Cerebral Cortex December 2010;20:2958–2970
doi:10.1093/cercor/bhq045
Advance Access publication April 9, 2010

Specialization along the Left Superior Temporal Sulcus for Auditory Categorization

Einat Liebenthal, Rutvik Desai, Michael M. Ellingson, Brinda Ramachandran, Anjali Desai and Jeffrey R. Binder

Department of Neurology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

Address correspondence to Einat Liebenthal, Department of Neurology, 8701 Watertown Plank Road, Medical College of Wisconsin, Milwaukee, WI 53226, USA. Email: einatl@mcw.edu.

The affinity and temporal course of functional fields in middle and posterior superior temporal cortex for the categorization of complex sounds was examined using functional magnetic resonance imaging (fMRI) and event-related potentials (ERPs) recorded simultaneously. Data were compared before and after subjects were trained to categorize a continuum of unfamiliar nonphonemic auditory patterns with speech-like properties (NP) and a continuum of familiar phonemic patterns (P). fMRI activation for NP increased after training in left posterior superior temporal sulcus (pSTS). The ERP P2 response to NP also increased with training, and its scalp topography was consistent with left posterior superior temporal generators. In contrast, the left middle superior temporal sulcus (mSTS) showed fMRI activation only for P, and this response was not affected by training. The P2 response to P was also independent of training, and its estimated source was more anterior in left superior temporal cortex. Results are consistent with a role for left pSTS in short-term representation of relevant sound features that provide the basis for identifying newly acquired sound categories. Categorization of highly familiar phonemic patterns is mediated by long-term representations in left mSTS. Results provide new insight regarding the function of ventral and dorsal auditory streams.

Keywords: auditory cortex, electroencephalograph, fMRI, speech, training

Introduction

A functional segregation for auditory and speech perception along an anterior-to-posterior axis in the temporal cortex was first advanced based on early neurological studies of aphasia (Wernicke 1874; Geschwind and Levitsky 1968). It is further suggested by the highly variable location of activation responses along the anterior–posterior axis of left superior temporal sulcus (STS) in neuroimaging studies using speech–nonspeech comparisons (Binder et al. 2000; Scott et al. 2000; Giraud and Price 2001; Narain et al. 2003; Dehaene-Lambertz et al. 2005; Liebenthal et al. 2005; Obleser et al. 2006; Desai et al. 2008). The anterior–posterior segregation in STS is often interpreted in the context of a functional dissociation between ventral and dorsal streams of auditory processing, oriented anteriorly and posteriorly from Heschl's gyrus (HG) along the temporal lobe, respectively. Electrophysiological studies in primates support a segregation into a ventral "what" stream concerned with auditory object identification based on analysis of spectral and temporal features of sounds and a dorsal "where" stream concerned with auditory object localization based on analysis of sound source location and motion (Rauschecker 1998; Romanski et al. 1999; Rauschecker and Tian 2000; Recanzone 2001), analogous to the pathways postulated in the visual system (Ungerleider and Mishkin 1982). A functional segregation in primates is also supported by the differential pattern of neuroanatomical connections from anterior auditory belt and parabelt regions projecting to anterior temporal lobe and ventrolateral frontal regions and from posterior auditory belt and parabelt regions projecting to posterior temporal and dorsolateral frontal regions (Kaas and Hackett 1999).

Although this evidence strongly suggests a distinction between anterior-ventral and posterior-dorsal auditory processing streams, the specific function and degree of specialization of each stream, particularly in the context of speech perception in humans, remains highly controversial (Belin and Zatorre 2000; Belin et al. 2000; Binder et al. 2000; Hickok and Poeppel 2000; Romanski et al. 2000; Scott et al. 2000; Wise et al. 2001; Middlebrooks 2002; Scott and Johnsrude 2003; Arnott et al. 2004; Hickok and Poeppel 2004; Liebenthal et al. 2005; Hickok and Poeppel 2007). Belin and Zatorre (2000), relying on their finding of a human voice–sensitive area in the middle and anterior STS bilaterally, postulated a dorsal route concerned with the perception of auditory spectrotemporal changes for semantic processing and a ventral pathway concerned with voice identification. Others have proposed that it is the anteriorly oriented ventral pathway that is mainly responsible for the extraction of meaning (Binder et al. 2000; Scott et al. 2000), with intermediate regions in left middle superior temporal lobe just ventral to HG specialized for phonemic perception (Liebenthal et al. 2005). Hickok and Poeppel (2000, 2004, 2007) postulate a dorsal pathway projecting to the temporoparietal boundary region for mapping sound onto articulatory representations and a ventral pathway for mapping sound onto meaning (though unlike in other models, this ventral pathway is oriented posteriorly and projects from the superior temporal gyrus bilaterally to the posterior middle temporal gyrus). Wise and colleagues (Wise et al. 2001) postulate a dorsal stream projecting to the posterior STS with a major role for this region in transient representation of the temporal structure of phonetic sequences and as an interface between phoneme perception and long-term lexical memory. A second more dorsal stream projecting to the temporoparietal junction is hypothesized to interface with the speech motor cortex, similar to the model of Hickok and Poeppel.

The goal of the present study was to test the hypothesis that an important determinant of the differentiation between ventral and dorsal streams of auditory processing in the temporal lobe is the familiarity and level of expertise of listeners with the sounds. To test this hypothesis, we compared the pattern of activation in subregions along the anterior–posterior axis of the left STS during categorization of familiar phonemic patterns and unfamiliar nonphonemic patterns. The nonphonemic patterns were closely matched with the phonemic patterns in terms of their acoustic properties but differed from them in that categories for these patterns were

linguistically irrelevant and were newly learned as part of the study procedures.

This hypothesis was motivated by a hierarchical neuroanatomical view of auditory and speech perception (recently reviewed by Obleser and Eisner 2009), the evidence reviewed above supporting a dissociation between anterior and posterior streams of auditory processing, and theoretical considerations regarding the neural basis of categorical perception (Harnad 1982, 1987). We propose that the categorization of highly familiar sound patterns is mediated by long-term representations in the left middle portion of the STS just ventral to HG (middle superior temporal sulcus [mSTS]), whereas the left STS posterior to HG (posterior superior temporal sulcus [pSTS]) plays a role in transient representation of relevant sound features that provide the basis for identifying newly acquired sound categories. In the left mSTS, neural representations of overlearned auditory patterns such as phonemes or phoneme combinations (syllables) are greatly abstracted from the analog spectrotemporal information in the speech signal, and they retain primarily the information that is invariant within phoneme categories. The analog spectrotemporal information in speech is represented in HG and surrounding auditory cortex and is mapped onto linguistically relevant abstract representations (i.e., phoneme codes) downstream, allowing for efficient retrieval of lexical–semantic information by still higher-order areas. Empirical evidence for the role of the left mSTS in phonemic perception comes from a previous functional magnetic resonance imaging (fMRI) study (Liebenthal et al. 2005) in which familiar speech sounds were found to activate this region more strongly than acoustically matched nonphonemic sounds. Because the phonemic and nonphonemic sounds were matched on spectrotemporal characteristics, the left mSTS activation has been interpreted as due to activation of abstract phoneme codes. In contrast, neural representations of newly learned nonphonemic auditory patterns or unfamiliar phonological sequences must retain some information regarding their spectrotemporal structure throughout the stream of auditory processing because these sounds cannot be associated with learned abstract representations stored in long-term memory. Neural representations for these sounds consist of low-level abstractions of the physical sensory properties of the sounds. Based on prior evidence implicating the left pSTS in the learning of unfamiliar nonnative or distorted speech sounds (Callan et al. 2003; Golestani and Zatorre 2004; Dehaene-Lambertz et al. 2005; Desai et al. 2008), and in transient storage and retrieval of phonological sequences (Hickok and Poeppel 2000; Wise et al. 2001; Indefrey and Levelt 2004; Buchsbaum et al. 2005), we hypothesized that this region transiently stores sensory-based representations of newly learned sound categories for which long-term abstract representations in mSTS have not been formed. These representations can be accessed by neurons in regions concerned with phonological processing and articulatory planning.

Participants in this study were trained to categorize a 7-token continuum of unfamiliar nonphonemic sounds with speech-like (i.e., human voice) acoustic properties into two discrete categories (A and B). This training was designed to enable the development of short-term category representations for the nonphonemic sounds. Participants were also trained with a continuum of speech syllables (/ba/ - /da/) to control for activity related to low-level auditory processing and for nonspecific training effects. Behavioral, fMRI, and event-related potential (ERP) measures were compared before and after training to gain insight into the spatiotemporal organization and the neural mechanisms governing phonemic and nonphonemic categorization in superior temporal cortex. For purposes of discussion, we defined the mSTS as the STS area immediately ventral to HG (Talairach $y = -5$ to -30) and the pSTS as the STS area posterior to HG (Talairach $y = -30$ to -55).

Materials and Methods

Participants

Twenty-five subjects (7 females) participated in the study. They ranged from 21 to 47 years of age (average 27.9) and were all right-handed (Oldfield 1971). Participants were native speakers of General American English, with normal hearing and no neurological symptoms. Data from six participants were excluded from group analysis due to lack of improvement in nonphonemic categorization with training, as indicated by a negative value for the difference in the nonphonemic categorization index (CI) after relative to before training ($PostCI^{NP} - PreCI^{NP} < 0$; five participants) or a negative value for the nonphonemic CI after training ($PostCI^{NP} < 0$; two participants). Note that data from one participant were excluded based on both criteria. Data from two additional participants were excluded from ERP analysis due to excessive artifact contamination. Thus, behavioral and fMRI results are reported from 19 participants and ERP results from 17 participants. Participants gave written informed consent, and the study was sanctioned by the Medical College of Wisconsin Institutional Review Board.

