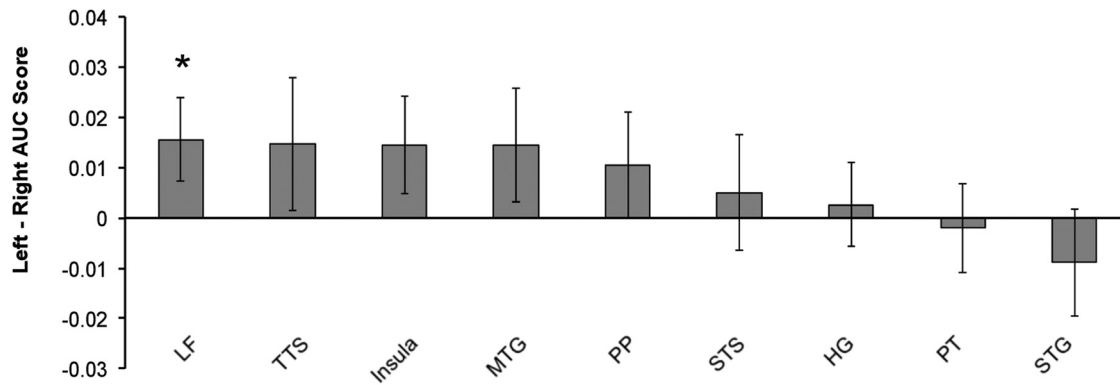
Figure 8. Hemispheric differences (left–right) in classification accuracy for place of articulation across nine auditory cortex ROIs. Although the general tendency reflects a leftward bias, only lateral fissure achieved significance ($p < 0.05$ one-tailed) on individual t tests. Error bars represent SEM. LF, Lateral fissure; TTS, transverse temporal sulcus; MTG, middle temporal gyrus; PP, planum polare; STS, superior temporal sulcus.

Role of motor cortex in speech processing

In contrast to some previous work (Pulvermüller et al., 2006; D'Ausilio et al., 2009; Möttönen et al., 2013), we did not observe reliable classification of place of articulation in the left motor/premotor cortex. We did, however, find modest evidence for place sensitivity in the subcentral gyrus and right IFG (Fig. 6). These effects were present in the tally of individual significant categories (≥ 3 categories with significant effect) but right IFG was not significant when place was treated as a single overarching feature and defined as a weighted average of its constituent categories (Fig. 7). Nevertheless, it is interesting to note that in monkeys, the secondary somatosensory and parietal ventral areas of the postcentral gyrus contain neurons that respond to somatosensory stimulation to the mouth, lips, and teeth (Padberg et al., 2005), an observation that may have relevance to the current finding but requires further investigation. In humans, Raizada and Poldrack (2007) used an adaptation approach and found sensitivity to categorical perception along the /ba/ to /da/ continuum in a similar region of the ventral parietal cortex.

While the univariate and multivariate analyses used in the current study provided slightly different types of information, neither provided positive evidence that the motor/premotor cortex is involved in phonological feature representation. Alternative roles that have been proposed for the motor system during speech perception have included turn taking (Scott et al., 2009), categorical processing (Lee et al., 2012), providing acoustic templates to aid in perceiving speech in noisy circumstances (Du et al., 2014), speech segmentation (Sato et al., 2009), and the automatic activation of articulatory information (Yuen et al., 2010). There is also recent evidence suggesting that motor/premotor cortex is specifically involved in phonological categorization. Krieger-Redwood and colleagues (2013) compared the effects of TMS to premotor cortex and STG using a phonological versus semantic categorization task that used the same stimuli (spoken words whose ending phoneme distinguished between two real words, such as "cart" and "carp"). They found that TMS to premotor cortex interfered with the phonological judgments but not the semantic task. Perhaps, then, the motor/premotor cortex is assisting in fine-grained categorical decisions by aiding participants in "sounding it out." When sounds cross perceptual boundaries, we often resort to subvocal rehearsal to repeat these sounds and make decisions on them. This notion is supported by fMRI evidence from Papoutsi and colleagues (2009), which identified the dorsal pars opercularis as being involved in syllabification.

