ACADEMIC PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

NeuroImage

NeuroImage 19 (2003) 64–79

www.elsevier.com/locate/ynimg

Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals☆

Marc F. Joanisse[a,*] and Joseph S. Gati[b]

[a] *Department of Psychology, University of Western Ontario, London, Ontario N6A 5C2, Canada*
[b] *Laboratory for Functional Magnetic Resonance Imaging, Robarts Research Institute, London, Ontario, Canada*

Received 25 April 2002; revised 2 December 2002; accepted 10 December 2002

Abstract

Speech perception involves recovering the phonetic form of speech from a dynamic auditory signal containing both time-varying and steady-state cues. We examined the roles of inferior frontal and superior temporal cortex in processing these aspects of auditory speech and nonspeech signals. Event-related functional magnetic resonance imaging was used to record activation in superior temporal gyrus (STG) and inferior frontal gyrus (IFG) while participants discriminated pairs of either speech syllables or nonspeech tones. Speech stimuli differed in either the consonant or the vowel portion of the syllable, whereas the nonspeech signals consisted of sinewave tones differing along either a dynamic or a spectral dimension. Analyses failed to identify regions of activation that clearly contrasted the speech and nonspeech conditions. However, we did identify regions in the posterior portion of left and right STG and left IFG yielding greater activation for both speech and nonspeech conditions that involved rapid temporal discrimination, compared to speech and nonspeech conditions involving spectral discrimination. The results suggest that, when semantic and lexical factors are adequately ruled out, there is significant overlap in the brain regions involved in processing the rapid temporal characteristics of both speech and nonspeech signals.

Introduction

Spoken language consists of a complex auditory signal encoding two critical elements of language in parallel: the phonetic form of an utterance and the semantic content that is conveyed through it. Recovering these two components of the speech signal is not trivial, as it involves decoding a set of rapidly changing and fast-fading acoustic events. The present work focuses on the process of perceiving the phonetic form of an auditory speech signal and the neural substrates that are engaged for this purpose. The phonetic form of speech is defined as an abstract code mapping the relationship between an acoustic signal and the articulatory gestures necessary to produce it. The process by which humans are able to recover this phonetic information is a matter of great debate, with theories falling into two broad categories. The first holds that humans possess neural systems that are uniquely adapted for mapping acoustic speech to internal representations of articulatory gestures. As such, the process by which humans perceive speech signals is thought to be radically different from how nonspeech auditory signals are perceived (Liberman and Mattingly, 1985). In contrast, other theories hold that more general auditory processing mechanisms are recruited for speech processing (Bregman, 1990; Massaro, 1997), suggesting that overlapping neural systems are responsible for processing both speech and nonspeech information.

Recent evidence from neuroimaging has again raised the question of whether "speech is special," given the discovery of regions in the brain that appear to be selectively activated during speech processing, compared to processing other types of auditory signals. Imaging research has demonstrated two specific regions that show apparent speech selectivity: the posterior portion of the superior temporal gyrus and sulcus (STG and STS) in both hemispheres and the left inferior frontal gyrus (L IFG). Imaging studies using

☆ This research is supported by grants from the Natural Sciences & Engineering Research Council of Canada and the Canadian Foundation for Innovation.

* Corresponding author. Fax: +1-519-661-3961.
E-mail address: marcj@uwo.ca (M.F. Joanisse).

1053-8119/03/$ – see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1053-8119(03)00046-6

positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have typically found these regions to be activated while listeners processed acoustic speech compared to nonspeech signals. As detailed below, these results might represent good evidence for neural substrates specialized for processing the phonetic form of speech.

An alternative explanation for these results comes from recent neuroimaging studies that have identified regions of superior temporal cortex that are apparently specialized for processing the rapidly time-varying aspects of auditory stimuli (Zatorre et al., 2002). The auditory speech signal relies heavily on this type of information, compared to other types of acoustic signals such as music or environmental noises. One possibility is that neural regions specialized for processing the phonetic form of speech are in fact functionally specialized for all types of auditory signals that involve processing rapidly changing acoustic features. In this theory, speech is special to the extent that it involves analyzing these dynamic characteristics of acoustic signals to a greater degree than other types of auditory information that humans encounter on a daily basis.