Stimuli and Paradigm

Stimuli consisted of 7-step phonemic (P) and nonphonemic (NP) continua. P tokens were composed of a continuum from /ba/ to /da/. NP tokens were created by spectrally inverting the first formant of the syllables in the phonemic continuum (Liebenthal et al. 2005). P and NP continua were matched on token duration, amplitude, spectrotemporal complexity, and overall formant structure but differed in that the stop-like onsets of the NP items were unfamiliar and not analogous to any English phoneme. Though lacking familiar phonemic information, the NP stimuli retained all the acoustic characteristics (e.g., fundamental frequency, harmonic and formant structure) of a typical human voice. Sounds were delivered binaurally using an Avotec SS-3100 pneumatic audio system at approximately 70 dB, adjusted individually to accommodate differences in hearing and in positioning of the eartips. Sound presentation was controlled using PsyScope software.

In the pretraining session before scanning, participants were first introduced only to the end tokens (1 and 7) of each continuum and briefly practiced to categorize them as /ba/ or /da/ (P continuum) and A or B (NP continuum), using 1 of 2 keys. This practice was performed first with the phonemic and then with the nonphonemic continuum and consisted of 10 trials per continuum with feedback provided after every trial. Participants were then scanned while performing the categorization task with all the tokens in each continuum. Tokens were repeated 10 times in random order in 4 runs alternating between P and NP (with run order counterbalanced between subjects).

In 4 subsequent training sessions occurring over the course of approximately 2 weeks, participants practiced categorizing all 7 tokens in each continuum into the two categories represented by the end points, through gradual introduction of token pairs closer to the category boundary. Table 1 summarizes the 4 steps of the training routine. In Step I, subjects practiced categorizing tokens 1 and 7; in Step II, they practiced tokens 1, 2, 6, and 7; and in Step III, they practiced tokens 1, 2, 3, 5, 6, and 7. In Step IV, categorization of the boundary tokens (3 and 5) was reinforced. The training trials were delivered in blocks of 24, and categorization accuracy was displayed at the end of every block. Each block was preceded by 3 presentations of the tokens trained in that step, in ascending order and accompanied by text on the computer screen identifying the category of each token (e.g., this is "A"). Ninety percent categorization accuracy was required

Table 1
Categorization training routine

Step	Trained tokens		Minimum accuracy (%)
	Category A	Category B	
I	1	7	90 × 1
II	1, 2	6, 7	90 × 1
III	1, 2, 3	5, 6, 7	90 × 3
IV	3	5	90 × 3

Note: In each of the 4 sessions, participants were trained to categorize the phonemic and then the nonphonemic sounds in 4 steps, starting with the end tokens and gradually introducing token pairs closer to the category boundary. In the final step, the boundary tokens were further reinforced. Ninety percent accuracy on a block of 24 trials was required once for steps I and II and three consecutive times for steps III and IV before advancing to the next step. The duration of a typical training session was approximately 1 h.

to proceed to the next step, on one block of trials in Steps I–II, and on 3 consecutive blocks of trials in Steps III–IV. The overall training time varied depending on performance and amounted to approximately 4 h. In the posttraining session, participants were scanned again using the same procedures as in the pretraining session. Delay between the pre- and postscans was 3–4 weeks.

An identification task was selected in this study because we hypothesized that effects of categorization training on discrimination would be secondary to (resultant from) the effects of categorization training on identification and may not be observable within 4 training sessions. We conjectured that effects of categorization training on identification reflect early stages of the formation of new categories, whereas effects of categorization training on discrimination would appear after longer-term exposure to the new sounds.

Behavioral CI

Indexes of categorization performance on P and NP continua (CI^P and CI^{NP}) before and after training were computed for each participant based on the respective identification curves. The index consisted of the beta coefficient in a logistic regression function fitted to individual identification curves. Logistic regression (Hosmer and Lemeshow 2004) fits an S-shaped curve to the data using the maximum likelihood method and generates coefficient estimates for the function that is most likely to describe the observed pattern of data. Under the logistic regression framework, the probability of a certain response (category 1) can be modeled as $P(\text{category 1}) = 1/(1 + e^{-(\alpha + \beta x)})$, where X is the predictor variable (here, the position of the token in the continuum). The coefficient β can be interpreted as the steepness or slope of the S-curve. High values of $|\beta|$ suggest a steep step-like curve characteristic of categorical perception. Low values suggest a more linear or continuously varying response, and values close to 0 indicate a flat response curve or chance performance (Desai et al. 2008; Morrison and Kondaurova 2009). For the linear correlation with functional activation levels, a square root transformation was applied to obtain a more linear increase in the indexes. Negative beta coefficient values were transformed to positive for square root computation and then reassigned a negative value.

Functional Magnetic Resonance Imaging

Images were acquired on a 3.0T GE Excite scanner (GE Medical Systems). T_2^*-weighted, gradient echo, echoplanar images (time echo = 20 ms, flip angle = 80°, number of excitations = 1) were collected using a clustered sequence (Edmister et al. 1999) with 2-s image acquisition time and 5-s intervals between images (Fig. 1). Sound stimuli were presented during the intervals at 3.5 s before the onset of each image to avoid their perceptual masking by the scanner acoustic noise and to maximize contribution of their blood oxygen level–dependent signal to the image. The functional images were constructed from 36 axially oriented contiguous slices with 3.44 × 3.44 × 3.5 mm^3 voxel dimensions, covering the whole brain. A total of 640 images were acquired, consisting of 140 images in each of the 4 experimental conditions (prephonemic, PreP; pre-nonphonemic, PreNP; postphonemic, PostP; post-nonphonemic, PostNP), split into 2 runs per condition, and 80 silence baseline images (10 images in each of the 4 pre- and 4 posttraining runs, randomly interspersed within the run). High-resolution anatomical images of the entire brain were obtained using a 3D spoiled gradient echo sequence ("SPGR"; GE Medical Systems), with 0.86 × 0.86 × 1 mm voxel dimensions.

Within-subject analysis was performed using AFNI (Cox 1996) and consisted of spatial coregistration of functional images within and between the sessions, deconvolution, and voxelwise multiple linear regression (Ward 2001) with reference functions representing the 4 experimental conditions (PreP, PreNP, PostP, and PostNP). Six motion parameters were also included as covariates of no interest. Individual maps were computed for the contrasts between stimulus conditions (PreP–PreNP; PostP–PostNP), training conditions (PostNP–PreNP; PostP–PreP) and for the interaction between the conditions ([PostNP–PreNP]–[PostP–PreP]). Individual data were smoothed with a Gaussian filter of 4 mm full-width at half-maximum. Functional images were aligned to the anatomical images using the align_epi_anat.py script in AFNI (Saad et al. 2009). Anatomical and functional images were then projected into standard stereotaxic space (Talairach and Tournoux 1988) using the auto_tlrc function in AFNI with the Colin N27 brain in Talairach space (TT_N27) as the reference anatomical template. In a random-effects analysis, coefficient maps were contrasted against a constant value of 0 to create group maps of z-scores. The group maps were thresholded at voxelwise $P < 0.025$. Clusters smaller than 414 μL were removed to achieve a corrected mapwise $P < 0.05$ as determined by Monte Carlo simulations (Ward 2000). To achieve increased sensitivity in the temporal lobes, masks containing the superior, middle, and inferior temporal gyri, HG, and the supramarginal gyrus were created for left and right hemispheres, using the Macro Label Atlas in AFNI. In these regions, a cluster size criterion of 240 μL was applied to achieve a corrected $P < 0.05$.

Mean activation levels relative to the rest baseline within the left mSTS and left pSTS were computed and entered into analyses of variance (ANOVAs) with factors of training, stimulus and region of interest (ROI). The ROIs for these analyses were defined functionally as spheres with an 8-mm radius placed at the peak of the positive activation in the left STS in PreP–PreNP (Talairach $x = -50$, $y = -23$, $z = -3$, in mSTS) and in PostNP–PreNP (Talairach $x = -56$, $y = -46$, $z = 2$, in pSTS). The left mSTS and pSTS ROIs were selected as the superior temporal regions most responsive to familiar phonemes and to newly acquired speech-like sounds, respectively. The purpose of the ROI analysis was to assess the direction and size of main effects and interactions of the factors of training and stimulus on the level of activation in each region.

A search for regions showing sensitivity to changes in individual behavioral categorization performance was conducted using exploratory voxelwise Pearson's correlation analyses between each fMRI stimulus contrast (PreP–PreNP, PostP–PostNP) and the corresponding CI stimulus contrast ($PreCI^P$–$PreCI^{NP}$, $PostCI^P$–$PostCI^{NP}$) and also between each fMRI training contrast (PostP–PreP, PostNP–PreNP) and the corresponding CI training contrast ($PostCI^P$–$PreCI^P$, $PostCI^{NP}$–$PreCI^{NP}$). In order to search for regions showing sensitivity to individual variation in both phonemic and nonphonemic categorization performance, a binary map (Fig. 5) was constructed from the conjunction of maps of regions showing a positive correlation between the level of activation in PostP and CI^P and a positive correlation between the level of activation in PostNP and CI^{NP}. Both the fMRI-CI correlation maps used to construct the conjunction map were thresholded at $P < 0.05$. For the group binary conjunction map, the centers of mass of the clusters in the left–right, anterior–posterior, and superior–inferior directions are reported (Supplementary Table 2).

Finally, correlation analyses for the same stimulus and training contrasts as for the fMRI-CI correlations were also conducted between the level of fMRI activation and the ERP P2 peak amplitude, in order to search for covariation in individual neurophysiological measures.