Interestingly, the above example as well as other fMRI and TMS studies that have implied a role for the motor cortex in speech per-

ception (Meister et al., 2007; D'Ausilio et al., 2009) use stimuli that vary according to place of articulation, which is most intuitively connected to the motor system but least empirically connected to behavioral measures of speech perception (as evidenced in the MDS analysis above, as well as in Peters, 1963; Shepard, 1972; and Rosen, 1992). Thus, it remains unclear whether the frontal-motor system is sensitive to phonological categorization in general or to the place of articulation feature in particular. One caveat to our null effects in motor cortex and other ROIs is that the use of multiple genders and a gender discrimination task potentially could have influenced pattern discrimination accuracy unequally across phonological categories (Ryalls et al., 1997; Munson, 2011; Bonte et al., 2014), a possibility that we are currently evaluating in an fMRI study that uses MVPA in a passive listening task.

Role of bilateral auditory cortices in speech processing

We further examined laterality differences for each distinctive feature to test the prediction of the AST model of speech perception that place of articulation should be preferentially processed in the left hemisphere, whereas manner and voicing features, which can be distinguished even in high-pass filtered signals (Shannon et al., 1995), need not show a left-hemisphere bias. We confirmed that the pooled effect across all auditory cortical ROIs was significantly greater in the left than in the right hemisphere for place of articulation, but not for manner or voicing. The laterality effect was pronounced around the transverse temporal sulcus and lateral fissure, regions that have previously been implicated with rapid temporal analysis of sounds (Zatorre and Belin, 2001; Meyer et al., 2005). It should be noted, however, that although the overall effect in auditory cortex was left biased, the right STG showed significant sensitivity to place of articulation and indeed showed nonsignificantly higher classification performance than the left STG. Thus, although modest hemispheric differences in pattern activity appear to be associated with the place of articulation feature, this general bias does not hold for every region of the superior temporal cortex.

Conclusion

We have found neural codes related to both perceptual confusability and phonological features throughout the bilateral auditory, but not frontal-motor, cortices. Our results suggest a modest functional asymmetry in the auditory cortex based on the time scales on which voicing, manner, and place are processed, which is consistent with the AST hypothesis (Poeppel, 2003). Further, we have shown that the neural representations associ-

ated with speech perception are well captured by patterns of perceptual confusability arising from classic work detailing the organization of consonants in "psychological space."

References

Abdi H (2010) Congruence: congruence coefficient, Rv-coefficient, and Mantel coefficient. In: Encyclopedia of research design (Salking NJ, Nougherty DM, Frey B, eds), pp 222–229. Thousand Oaks, CA: Sage.

Aguirre GK (2007) Continuous carry-over designs for fMRI. Neuroimage 35:1480–1494. CrossRef Medline

Ahdesmäki M, Strimmer K (2010) Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. Ann Appl Stat 4:503–519. CrossRef

Avants BB, Epstein CL, Grossman M, Gee JC (2008) Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med Image Anal 12: 26–41. CrossRef Medline

Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E (2014) Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. J Neurosci 34:4548–4557. CrossRef Medline

Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Patt Recog 30:1145–1159. CrossRef

Brouwer GJ, Heeger DJ (2009) Decoding and reconstructing color from responses in human visual cortex. J Neurosci 29:13992–14003. CrossRef Medline

Chevillet MA, Jiang X, Rauschecker JP, Riesenhuber M (2013) Automatic phoneme category selectivity in the dorsal auditory stream. J Neurosci 33:5208–5215. CrossRef Medline

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comp Biomed Res 29:162–173. CrossRef

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis I. Segmentation and surface reconstruction. Neuroimage 194:179–194. Medline

D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L (2009) The motor somatotopy of speech perception. Curr Biol 19:381–385. CrossRef Medline

Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53:1–15. CrossRef Medline

Du Y, Buchsbaum BR, Grady CL, Alain C (2014) Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc Natl Acad Sci U S A 111: 7126–7131. CrossRef Medline

Dubno JR, Schaefer AB (1992) Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. J Acoust Soc Am 91:2110–2121. CrossRef Medline

Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322: 970–973. CrossRef Medline

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

Kilian-Hütten N, Valente G, Vroomen J, Formisano E (2011) Auditory cortex encodes the perceptual interpretation of ambiguous sound. J Neurosci 31:1715–1720. CrossRef Medline

Krieger-Redwood K, Gaskell MG, Lindsay S, Jefferies E (2013) The selective role of premotor cortex in speech perception: A contribution to phoneme judgements but not speech comprehension. J Cogn Neurosci 25:2179–2188. CrossRef Medline

Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863–3868. CrossRef Medline

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—connecting the branches of systems neuroscience. Front Syst Neurosci 2:1–28. CrossRef Medline

Lee YS, Turkeltaub P, Granger R, Raizada RD (2012) Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. J Neurosci 32:3942–3948. CrossRef Medline

Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. Psychol Rev 74:431–461. CrossRef Medline

Lotto AJ, Hickok GS, Holt LL (2009) Reflections on mirror neurons and speech perception. Trends Cogn Sci 13:110–114. CrossRef Medline

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M (2007) The essential role of premotor cortex in speech perception. Curr Biol 17:1692–1696. CrossRef Medline

Mesgarani N, Cheung C, Johnson K, Chang EF (2014) Phonetic feature encoding in human superior temporal gyrus. Science 343:1006–1010. CrossRef Medline

Meyer M, Zaehle T, Gountouna VE, Barron A, Jancke L, Turk A (2005) Spectro-temporal processing during speech perception involves left posterior auditory cortex. Neuroreport 16:1985–1989. CrossRef Medline

Miller GA, Nicely PE (1955) An analysis of perceptual confusions among some English consonants. J Acoust Soc Am 27:338–352. CrossRef

Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. J Neurosci 32:14205–14216. CrossRef Medline

Möttönen R, Dutton R, Watkins KE (2013) Auditory-motor processing of speech sounds. Cereb Cortex 23:1190–1197. CrossRef Medline

Munson B (2011) The influence of actual and imputed talker gender on fricative perception, revisited (L). J Acoust Soc Am 130:2631–2634. CrossRef Medline

Padberg J, Disbrow E, Krubitzer L (2005) The organization and connections of anterior and posterior parietal cortex in titi monkeys: do New World monkeys have an area 2? Cereb Cortex 15:1938–1963. CrossRef Medline

Papoutsi M, de Zwart JA, Jansma JM, Pickering MJ, Bednar JA, Horwitz B (2009) From phonemes to articulatory codes: an fMRI study of the role of Broca's area in speech production. Cereb Cortex 19:2156–2165. CrossRef Medline

Peters RW (1963) Dimensions of perception for consonants. J Acoust Soc Am 35:1985–1989. CrossRef

Poeppel D (2003) The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time." Speech Commun 41:245–255.

Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006) Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci U S A 103:7865–7870. CrossRef Medline

Raizada RD, Poldrack RA (2007) Selective amplification of stimulus differences during categorical processing of speech. Neuron 56:726–740. CrossRef Medline

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373. CrossRef Medline

Ryalls J, Zipprer A, Baldauff P (1997) A preliminary investigation of the effects of gender and race on voice onset time. J Speech Lang Hear Res 40:642–645. CrossRef Medline

Sato M, Tremblay P, Gracco VL (2009) A mediating role of the premotor cortex in phoneme segmentation. Brain Lang 111:1–7. CrossRef Medline

Scott SK, McGettigan C, Eisner F (2009) A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. Nat Rev Neurosci 10:295–302. CrossRef Medline

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

Shepard RN (1972) Psychological representation of speech sounds. In: Human communication: a unified view (David E, Denes P, eds), pp 67–113. New York: McGraw-Hill.

Shepard RN, Chipman S (1970) Second-order isomorphism of internal representations: shapes of states. Cogn Psych 1:1–17. CrossRef

Steinschneider M, Nourski KV, Kawasaki H, Oya H, Brugge JF, Howard MA 3rd (2011) Intracranial study of speech-elicited activity on the human posterolateral superior temporal gyrus. Cereb Cortex 21:2332–2347. CrossRef Medline

Szenkovits G, Peelle JE, Norris D, Davis MH (2012) Individual differences in premotor and motor recruitment during speech perception. Neuropsychologia 50:1380–1392. CrossRef Medline

Venables WN, Ripley BD (2002) Modern applied statistics with S. New York: Springer.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7:701–702. CrossRef Medline

Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. Nat Methods 8:665–670. CrossRef Medline

Yuen I, Davis MH, Brysbaert M, Rastle K (2010) Activation of articulatory information in speech perception. Proc Natl Acad Sci U S A 107:592–597. CrossRef Medline

Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946–953. CrossRef Medline