The present work explored the theory of a functional overlap in the neural regions responsible for processing the phonetic form of speech and other types of dynamic auditory information. Event-related FMRI was used to measure regional changes in blood oxygenation while participants discriminated synthesized speech or nonspeech signals. It was theorized that manipulating the acoustic characteristics of the sounds being discriminated would reveal regions involved in processing rapidly changing auditory signals either in speech or across all types of acoustic stimuli. Of particular interest were the two brain regions commonly associated with processing the phonetic form of speech: the posterior portion of superior temporal cortex and the left inferior frontal gyrus. Below we outline previous studies that have implicated these regions in phonetic processing, along with evidence suggesting that these regions are functionally specialized for processing either speech or other types of rapid temporal information.

Superior temporal gyrus and sulcus

The first brain region of interest in this study was the posterior portion of the superior temporal lobe adjacent to primary auditory cortex. Multiple studies have identified bilateral regions of increased activation in the posterior STG and STS during speech processing, compared to nonspeech processing. This result has been observed across many types of speech processing tasks, including listening passively to speech compared to noise (Jäncke et al., 2002); listening to vocal sounds of all types compared to nonspeech environmental noises, distorted speech, noise with a speech-like amplitude envelope, and bells (Belin et al., 2000); listening to intelligible sentences compared to the same sentences that have been spectrally modified to destroy intelligibility (Scott et al., 2000) nonwords compared to sinewave analogues (Vouloumanos et al., 2001); listening to sentences, lists of words, and pseudowords compared to tones or noise (Binder et al., 2000); listening to words in one's native language compared to words in a foreign language (Schlosser et al., 1998); rhyme and phoneme monitoring compared to tones or noise (Demonet et al., 1992; Jäncke et al., 2002; Zatorre et al., 1992); and discriminating the initial phoneme in a pair of words compared to tones (Burton et al., 2000).

One interpretation of these data is that this region is specifically involved in processing the phonetic form of speech. However, this conclusion is complicated by the need to control for semantic processing engaged by intelligible speech. This is because the superior temporal gyrus lies at the juncture of two proposed neural streams involved in speech processing (Binder et al., 1997; Hickok and Poeppel, 2000): a "ventral" pathway that connects primary and associative auditory cortex in STS and STG to left middle temporal regions involved in processing semantic information (e.g., Shaywitz et al., 1994) and a "dorsal" pathway that connects auditory cortex to temporal and frontal regions when accessing phonological segments such as phonemes. This raises a critical concern for neuroimaging studies of speech: since intelligible speech contains both phonetic and semantic information, it engages both of these pathways. It is also likely that these pathways tend to interact during processing, similar to what is proposed for dorsal and ventral pathways in vision (Goodale and Milner, 1992). This potential interactivity makes it difficult to accurately subtract the effects of semantic processes in a neuroimaging experiment of speech. For instance, it might not be possible to tease apart "semantic" and "phonological" brain regions strictly by comparing neural activation obtained for intelligible speech to an unintelligible nonspeech signal; differences in activation might be due to either phonetic or semantic factors or an interaction of the two. This makes it difficult to determine whether activation obtained by comparing intelligible speech to nonspeech or unintelligible speech baselines actually corresponds to regions involved in phonetic processing (e.g., Belin et al., 2000; Schlosser et al., 1998; Scott et al., 2000; Vouloumanos et al., 2001). Speech/nonspeech differences in such studies might be due to either semantic or phonetic processing or an interaction among these factors.