Event-Related Potentials

ERP data were collected continuously during fMRI using the 64-channel Maglink system at 500-Hz digitization rate, band-limited from 0 to 100 Hz (DC mode), and analyzed using the Scan 4.3 software

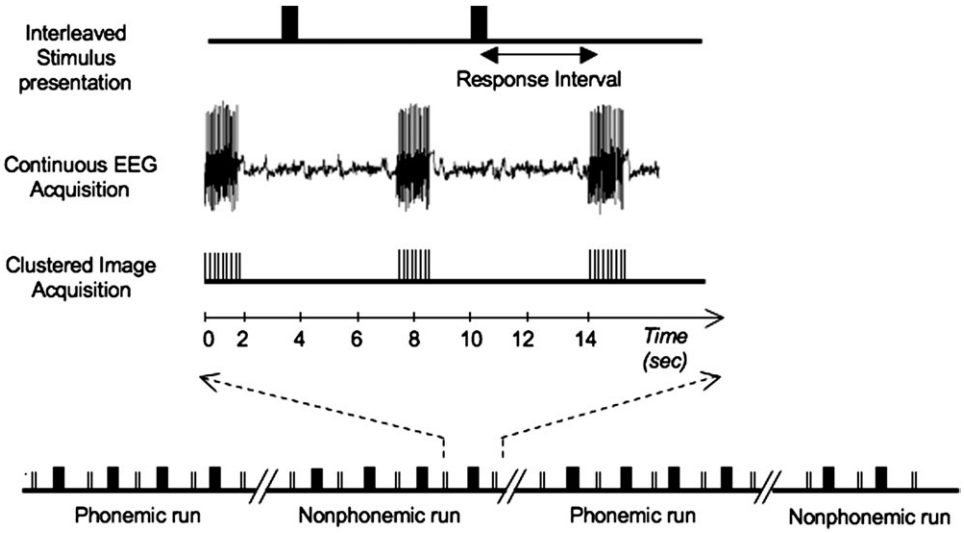

Figure 1. Experimental paradigm: Sound stimuli were presented in 4 runs alternating between the phonemic and nonphonemic conditions, in each of the 2 scanning sessions (pre- and posttraining). Image acquisition was clustered to 2 s at 7-s periods. Electroencephalograph was acquired continuously. Sounds were presented during the intervals between image acquisitions.

package (Compumedics Neuroscan). Electrode sites conformed to the 10-20 system with CPZ serving as the reference and two other sites designed as the electrooculogram and electrocardiogram channels. Offline, data from each participant were bandpass-filtered from 0.3 to 30 Hz, treated for ballistocardiogram artifact removal, and epoched from −100 to 500 ms from stimulus presentation. Epochs with artifacts larger than ±100 μV were removed. The mean number of epochs accepted per condition was 97 and did not differ between the conditions. Accepted epochs were sorted and averaged according to condition and then digitally re-referenced to the mastoids. Grand-average ERP waveforms in each condition were constructed by averaging the individual waveforms.

Peaks of the N1, P2, and P3 components in individual data were automatically identified as the largest negativity between 100 and 160 ms at FZ for N1, the largest positivity between 185 and 275 ms at FZ for P2, and the largest positivity between 300 and 400 ms at PZ for P3. Peak amplitude values of N1 and P2 were measured at frontal and frontocentral electrodes (FZ, F1, F2, F3, F4, FC3, and FC4) and of P3 at parietal electrodes (PZ, P1, and P2). These values were entered into an ANOVA with factors of stimulus, training and electrode as a repeated measure. Pointwise paired *t*-tests were conducted at −100 to 500 ms for the 4 contrasts of interest (PreP–PreNP, PostP–PostNP, PostP–PreP, and PostNP–PreNP) at all electrode sites. At least 11 consecutive data points (20 ms) in which the *t*-test exceeded $P < 0.05$ were required for statistical significance of the difference potentials (Guthrie and Buchwald 1991). Because P2 showed effects of stimulus and training, correlation analyses between individual P2 peak amplitude values in the stimulus and training contrasts and the corresponding CI measures were also conducted.

Current density reconstruction (CDR) of the grand-average ERP data in PostP and PostNP in the period 180–300 ms after sound presentation were performed in Curry 5 (Compumedics Neuroscan). These two conditions were selected for CDR because their average noise level was lowest and similar (0.262 and 0.269 μV, respectively). A realistic three-compartment boundary element model constructed from the Montreal Neurological Institute (MNI) ICBM152 brain and implemented as a reference brain in Curry 5 was used as a volume conductor. Sensor locations were determined by label-matching. The standardized low resolution brain electromagnetic tomography Minimum Norm Least Squares approach (Pascual-Marqui 2002) was used for distributed source modeling. Sources with current strength below 75% of the maximum were clipped in the display (Fig. 8). To enable comparison with peaks of fMRI activation, the coordinates of CDR activation peaks were transformed from the internal Curry coordinates into SPM99 MNI coordinates and then into Talairach coordinates, using affine transformations provided in Curry.

Results

Behavioral

In general, training for approximately 4 h over 4 separate sessions resulted in a similar level of improvement in the categorization of both P and NP sounds, but P categorization was near perfect posttraining, whereas NP categorization accuracy was lower (Fig. 2*a*). This was observed as lower average categorization accuracy, lower average CI (computed from the slopes of logistic regression functions fitted to individual identification curves) and longer average response time (RT) for the NP sounds.

Pretraining, subjects classified P sounds into 2 discrete categories composed of tokens 1–3 (/ba/) and 5–7 (/da/), as evident from the step-like identification curve. Posttraining, the average identification accuracy increased from 87% to 98% ($t = -3.02$, $P < 0.01$) and the average identification RT decreased from 741 to 578 ms ($t = 4.07$, $P < 7 \times 10^{-4}$). The slope of the identification curve across the category boundary (as measured by the CI for P, CI^P) increased posttraining from 1.66 to 2.37 ($t = 2.42$, $P < 0.03$). Across the NP continuum pretraining, perception gradually changed from the A to the "B" end with no defined category boundary in the identification or RT curves. After training, the average identification accuracy of the two categories (defined by the training procedure as A for tokens 1–3 and B for tokens 5–7) increased from 62% to 88% ($t = -6.98$, $P < 2 \times 10^{-6}$), and the average RT decreased from 964 to 787 ms ($t = 4.18$, $P < 6 \times 10^{-4}$). The slope of the identification curve across the category boundary (as measured by the CI for NP, CI^{NP}) increased posttraining from 0.40 to 1.30 ($t = 4.38$, $P < 4 \times 10^{-4}$).

ANOVA with factors of stimulus (P, NP) and training (Pre, Post) showed main effects of both factors on the CI ($F_{1,72} = 34.24$, $P < 10^{-4}$; $F_{1,72} = 16.40$, $P < 10^{-4}$, respectively) with no significant interaction ($F_{1,72} = 0.23$, $P < 0.6$), suggesting that the extent of categorization improvement with training in both continua was similar (Fig. 2*b*).

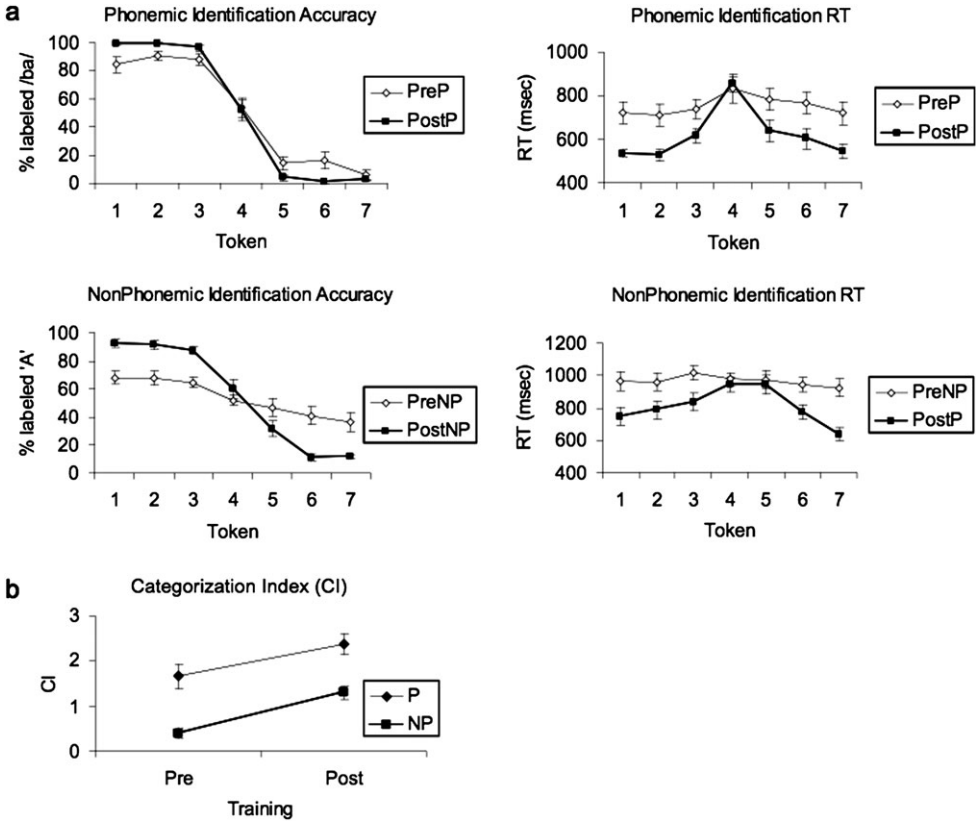

Figure 2. (*a*) Tokenwise categorization and RT curves for the phonemic and nonphonemic continua prior to training (PreP and PreNP, respectively) and following training (PostP and PostNP, respectively). RT was calculated from the onset of the sound. (*b*) Effect of training on the CI for P and NP.

Functional Magnetic Resonance Imaging

In the left temporal lobe, the mSTS was activated only during categorization of P sounds, whereas categorization of NP sounds engaged the pSTS.

Group contrast maps showing the effects of stimulus (PreP–PreNP, PostP–PostNP; Fig. 3*a*) revealed stronger activation for P compared with NP in left mSTS (peak at Talairach $x = -50$, $y = -23$, $z = -3$) prior to training. Additional regions in the left inferior frontal and parietal, right temporal, and cingulate cortex were also activated. Following training, the P–NP difference map was dominated by negative activation in bilateral and distributed frontal, parietal, and posterior temporal areas indicating stronger engagement of these regions during NP categorization. Only the angular gyri (AGs) and the right superior frontal gyrus were positively activated in the stimulus contrast posttraining.

The effects of training (PostP–PreP, PostNP–PreNP; Fig. 3*b*) were observed in NP as increased activity predominantly in left posterior STS and middle temporal gyrus (peak at Talairach $x = -55$, $y = -49$, $z = -10$), with smaller foci of increased activation in left inferior frontal cortex and bilateral parietal cortex. In P, the effect of training was observed as decreased activity in distributed bilateral frontal, parietal, and posterior superior temporal regions indicating stronger activation pretraining in these regions, consistent with the PostP–PostNP map.

The compound interaction ([PostNP–PreNP]–[PostP–PreP]; Fig. 3*c*) was intended to identify the specific effects of categorization training with the unfamiliar NP stimuli, over and beyond the more general effects of categorization training observed with the familiar P sounds. In order to identify the regions more strongly engaged in NP categorization after training, the interaction map was masked to show only those voxels displaying positive activation ($P < 0.05$) in PostNP–PreNP (and not the voxels showing positive activation in the interaction map due to negative activation in PostP–PreP). The masked interaction contrast revealed a cluster in the left pSTS (peak at Talairach $x = -59$, $y = -46$, $z = 2$) overlapping with the dorsal portion of the left posterior temporal cluster revealed in PostNP–PreNP. Additional clusters were observed in left inferior frontal gyrus (IFG), left insula, and right superior parietal cortex. Supplementary Table 1 lists all the activation foci in the 5 experimental contrasts shown in Figure 3.

To further examine the role of left anterior and posterior superior temporal regions in phonemic categorization and in the learning of new auditory categories, an ROI analysis was conducted. The left mSTS and pSTS ROIs were defined functionally as the superior temporal regions most responsive to familiar phonemes (derived from PreP–PreNP) and to newly acquired speech-like sounds (derived from PostNP–PreNP), respectively. The mean activation in each ROI in the 4 experimental conditions (relative to baseline) is shown in Figure 4. A three-way ANOVA of mean activation with factors of stimulus (P, NP), training (Pre, Post), and ROI (left mSTS, left pSTS) showed that the three-way interaction between these factors was not significant. However, there was an interaction between stimulus and training due to a decrease in mean activation with training for P and an increase for NP ($F_{1,144} = 6.195$, $P < 0.01$). Two-way ANOVA, performed separately for each ROI, with factors of stimulus and training showed a main

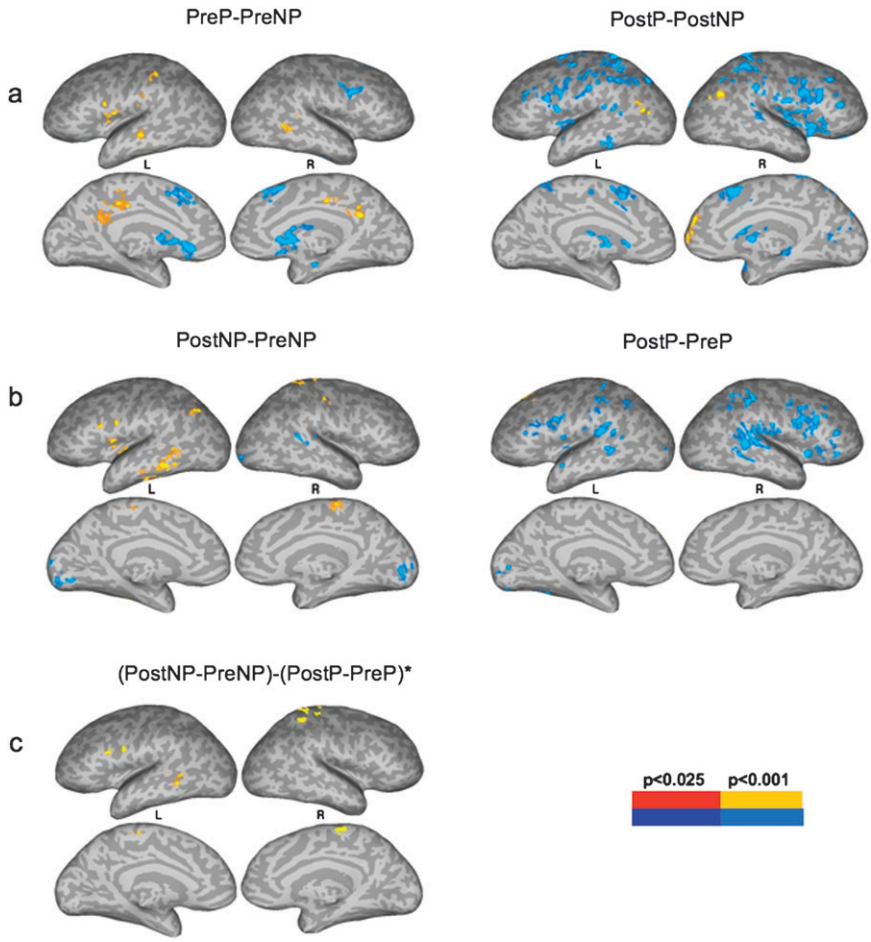

Figure 3. Group fMRI contrast maps showing the effects of stimulus type before and after training (*a*), the effects of training status for P and NP (*b*), and the interaction between stimulus and training (*c*), overlaid over an inflated template of the N27 brain. *The interaction map was masked to show only voxels displaying positive activation ($P < 0.05$) in PostNP–PreNP.

effect of stimulus with stronger activation in left mSTS for P compared with NP ($F_{1,72} = 7.36$, $P < 0.01$) and an interaction in left pSTS whereby the mean activation decreased for P and increased for NP with training ($F_{1,72} = 3.92$, $P < 0.05$). Two-way ANOVA with factors of training and ROI showed an effect of ROI on mean activation for NP whereby the activation was stronger in left pSTS compared with left mSTS ($F_{1,72} = 5.82$, $P < 0.02$). Two-way ANOVA with factors of stimulus and ROI showed a main effect of stimulus whereby mean activation pretraining was higher for P ($F_{1,72} = 6.72$, $P < 0.01$). Other effects in these analyses were not significant ($P > 0.06$).

Voxelwise exploratory correlation maps between individual levels of activation and the behavioral CI in the two stimulus contrasts and the two training contrasts all revealed negative activation in the left IFG or insula, consistent with a stronger fMRI-CI correlation in this region in pretraining compared with posttraining conditions and for NP compared with P sounds. However, these maps did not reveal significant activation in the temporal lobe ROIs in this study. A similar result was obtained in the voxelwise correlation analyses between individual levels of activation and the peak amplitude of the ERP P2 component in the stimulus and training contrasts. Negative activation was observed in PostNP–PreNP in the left insula and the superior temporal plane, bilaterally, but there were no regions of significant positive activation in the temporal lobes. Finally, the conjunction map of regions showing a positive correlation between their level of activation and both the phonemic and the nonphonemic posttraining CIs revealed positive activation in the left parietal cortex (superior parietal lobule [SPL] and supramarginal gyrus) and the right pSTS (Fig. 5 and Supplementary Table 2).

Event-Related Potentials

Grand-average auditory cortical ERPs elicited in all the stimulus and training conditions (Fig. 6) displayed a typical sequence of components composed of a large frontocentral negativity peaking at 140 ms (N1), a frontocentral positivity peaking at 224 ms (P2), and a parietal positivity peaking at 358 ms (P3). The amplitude of P2 elicited by NP sounds increased with training. The P2 response was consistent with left temporal generators in a location more posterior for NP compared with P sounds.

ANOVA of individual peak amplitude measurements with factors of training, stimulus, and frontocentral electrode sites as a repeated measure revealed a trend for an effect of training on P2 peak amplitude ($F_{1,48} = 3.57$, $P < 0.065$). Pointwise *t*-tests of individual responses at −100 to 500 ms in all active electrodes contrasting the relevant pairs of experimental conditions show that the amplitude of P2 was smaller in PreNP compared with PreP (Fig. 7, left panel). The amplitude of P2 was also smaller in

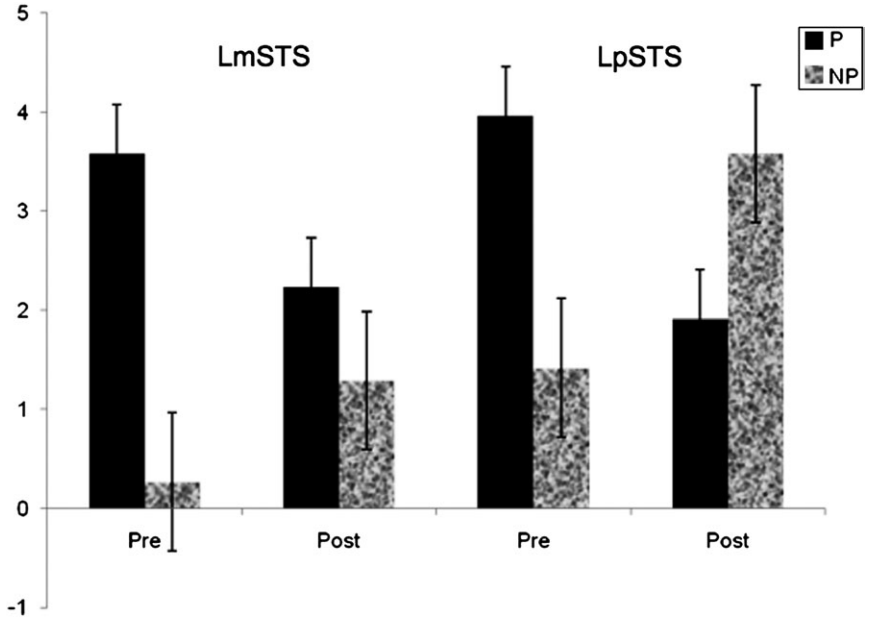

Figure 4. Mean activation relative to baseline in left middle superior temporal sulcus (LmSTS) and left posterior superior temporal sulcus (LpSTS) before (Pre) and after (Post) training to categorize the P and the NP continua.