Another consideration for understanding the function of posterior superior temporal regions in speech perception is the extent to which it is made up of areas that are functionally specialized for processing phonetic information. Spoken language is composed of multiple periodic and aperiodic spectral events that tend to fall into two broad categories. The first are characterized as brief acoustic events (typically, 15–50 ms) during which a signal's spectral components rapidly change states, which we refer to here as rapid temporal or dynamic acoustic features. The second is typified by steady-state periods during which a

signal's spectral components remain relatively stable, which we refer to as spectral acoustic features. For instance, the syllable [ba] can be described as having both a rapid temporal component during which the spectral formants rapidly change frequency, critical in recognizing the [b] consonant, and a period during which formant frequencies remain relatively stable, which is critical in recognizing the vowel [a]. Ultimately, however, both types of information are important for recognizing the constituent phonemes of a speech stream. For instance, many languages use durational cues to contrast long and short vowels; likewise, frequency peaks are an important cue for recognizing fricatives. As such, it is unrealistic to dichotomize vowels and consonants as relying uniquely on a single type of cue. The present experiment simply sought to differentiate the effects of the two types of acoustic features as much as possible, in both speech and nonspeech signals, in order to better understand how they might be differently processed in the brain.

Evidence from fMRI studies indicates functional specialization in superior temporal cortex for the dynamic or spectral aspects of auditory signals. A posterior region of left superior temporal gyrus appears to be sensitive to rapidly changing auditory events, whereas a similar area in the right superior temporal gyrus shows greater activation for signals containing more steady-state events (Zatorre and Belin, 2001). This raises the possibility that apparently speech-specific activation might be indicative of regions specialized for processing the types of rapidly changing features that typify speech signals. A study by Scott et al. (2000) addressed this issue by using PET to measure neural activation while participants listened to either auditory sentences or speech signals that were acoustically distorted in ways that destroyed intelligibility. This study identified regions of activation in the left STG lateral and anterior to A1 cortex while participants listened to either intelligible speech or unintelligible acoustic signals (speech that has been spectrally rotated, Blesser, 1972) that contained dynamic acoustic features similar to those found in speech. This raised the possibility that both types of signals engage brain regions that respond to the phonetic components of the speech signals. However, this conclusion assumes that spectrally rotated signals used in the study did in fact retain the phonetic characteristics of normal speech. While some acoustic cues for consonants and vowels are maintained in rotated speech (e.g., vowel-like steady-state periods, acoustic noise typical of fricatives, and short gaps that might signal consonant voicing), participants typically require a great deal of training to perceive the phonemic content of these signals (Blesser, 1972). An alternative explanation is that rotated speech preserves the basic acoustic components of speech—the spectral and dynamic events that are typical of spoken language—but lacks the actual phonetic content that signals identifiable and discriminable phonemes or syllables. If this is the case, results from Scott et al. might instead be indicative of brain regions that are specialized for processing dynamic auditory information, including speech and nonspeech sounds that share key acoustic characteristics of speech.

Some evidence exists to support this possibility. Binder et al. (2000) used fMRI to identify regions of neural activation while participants listened to lists of real words, pseudowords, and reversed speech, compared to pure tones or noise. This study revealed common regions of activation for the three speech conditions that were not activated during the nonspeech conditions, even though two of the three speech conditions were specifically designed to preclude semantic activation. The authors suggested that this result could be due to the dynamic characteristics of the speech stimuli that were not present in the nonspeech conditions. However, they also conceded that these regions of activation could have been due to the phonetic content of the signals, given that listeners might have been able to extract phonetic information from both the pseudowords and the reversed words.

In summary, it remains unclear whether superior temporal regions exist that are functionally specialized for processing the phonetic form of speech. While some studies have revealed apparently speech-selective areas, these may be due to either semantic factors or acoustic differences between speech and nonspeech signals. The present study seeks to address this question by directly comparing discrimination of either dynamic or steady-state information in speech and nonspeech signals. In addition, the stimuli and task were designed to minimize the extent to which they engaged semantic processes, in order to attenuate the impact of this type of processing.

Inferior frontal gyrus

Aphasia and neuroimaging studies have also implicated IFG regions in speech processing. The classical Broca's area, encompassing areas 44 and 45 of left IFG, has long been considered a key area in expressive language. However, an emerging body of evidence also suggests that this region is involved in receptive language processing. For instance, while studies have typically found speech production deficits in Broca's aphasics, there is also evidence that these patients have abnormal receptive language abilities. Patients with Broca's aphasia demonstrate deficits in discriminating some speech contrasts (Blumstein, 1998) and atypical priming effects during auditory lexical decision (Utman et al., 2001). They also have difficulty with higher-order phonological processing, such as auditory rhyme judgments (Blumstein et al., 2000) and phoneme blending and deletion (Berndt et al., 1996).