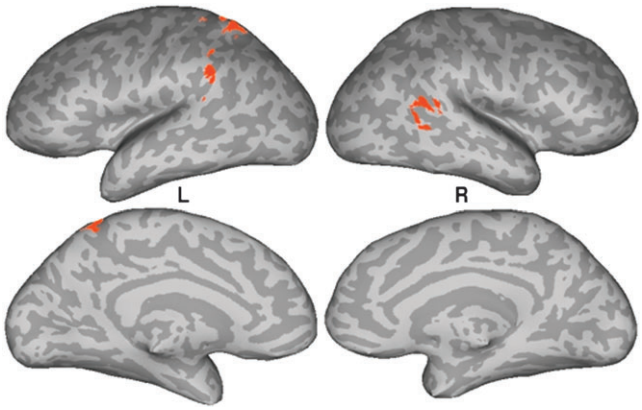

Figure 5. Binary fMRI conjunction map of areas showing a positive correlation between the level of activation in PostP and the phonemic CI^{P} and a positive correlation between the level of activation in PostNP and the nonphonemic CI^{NP}. Both the PostP–CI^{P} and PostNP–CI^{NP} correlation maps, which were used to construct the conjunction map, were thresholded at $P < 0.05$.

PreNP compared with PostNP (Fig. 7, right panel). There were no significant amplitude differences in the PostP–PostNP and PostP–PreP contrasts. There were also no significant effects of stimulus or training on the peak amplitudes of N1 or P3 or in the pointwise comparisons at latencies outside the P2 range.

Correlation analyses between the P2 peak amplitude and CI were also conducted for the stimulus and training contrasts. Results revealed a positive relationship between P2 and CI in PostNP–PreNP ($r = 0.44$, $P < 0.04$) but not in PostP–PreP ($r = -0.17$, $P < 0.26$) and a significant interaction between the nonphonemic and phonemic training contrasts ($r = 0.44$, $P < 0.04$). The correlations in the stimulus contrasts were not significant.

CDR of the grand-average ERP waveforms during the P2 time range (180–300 ms) in PostP and PostNP revealed left-lateralized foci in the temporal lobe with largest current strength around 230 ms in both conditions (Fig. 8). The CDR peak activations were generally more posterior than the corresponding fMRI peaks in left mSTS and pSTS, but the peak in PostNP (Talairach coordinates $x = -62$, $y = -56$, $z = 0$) fell 20 mm posterior to that in PostP (Talairach coordinates $x = -68$, $y = -36$, $z = 5$), similar to the fMRI pattern. Other local maxima in each condition are detailed in Supplementary Table 3.

Discussion

Participants were able to categorize the phonemic but not the nonphonemic sounds prior to training, consistent with previous work using the same stimuli (Liebenthal et al. 2005). Training improved categorization performance with both sound types, to a near-perfect level with the syllables and to a lower level with the nonphonemic sounds. A lower level of performance was also reported in prior categorization training studies for nonnative compared with native sounds, even when using intensive high-variability training procedures over prolonged training periods (Lively et al. 1993; Callan et al. 2003; Golestani and Zatorre 2004). This difficulty in attaining a "native" level of performance through training with nonphonemic or nonnative sounds is consistent with the idea that categorization of sounds that are learned early in development and are constantly reiterated through frequent usage, such as native speech syllables, relies on a dedicated neural mechanism that is distinct from that mediating the categorization of other sounds that are learned later in life or do not carry the same communicative value.

Phonemic Categorization

The increased activity in left mSTS for phonemic compared with nonphonemic sounds pretraining implicates this region in phonemic perception, consistent with prior studies showing activation in this region for syllables, words, and sentences over a variety of control sounds (Binder et al. 2000; Scott et al. 2000; Giraud and Price 2001; Jancke et al. 2002; Davis and Johnsrude 2003; Desai et al. 2005; Liebenthal et al. 2005; Obleser et al.

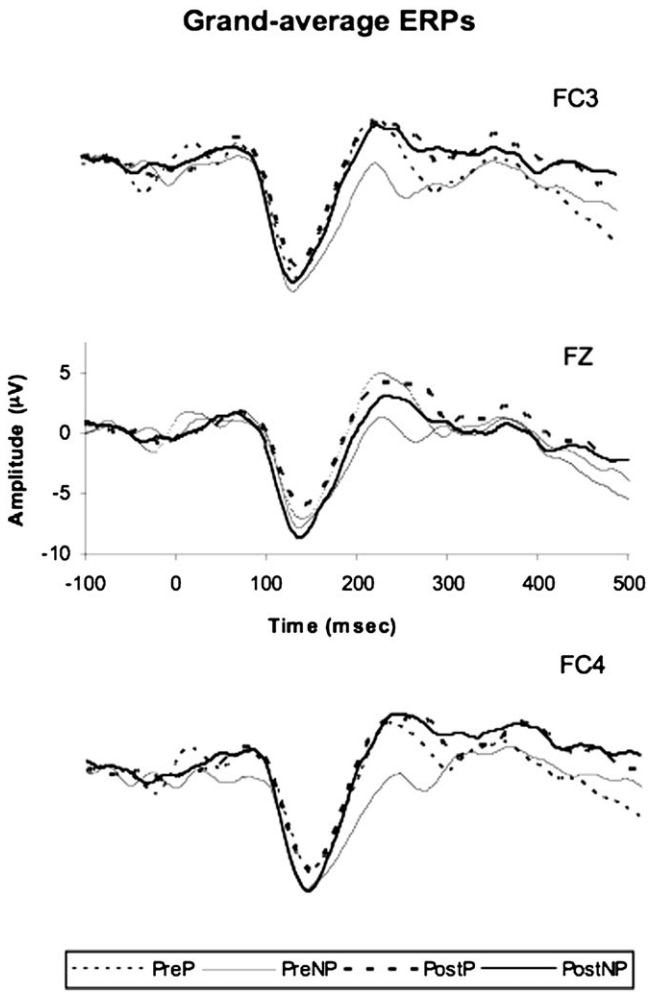

Figure 6. Grand-average ERPs ($N = 17$) at left (FC3), midline (FZ), and right (FC4) frontal electrode sites, during identification of phonemic and nonphonemic tokens before (PreP and PreNP, respectively) and after (PostP and PostNP, respectively) training.

2006). The left mSTS is specifically activated when contrasting sublexical phonemes with acoustically comparable nonphonemic sounds (this study; Liebenthal et al. 2005). The nonphonemic sounds used in this and our prior study are similar to glottal stops that are phonemic in some languages (though not in English). These sounds could conceivably be produced by a vocal tract, and they retain a human voice quality. However, the pattern of activation observed here in mSTS differs from that associated with voice recognition (Belin et al. 2000; Belin and Zatorre 2003) in that it is left lateralized and does not extend to anterior portions of the STS. Thus, the left mSTS appears to be truly sensitive to the phonemic properties of speech (i.e., the set of sounds that compose the native language) rather than to the prephonemic-physical or lexical–semantic properties of speech. Nevertheless, further research is warranted to examine whether this region is more broadly tuned and responsive to other familiar sound categories in addition to speech phonemes.

Additional areas were activated in PreP–PreNP, including the left IFG and anterior parietal lobe, the right pSTS, and the middle and posterior cingulate regions. These areas were also activated more strongly for NP sounds in other contrasts and therefore appear to play a nonspecific role in phonemic perception. The left IFG was activated more strongly in PostNP relative to PreNP and relative to PostP and in PreP relative to PostP. This region has been implicated in phonological processing and phonemic categorization, with the level of activation increasing as a function of categorization difficulty (Binder et al. 2004; Blumstein et al. 2005; Myers et al. 2009). However, portions of the left inferior frontal cortex have also been shown to be responsive during auditory (nonphonetic) decision making (Locasto et al. 2004; Burton and Small 2006). The pattern of left IFG activation in the present study is consistent with a role for this region in the temporary storage and comparison of complex sounds for decision making. The left anterior parietal cortex (postcentral gyrus and sulcus) was similarly activated more strongly in PostNP relative to PostP and in PreP relative to PostP. The right pSTS was also activated in the conjunction correlation map with both the phonemic and nonphonemic posttraining CIs, suggesting that activation in this region was generally sensitive to auditory categorization. While the contribution of these latter regions to phonemic perception remains unclear, it also appears to be domain nonspecific.

The effect of training on phonemic categorization (PostP–PreP) was observed as a reduction in activity in a broad network of posterior temporal, parietal, and frontal regions, bilaterally. There was little change in left middle temporal regions implicated in phonemic perception. Activation in parietal and frontal regions is associated with the level of executive control imposed by cognitive tasks, including perceptual difficulty, attentional demands, working memory load, and response selection difficulty (Petersen et al. 1998; Duncan and Owen 2000; Culham and Kanwisher 2001; Binder et al. 2004). The reduction in activation in those regions with training (as well as in PostP–PostNP) may reflect the reduced demands on executive control as task performance becomes more skilled and automated and requires less monitoring (Poldrack 2000). The decrease in activation in right temporoparietal cortex with categorization training (also observed with the nonphonemic sounds) could reflect decreased attention to the local spectral properties of the sounds due to the reinforcement of their global categorical properties with training (Fink et al. 1996; Brechmann and Scheich 2005; Geiser et al. 2008). The reduction in activation of left temporoparietal cortex could reflect lesser reliance on phonological or prearticulatory codes to categorize the speech sounds posttraining, consistent with the implication of these regions in a sensory-motor circuit (Hickok and Poeppel 2004; Buchsbaum et al. 2005). There was also a reduction in activation in the left pSTS, and this finding is discussed further in the next section. Importantly, the lack of significant change in left mSTS with phonemic categorization training is consistent with the idea that the performance improvement induced by training was achieved primarily through more efficient executive control and diminished involvement of posterior temporal pathways, rather than by changes in the level of activation of the left middle temporal regions specializing in phonemic perception.

One region that was positively activated in PostP–PostNP was the AG, bilaterally. The AG is strongly implicated in lexico-semantic processing (see Binder et al. 2009 for a review). Recent evidence also suggests that the AGs are important for skilled reading and that learning to read strengthens the connectivity between the left and right AG (Carreiras et al.

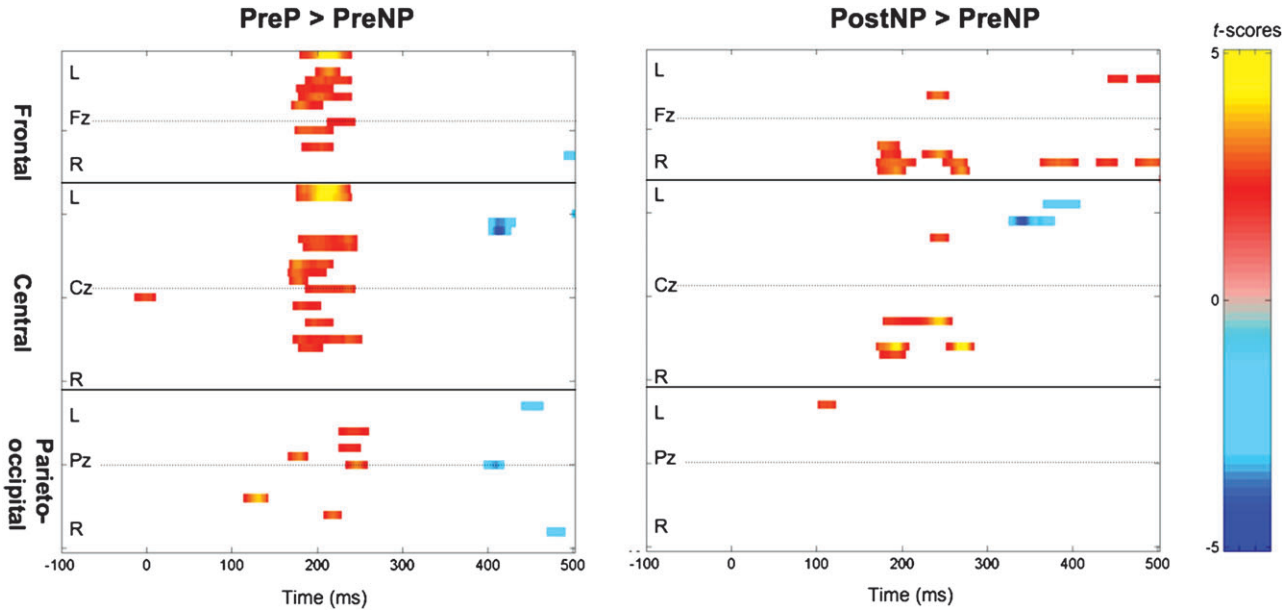

Figure 7. Pointwise *t*-score maps from 60 electrodes showing the effect of stimulus pretraining (PreP–PreNP; left panel) and the effect of training on nonphonemic categorization (PostNP–PreNP; right panel). The electrodes are grouped according to their anterior–posterior position on the scalp from frontal (top third) to central (middle third) and parieto-occipital (bottom third) sites. Within each group of sites, electrodes are sorted according to their lateral position from left (L, upper half) to right (R, lower half). The midline electrode in each group is also indicated (Fz, Cz, Pz). Maps are thresholded at *t*-scores corresponding to $P < 0.05$ sustained for at least 11 consecutive data points (see Materials and Methods for further details).

2009). The present finding suggests that with categorization training, listeners developed a stronger tendency to associate the P sounds with words or meanings compared with the NP sounds. It also raises the possibility that one mechanism by which the AG contributes to skilled reading is through its engagement in overlearning of phonemic categories.

Unfamiliar Sound Categorization

In contrast to phonemic categorization training, which engaged primarily the left mSTS, learning to categorize the nonphonemic sounds engaged the left pSTS with little change in left mSTS. This was evident from the positive activation observed in left pSTS in the NP (but not P) training contrast (Fig. 3*b*) and in the voxelwise (Fig. 3*c*) and mean activation in ROI (Fig. 4) analyses of the interaction between these contrasts showing an increase in activation with training for NP and a decrease for P in left pSTS (but not in mSTS). Taken together, these results implicate the left pSTS in the learning and neural representation of unfamiliar sounds. Nevertheless, it is interesting that the left pSTS was also activated for P sounds (Fig. 4). This finding may reflect the fact that the particular instances of P sounds used in the study, despite representing highly familiar phonemic categories, were not initially familiar to the listeners in terms of their specific acoustic (indexical) properties. It is possible that perception of the P sounds, especially in the context of the sublexical categorization task, initially required more reliance on their acoustic properties, thereby engaging the left pSTS (in addition to the left mSTS). However, contrary to the left pSTS activation for NP sounds, the activation for P sounds was reduced after training (Figs. 3*b* and 4), consistent with the alleged role of this region in the categorization of newly learned sounds.

The left pSTS is also activated during the discrimination or identification of sinewave speech analogs that have unfamiliar and peculiar acoustic properties (Dehaene-Lambertz et al. 2005; Mottonen et al. 2006; Desai et al. 2008), consistent with the above interpretation of the left pSTS activation for P in the present study. In another recent study of auditory categorization training on artificial nonspeech sounds using a video game, Leech et al. (2009) found that the level of activation in this region increased in proportion to the individual improvement in categorization performance. These authors suggested that the increased expertise in categorization of nonspeech sounds prompted a speech-like pattern of activation in this region. However, a condition of training with speech sounds that would allow testing this hypothesis was not included in that study. The present results, showing an increase in left pSTS activation for NP and a decrease for P after categorization training, suggest that the left pSTS is engaged specifically during the categorization of novel sounds, whether they are speech or not.

The voxelwise fMRI-CI correlations in the training and stimulus contrasts, designed to identify changes in activation that are related to individual differences between the phonemic and nonphonemic CIs or changes in the indexes induced by training, did not reveal activation in left temporal regions. The left frontal regions negatively activated in these correlations may reflect the greater executive demands imposed by nonphonemic categorization posttraining and by phonemic categorization pretraining and not changes in auditory categorization performance per se (similar to the effects of training in frontal cortex discussed earlier). One plausible cause for this result is that the degree of individual variation in the difference of behavioral indexes in this study was relatively small and therefore limited the ability to reliably detect related changes in fMRI activation. In fact, we decided to exclude data from participants who showed no training-induced improvement in categorization performance, in order to increase the sensitivity of the group analysis to training effects. In support of this explanation, in another training study in which data from participants performing below chance level

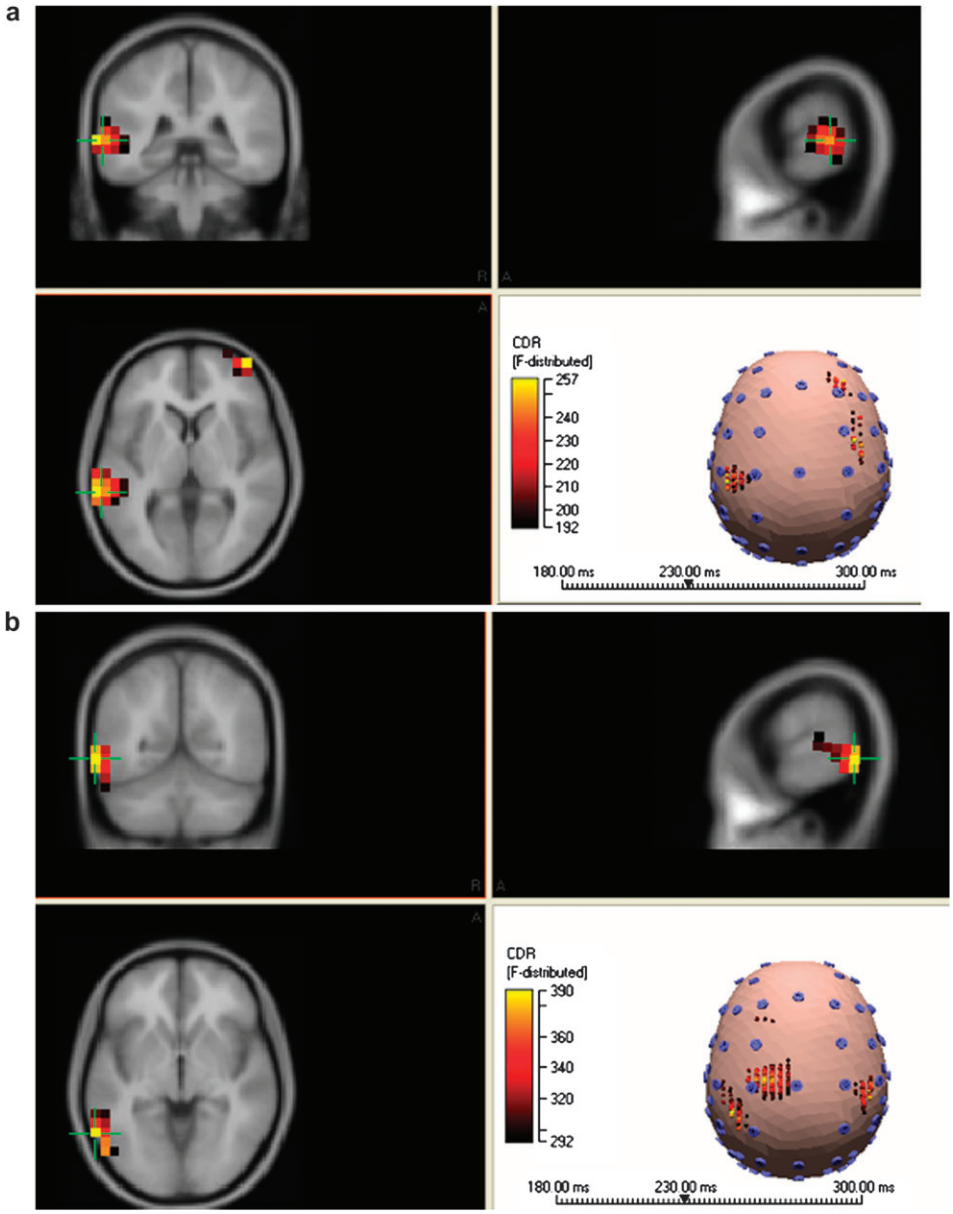

Figure 8. Current density reconstruction (CDR) of grand-average ERPs in PostP (*a*) and PostNP (*b*) at 230 ms, displayed on coronal, sagittal, and axial slices of the MNI ICBM152 template brain (upper quadrants and left lower quadrant) as implemented in Curry 5 (Compumedics Neuroscan). The cursor is positioned near the peak of the activation in the left superior temporal cortex. Bottom right quadrant: CDR shown within the standardized boundary element model volume conductor computed from the template brain (top view, transparent rendering). Activation below 75% of the maximum is clipped.

after training were included, a positive correlation with improvement in categorization performance was obtained in the left pSTS but there were no group training effects (Leech et al. 2009).