Neuroimaging evidence also indicates the involvement of L IFG in speech processing. PET and fMRI studies have found left IFG activation during phoneme monitoring (Demonet et al., 1992; Jäncke et al., 2002), syllable counting (Poldrack et al., 1999), phoneme discrimination (Burton et al., 2000; Fiez et al., 1995), and rhyme judgment (Poldrack et al., 2001; Zatorre et al., 1992). Burton (2001) has sug-

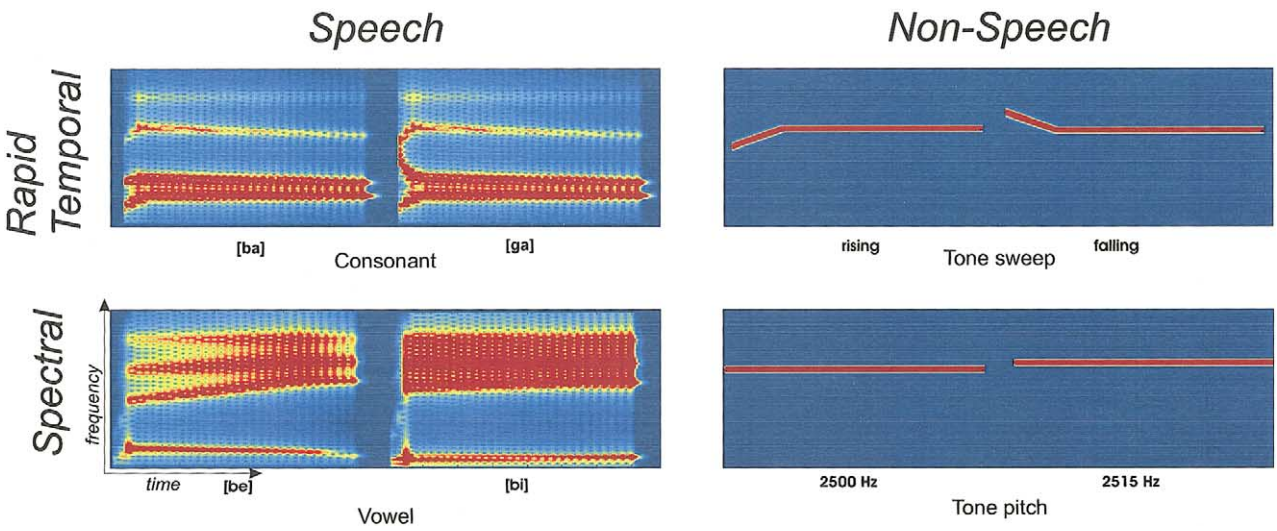

Fig. 1. Stimulus types used in the four conditions of this experiment.

gested that the common factor among these studies is that they require participants to access higher-order phonological representations such as phonemes and syllables. For instance, in a PET study of speech processing, Zatorre et al. (1992) found IFG activation when participants were asked to monitor pairs of words for a common final phoneme (e.g., fa*t*–ti*d*), but not when they passively listened to these same stimuli. One possibility is that left IFG is involved in accessing higher-order phonological representations such as articulatory gestures and that superior temporal regions are responsible for accessing more basic characteristics of speech such as phonetic features. Burton et al. (2000) tested this theory using an auditory speech discrimination task and fMRI. In one experiment, participants were asked to judge whether a pair of otherwise identical words had the same initial phoneme (e.g., *tip–tip* vs *tip–dip*); the second experiment was similar except that stimulus words differed in terms of several phonemes not involved in discrimination (*tip–tomb* vs *tip–doom*). It was found that regions of IFG were only activated during the second experiment and not the first. The authors theorized that this difference was due to the fact that items in the second experiment (*tip–doom*) required the segmentation of the speech stream into its constituent phonemes, whereas the first experiment (*tip–dip*) did not. This would seem to support the theory that left IFG is specifically engaged by tasks requiring access to higher-order phonological representations, rather than more basic auditory or phonetic information.

An alternative explanation for these differences is that speech pairs requiring segmentation tend to place greater demands on secondary capacities such as working memory, which is reflected in greater activation in IFG. In this theory, any increase in task demands might also tend to increase left IFG activation without invoking higher-order representations such as segmentation. Results from Poldrack et al. (2001) are suggestive in this regard. In their study, fMRI participants performed an auditory sentence verification task on speech samples that were temporally compressed to different degrees. This temporal compression had the effect of shortening the duration of many acoustic features of speech, which resulted in poorer performance on this task. It was found that activation in regions of left IFG increased as a function of the degree of acoustic compression as long as the signal remained intelligible. At higher levels of compression that eliminated intelligibility, activation levels decreased. Further analyses revealed that this region of IFG was also activated during a (visual) pseudoword rhyme judgment task. These results suggest that left IFG activation might tend to increase as a function of any manipulation that results in increased processing demands.

As is the case for superior temporal cortex, existing studies leave it unclear whether areas of IFG represent language-specific processing centers. Many studies have found greater activation in left IFG for speech processing, compared to similar tasks involving nonspeech signals (Burton, 2001; Demonet et al., 1992; Zatorre et al., 1992); these differences might reflect the relative ease in processing the nonspeech signals or differences in the acoustic characteristics of the speech and nonspeech stimuli. Data from a PET study by Fiez et al. (1995) seem to support this theory; their study found bilateral IFG activation across multiple speech and nonspeech conditions involving monitoring for words, syllables, vowels, and tone triplets. The common factor among these different stimuli would seem to be the rapid modulation of the signals, suggesting that this aspect of the signals was responsible for the observed pattern of activation in IFG. Other neuroimaging studies appear to support the observation that rapidly changing nonspeech stimuli also engage IFG (Johnsrude et al., 1997; Müller et al., 2001). These studies have found significant left IFG activation when participants monitored for nonspeech auditory signals involving dynamic information, compared to

similar signals that were either less rapid or completely steady state.

Of interest in the present study was the extent to which regions of IFG will tend to be activated during speech processing that does not involve higher-order phonological processing such as segmentation. As we discuss below, the speech stimuli used in this study were designed to bear minimally on phonological processes such as rhyme judgment and segmentation and are instead argued to be engaging acoustic–phonetic knowledge of phoneme identities. A second question regarded the extent to which any IFG activation during speech processing would also tend to overlap with areas of activation while processing dynamic aspects of nonspeech signals. This again would help inform the theory of whether neural regions specialized for speech processing exist in IFG.

Rationale for the present study

This study sought to address issues related to brain regions involved in processing the phonetic form of auditory speech and other types auditory signals that share the acoustic features of speech. fMRI was used to determine whether neural systems exist that are specialized for processing phonetic information. It also investigated an alternative hypothesis, that regions engaged during speech processing are also involved in processing other types of acoustic signals that have dynamic features similar to those found in speech. Auditory discrimination was chosen as a behavioral task because it emphasizes the phonetic/acoustic properties of stimuli (Liberman et al., 1957). It was expected that semantic and lexical processing would thus be minimized during task intervals, allowing for a more precise specification of phonetic/acoustic processing subsystems. Nevertheless, we concede that this task—like any other cognitive task—will tend to invoke some amount of semantic activation since participants are likely always thinking about the meaningfulness of a task and stimuli, even when they are being asked to actively attend to nonsemantic aspects of the stimuli.

The stimuli that were used represented the crossing of the two variables of interest in this study: speech vs nonspeech and dynamic vs spectral (Fig. 1). In the speech conditions, participants discriminated pairs of CV syllables that differed with respect to either a brief and rapidly changing acoustic feature (formant transitions that signal consonant place of articulation) or a steady-state spectral feature (signaling vowel quality). The use of CV syllables was intended to further minimize the potential impact of lexical effects in the imaging results. In the case of the consonant condition, this was reinforced by the use of a minimal pair of nonwords ([ba] and [ga]). Stimuli in the vowel condition were [bi] and [be], which did correspond to familiar words. It was not possible to use nonwords in the case of the vowel stimuli, as no minimal pair of English-like nonword CV syllables minimally contrasts along a single acoustic dimension.[1] As discussed further below, the extent to which this affected the imaging results must be taken into consideration when interpreting the imaging data. Two sets of nonspeech stimuli were used, which were also distinguished using either dynamic or spectral information. These consisted of pairs of sinewave tone sweeps or steady-state tone pitches that were intended to capture the general auditory distinction between the consonant or vowel conditions, but were not recognizable as speech.