The conjunction correlation approach was more inclusive in that it searched for brain regions displaying sensitivity to individual changes in both phonemic and nonphonemic CIs, rather than to individual changes in the difference between the indexes. The conjunction map of regions showing sensitivity to individual variation in both phonemic and nonphonemic CIs revealed positive activation in the left SPL, left supramarginal gyrus (SMG), and right pSTS, suggesting that these regions were more strongly activated in individuals with stronger phonemic and nonphonemic categorization performance. A similar result was obtained in a previous study from our group, in which activation in left SMG was found to be correlated with individual improvement in categorization performance of phonemic and of nonphonemic sinewave replicas (Desai et al. 2008). This region has also been associated with the learning of nonnative phonetic contrasts (Golestani et al. 2002; Callan et al. 2003; Golestani and Zatorre 2004) and with phonological processing of visual words (Xu et al. 2002; Katzir et al. 2005). The left SMG has been suggested to be part of a multimodal dorsal stream for sensory-motor integration (Hickok and Poeppel 2007). Increased activation in left SMG during the learning of new sounds could reflect the acquisition of auditory–articulatory mappings as a means of learning.

The left pSTS has been proposed to serve as a short-term memory buffer for phonological sequences and as an interface between sound perception and sound rehearsal (Wise et al.

2001; Buchsbaum et al. 2005; Jacquemot and Scott 2006; Scott et al. 2006; Buchsbaum and D'Esposito 2008; Obleser and Eisner 2009). This region is implicated in the learning of new sounds (Dehaene-Lambertz et al. 2005; Mottonen et al. 2006; Desai et al. 2008; Leech et al. 2009). The present results go a step further in suggesting that the left pSTS plays a specific role in extracting and representing the relevant (trained) sound features that provide the basis for categorizing newly acquired sounds. We hypothesize that the degree of abstraction from the analog sequential spectrotemporal information of sounds in this region is relatively small, such that these representations essentially reflect the range of variation in spectrotemporal properties in the specific set of trained sounds. The left pSTS may feed into other left parietal regions including the SMG, which play a role in multimodal sensory-motor integration.

Temporal Course of Categorization

Similar to the left mSTS activation, the P2 auditory–evoked response was larger for P compared with NP sounds pretraining, with no significant effect of training. Similar to the left pSTS activation, the P2 response to NP sounds increased with training. Individual measures of the increase in P2 peak amplitude with training were also correlated with individual measures of the increase in NP categorization accuracy. Current source density reconstruction suggested that the P2 source location for both sound types was in superior temporal cortex posterior to HG, as previously described for the P2 response peaking around 220 ms (Verkindt et al. 1994; Godey et al. 2001). Also consistent with the fMRI activation, the P2 response in left temporal cortex to NP sounds was more posterior by approximately 20 mm relative to the response to P sounds, though the location of the CDR peaks was generally more posterior than that of the fMRI peaks. The source localization results are admittedly limited by the low spatial resolution of ERP. The difference between the location of fMRI and CDR peaks could also be due at least in part to inaccuracies in the affine transformation from Curry to Talairach space, related to misregistration between the different reference brains. However, the distance measurement between CDR peaks in left mSTS and pSTS is not affected by this transformation because it is conducted within the original Curry coordinate system. Despite the inherent limitations of current source density reconstruction, the similar patterns of left temporal fMRI and P2 dependence on stimulus type, training level, and categorization performance together support the idea that P2 reflects the activation of neural representations of sound categories in both left mSTS and left pSTS. The ERP peak locations further provide tentative converging evidence for an anterior–posterior segregation in the temporal cortex, akin to that suggested by the fMRI results.

An increase in P2 following discrimination or identification training was previously described for syllables (Tremblay et al. 2001; Reinke et al. 2003; Sheehan et al. 2005) and tones (Bosnyak et al. 2004) and in trained musicians compared with nonmusicians during passive listening (Shahin et al. 2003, 2005). The reported source location of the P2 training effect varies between the studies, likely reflecting differences in the stimuli and tasks that were used and limitations in the spatial resolution of ERPs. Nevertheless, it is noteworthy that large increases in P2 with effective discrimination training of unfamiliar speech contrasts were reported in central and left temporal electrode sites. In contrast, smaller increases in P2 not specific to training (i.e., after repeated exposure to sounds with no training and that yielded no improvement in behavioral performance) were reported in frontal and right temporal sites (Reinke et al. 2003; Sheehan et al. 2005). The present findings are consistent with these previous results and further suggest that there is a specific contribution of left pSTS neurons to the P2 training effect.

In terms of neural mechanism, the increase in P2 with training has tentatively been attributed to increased neural synchrony and strengthening of neural connections (Tremblay et al. 2001) or to the recruitment of new neurons (Reinke et al. 2003) in stimulus feature maps, analogous to the reorganization with training described for primary sensory cortical regions (Recanzone et al. 1993). Linear increases in P2 amplitude have also been reported for increases in memory load and were attributed in these cases to increased reliance on phonological short-term memory (Conley et al. 1999; Wolach and Pratt 2001). The results of the present study are consistent with the activation of new short-term neural representations of novel auditory categories in the left pSTS as a source for the P2 training effect.

Model of Auditory Categorization in Left Temporal Lobe

We propose that a main factor distinguishing the processing in left mSTS and pSTS is the affinity of the former to highly abstract long-term representations and of the latter to relatively veridical (with low level of abstraction) short-term representations of sound categories.

Because of its highly dynamic nature and crucial communicative value, the speech signal must be abstracted from the analog detailed physical information to permit efficient extraction of the phonetic information that is consistent within phoneme categories. Harnad (1982, 1987) postulates a system for speech perception in which input is represented both as analog representations that are faithful to the physical spectrotemporal properties and instance-to-instance variations in the signal and as a categorical representation that retains only the invariant information across different instances of the category. The categorical representations are postulated to be highly reduced (filtered) reflections of the input structure, and they are also associated with arbitrary (symbolic) category labels that constitute the basis for the lexicon of a language.

We suggest that prelexical abstract long-term representations of highly familiar and overlearned sounds such as native speech syllables are coded in a ventral pathway originating in HG and projecting to the left mSTS. The left mSTS stores representations that are highly abstracted from the analog detailed information in HG and surrounding sensory cortex. The left mSTS representations can be accessed and processed for meaning by ventral and anterior temporal regions and the AG at a semantic and syntactic level of analysis.

In contrast, sequences of unfamiliar complex sounds cannot be mapped onto long-term abstract representations and must instead be stored as consecutive segments of information for subsequent processing. We suggest that short-lived neural representations of recently trained sounds are stored in the left pSTS and have relatively low levels of abstraction from the detailed spectrotemporal information represented in sensory auditory cortex. The low level of abstraction reflects the

relevant (recently trained) sound features that provide the basis for processing newly acquired sound categories. However, these neural representations also retain much of the instance-to-instance variation in spectrotemporal information within the category. As a result, perception of novel sounds as mediated by the left posterior temporal cortex is expected to be less categorical, in the sense that discrimination within category remains relatively high, and identification does not generalize well to new instances of the category. These neural representations are short-lived in the sense that they are formed in the context of learning new sounds and may be lost if not reinforced by training or may eventually be replaced by long-term representations in left mSTS. Repeated exposure to multiple instances of the category can facilitate the extraction of the invariant properties of the category and the formation of long-term categorical representations in the left mSTS. Interestingly, the activation of phonemic and nonphonemic category representations in the left superior temporal sulcus occurs within a similar time window of about 220 ms.

In conclusion, we suggest that the left pSTS plays a role in short-term representation of relevant sound features that provide the basis for identifying newly acquired sound categories. The neural representations in left pSTS consist of low-level abstractions of the sensory input information. In contrast, categorization of familiar phonemic patterns is mediated by long-term, highly abstract, and categorical representations in left mSTS.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/

Funding

National Institute on Deafness and Other Communication Disorders (R01 DC006287 to E.L.); National Institutes of Health (M01 RR00058).

Notes

The authors wish to thank Natasha Tirko and Mark Mulcaire-Jones for their help in testing subjects. *Conflict of Interest*: None declared.

References

Arnott SR, Binns MA, Grady CL, Alain C. 2004. Assessing the auditory dual-pathway model in humans. Neuroimage. 22:401–408.

Belin P, Zatorre RJ. 2000. 'What', 'where' and 'how' in auditory cortex. Nat Neurosci. 3:965–966.

Belin P, Zatorre RJ. 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport. 14:2105–2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. Nature. 403:309–312.