Methods

Subjects

Participants were seven healthy right-handed adults with a mean age of 27 (range: 20–32 years); four were female. All were native speakers of English and were prescreened for normal hearing and neurological and cognitive disorders. Participants gave written consent before participating and were paid for participating in the study. Behavioral testing protocols were reviewed and approved by the University of Western Ontario Office of Research Ethics.

Stimuli

Four sets of acoustic stimuli were used. All sets consisted of eight digital waveforms (16-bit, 8000-Hz sample rate) that formed an evenly spaced continuum varying along a specific acoustic dimension (Fig. 1). The four sets represented the crossing of two manipulations: speech vs nonspeech signals and frequency vs spectral discrimination. The speech stimuli consisted of synthetic syllables created using the Klatt cascade/parallel formant synthesizer (Klatt, 1980). Two speech continua were created in which either the consonant or the vowel portion of a CV syllable was manipulated along a single acoustic dimension. The duration of items in both sets was 330 ms. Stimuli in the Consonant condition consisted of the syllables “ba” and “ga,” in which the onset frequency of the syllable’s second formant (F2) transition was manipulated. The F2 transition consisted of a 40-ms sweep with an onset frequency varying from 1100 to 1800 Hz in 100-Hz steps. The extrema of this continuum represented canonical North American English instances of ga at 1100 Hz and ba at 1800 Hz. All other acoustic variables were kept constant across items, including the onset and offset frequencies of all other formant transitions (F2 offset: 1200 Hz; first formant (F1) onset, offset: 400, 750 Hz; third formant (F3) onset, offset: 2000, 2500 Hz) and the following [a] vowel (F1: 750 Hz; F2: 1200 Hz; F3: 2500 Hz).

Stimuli used in the Vowel condition consisted of an

[1] The exception are light CV syllables such as [bɪ] and [bɛ] (*bih*, *beh*) that are phonotactically prohibited in isolation in English and are therefore inappropriate for the present study.