Binder JR, Desai RH, Graves WW, Conant LL. 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb Cortex. 19:2767–2796.

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.

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. 2004. Neural correlates of sensory and decision processes in auditory object identification. Nat Neurosci. 7:295–301.

Blumstein SE, Myers EB, Rissman J. 2005. The perception of voice onset time: an fMRI investigation of phonetic category structure. J Cogn Neurosci. 17:1353–1366.

Bosnyak DJ, Eaton RA, Roberts LE. 2004. Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. Cereb Cortex. 14:1088–1099.

Brechmann A, Scheich H. 2005. Hemispheric shifts of sound representation in auditory cortex with conceptual listening. Cereb Cortex. 15:578–587.

Buchsbaum BR, D'Esposito M. 2008. Repetition suppression and reactivation in auditory-verbal short-term recognition memory. Cereb Cortex. 19:1474–1485.

Buchsbaum BR, Olsen RK, Koch P, Berman KF. 2005. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron. 48:687–697.

Burton MW, Small SL. 2006. Functional neuroanatomy of segmenting speech and nonspeech. Cortex. 42:644–651.

Callan DE, Tajima K, Callan AM, Kubo R, Masaki S, Akahane-Yamada R. 2003. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. Neuroimage. 19:113–124.

Carreiras M, Seghier ML, Baquero S, Estevez A, Lozano A, Devlin JT, Price CJ. 2009. An anatomical signature for literacy. Nature. 461:983–986.

Conley EM, Michalewski HJ, Starr A. 1999. The N100 auditory cortical evoked potential indexes scanning of auditory short-term memory. Clin Neurophysiol. 110:2086–2093.

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

Culham JC, Kanwisher NG. 2001. Neuroimaging of cognitive functions in human parietal cortex. Curr Opin Neurobiol. 11:157–163.

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

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

Desai R, Liebenthal E, Possing ET, Waldron E, Binder JR. 2005. Volumetric vs. surface-based alignment for localization of auditory cortex activation. Neuroimage. 26:1019–1029.

Desai R, Liebenthal E, Waldron E, Binder JR. 2008. Left posterior temporal regions are sensitive to auditory categorization. J Cogn Neurosci. 20:1174–1188.

Duncan J, Owen AM. 2000. Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci. 23:475–483.

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM. 1999. Improved auditory cortex imaging using clustered volume acquisitions. Hum Brain Mapp. 7:89–97.

Fink GR, Halligan PW, Marshall JC, Frith CD, Frackowiak RS, Dolan RJ. 1996. Where in the brain does visual attention select the forest and the trees? Nature. 382:626–628.

Geiser E, Zaehle T, Jancke L, Meyer M. 2008. The neural correlate of speech rhythm as evidenced by metrical speech processing. J Cogn Neurosci. 20:541–552.

Geschwind N, Levitsky W. 1968. Human brain: left-right asymmetries in temporal speech region. Science. 161:186–187.

Giraud AL, Price CJ. 2001. The constraints functional neuroimaging places on classical models of auditory word processing. J Cogn Neurosci. 13:754–765.

Godey B, Schwartz D, de Graaf JB, Chauvel P, Liegeois-Chauvel C. 2001. Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. Clin Neurophysiol. 112:1850–1859.

Golestani N, Paus T, Zatorre RJ. 2002. Anatomical correlates of learning novel speech sounds. Neuron. 35:997–1010.

Golestani N, Zatorre RJ. 2004. Learning new sounds of speech: reallocation of neural substrates. Neuroimage. 21:494–506.

Guthrie D, Buchwald JS. 1991. Significance testing of difference potentials. Psychophysiology. 28:240–244.

Harnad S. 1982. Metaphor and mental duality. In: Simon TW, Scholes RJ, editors. Language, mind and brain. Hillsdale (NJ): Erlbaum. p. 189–211.
Harnad S. 1987. Category induction and representation. In: Harnad S, editor. Categorical perception: the groundwork of cognition. New York: Cambridge University Press. p. 535–565.
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.
Hosmer DJ, Lemeshow S. 2004. Applied logistic regression. NY: Wiley.
Indefrey P, Levelt WJ. 2004. The spatial and temporal signatures of word production components. Cognition. 92:101–144.
Jacquemot C, Scott SK. 2006. What is the relationship between phonological short-term memory and speech processing? Trends Cogn Sci. 10:480–486.
Jancke L, Wustenberg T, Scheich H, Heinze HJ. 2002. Phonetic perception and the temporal cortex. Neuroimage. 15:733–746.
Kaas JH, Hackett TA. 1999. 'What' and 'where' processing in auditory cortex. Nat Neurosci. 2:1045–1047.
Katzir T, Misra M, Poldrack RA. 2005. Imaging phonology without print: assessing the neural correlates of phonemic awareness using fMRI. Neuroimage. 27:106–115.
Leech R, Holt LL, Devlin JT, Dick F. 2009. Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. J Neurosci. 29:5234–5239.
Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA. 2005. Neural substrates of phonemic perception. Cereb Cortex. 15:1621–1631.
Lively SE, Logan JS, Pisoni DB. 1993. Training Japanese listeners to identify English /r/ and /l/. II: the role of phonetic environment and talker variability in learning new perceptual categories. J Acoust Soc Am. 94:1242–1255.
Locasto PC, Krebs-Noble D, Gullapalli RP, Burton MW. 2004. An fMRI investigation of speech and tone segmentation. J Cogn Neurosci. 16:1612–1624.
Middlebrooks JC. 2002. Auditory space processing: here, there or everywhere? Nat Neurosci. 5:824–826.
Morrison GS, Kondaurova MV. 2009. Analysis of categorical response data: use logistic regression rather than endpoint-difference scores or discriminant analysis. J Acoust Soc Am. 126:2159–2162.
Mottonen R, Calvert GA, Jaaskelainen IP, Matthews PM, Thesen T, Tuomainen J, Sams M. 2006. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. Neuroimage. 30:563–569.
Myers EB, Blumstein SE, Walsh E, Eliassen J. 2009. Inferior frontal regions underlie the perception of phonetic category invariance. Psychol Sci. 20:895–903.
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.
Obleser J, Boecker H, Drzezga A, Haslinger B, Hennenlotter A, Roettinger M, Eulitz C, Rauschecker JP. 2006. Vowel sound extraction in anterior superior temporal cortex. Hum Brain Mapp. 27:562–571.
Obleser J, Eisner F. 2009. Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn Sci. 13:14–19.
Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9:97–113.
Pascual-Marqui RD. 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find Exp Clin Pharmacol. 24(Suppl D):5–12.
Petersen SE, van Mier H, Fiez JA, Raichle ME. 1998. The effects of practice on the functional anatomy of task performance. Proc Natl Acad Sci U S A. 95:853–860.
Poldrack RA. 2000. Imaging brain plasticity: conceptual and methodological issues—a theoretical review. Neuroimage. 12:1–13.
Rauschecker JP. 1998. Parallel processing in the auditory cortex of primates. Audiol Neurootol. 3:86–103.
Rauschecker JP, Tian B. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A. 97:11800–11806.
Recanzone GH. 2001. Spatial processing in the primate auditory cortex. Audiol Neurootol. 6:178–181.
Recanzone GH, Schreiner CE, Merzenich MM. 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J Neurosci. 13:87–103.
Reinke KS, He Y, Wang C, Alain C. 2003. Perceptual learning modulates sensory evoked response during vowel segregation. Brain Res Cogn Brain Res. 17:781–791.
Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci. 2:1131–1136.
Romanski LM, Tian B, Fritz JB, Mishkin M, Goldman-Rakic PS, Rauschecker JP. 2000. Reply to 'What', 'where' and 'how' in auditory cortex'. Nat Neurosci. 3:966.
Saad ZS, Glen DR, Chen G, Beauchamp MS, Desai R, Cox RW. 2009. A new method for improving functional-to-structural MRI alignment using local Pearson correlation. Neuroimage. 44:839–848.
Scott SK, Blank CC, Rosen S, Wise RJ. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 123(Pt 12):2400–2406.
Scott SK, Johnsrude IS. 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26:100–107.
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.
Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci. 23:5545–5552.
Shahin A, Roberts LE, Pantev C, Trainor LJ, Ross B. 2005. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. Neuroreport. 16:1781–1785.
Sheehan KA, McArthur GM, Bishop DV. 2005. Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? Brain Res Cogn Brain Res. 25:547–553.
Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers.
Tremblay K, Kraus N, McGee T, Ponton C, Otis B. 2001. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. Ear Hear. 22:79–90.
Ungerleider LG, Mishkin M. 1982. Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield RJW, editors. Analysis of visual behavior. Cambridge (MA): MIT Press. p. 549–586.
Verkindt C, Bertrand O, Thevenet M, Pernier J. 1994. Two auditory components in the 130–230 ms range disclosed by their stimulus frequency dependence. Neuroreport. 5:1189–1192.
Ward BD. 2000. Simultaneous inference for fMRI data. Available from http://afni.nimh.nih.gov/pub/dist/doc/manual/AlphaSim.pdf.
Ward BD. 2001. Deconvolution analysis of FMRI time series data. Available from http://afni.nimh.nih.gov/afni/doc/manual/3dDeconvolve.
Wernicke C. 1874. Der aphasische symptomencomplex: eine psychologische studie auf anatomischer basis. In: Wernicke's works on aphasia: a sourcebook and review. The Hague, The Netherlands: Mouton. p. 91–147.
Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. 2001. Separate neural subsystems within 'Wernicke's area'. Brain. 124:83–95.
Wolach I, Pratt H. 2001. The mode of short-term memory encoding as indicated by event-related potentials in a memory scanning task with distractions. Clin Neurophysiol. 112:186–197.
Xu B, Grafman J, Gaillard WD, Spanaki M, Ishii K, Balsamo L, Makale M, Theodore WH. 2002. Neuroimaging reveals automatic speech coding during perception of written word meaning. Neuroimage. 17:859–870.