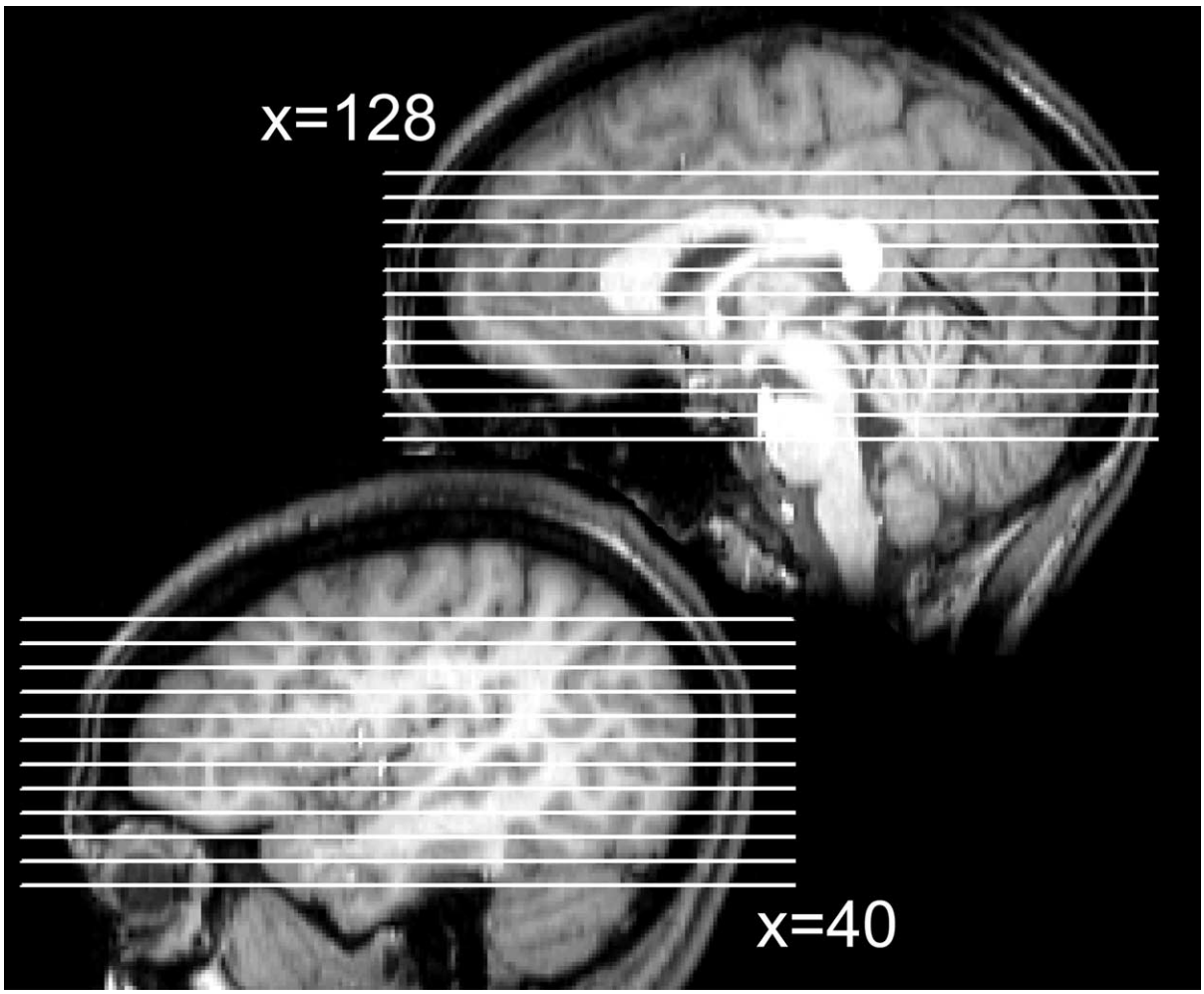

Fig. 2. Location and orientation of functional slices used in this experiment.

evenly spaced continuum between the syllables [be] and [bi], which are typically differentiated by the configuration of the steady-state F1 and F2 components of the speech signal. F1 values varied from 420 to 280 Hz in 20-Hz steps and F2 from 1720 to 2100 Hz in ~54-Hz steps. All other acoustic cues were held constant across the items in this continuum, including the formant transitions signaling the initial [b] in the syllable (F1 onset, offset: 200, 280 Hz; F2 onset, offset: 900, 2100 Hz; F3 onset, offset: 2100, 2700 Hz) and the steady-state F3 component of the vowel (2300 Hz).

The nonspeech stimuli consisted of two sets of sinewave tones that were varied along either dynamic or spectral acoustic dimensions, matching those manipulated in the consonant and vowel conditions. The Tone Sweep stimuli consisted of a 320-ms 2500-Hz steady-state sinewave tone preceded by an 80-ms frequency sweep. The onset frequency was varied from 2000 to 3000 Hz in 125-Hz steps, resulting in eight waveforms varying solely in the extent to which the sweep was rising or falling in pitch. The Tone Pitch stimuli consisted of steady-state sinewave tones of the same duration that varied in frequency from 2500 Hz to 2600 Hz in ~14-Hz increments.

Behavioral procedure

During scanning, participants performed a standard AX discrimination task in which they listened to pairs of stimuli that they were asked to judge as sounding the "same" or "different." Responses were recorded by pressing one of two buttons on an MRI-compatible numerical keypad. Each of the four conditions was presented in 30-trial segments that lasted approximately 6 min. All participants performed all four conditions, with condition order randomized across subjects. Functional scans for Subject 2 for the *sine-pitch* condition were excluded from analyses due to equipment failure. Stimuli were presented using an MRI-compatible headset connected to a set of piezoelectric speakers, located outside the magnet, using a pair of PVC tubes. The headset also served to attenuate EPI noise (but not stimulus sounds) such that subjects reported hearing stimuli clearly above the gradient noise at a safe and comfortable level of amplifica-

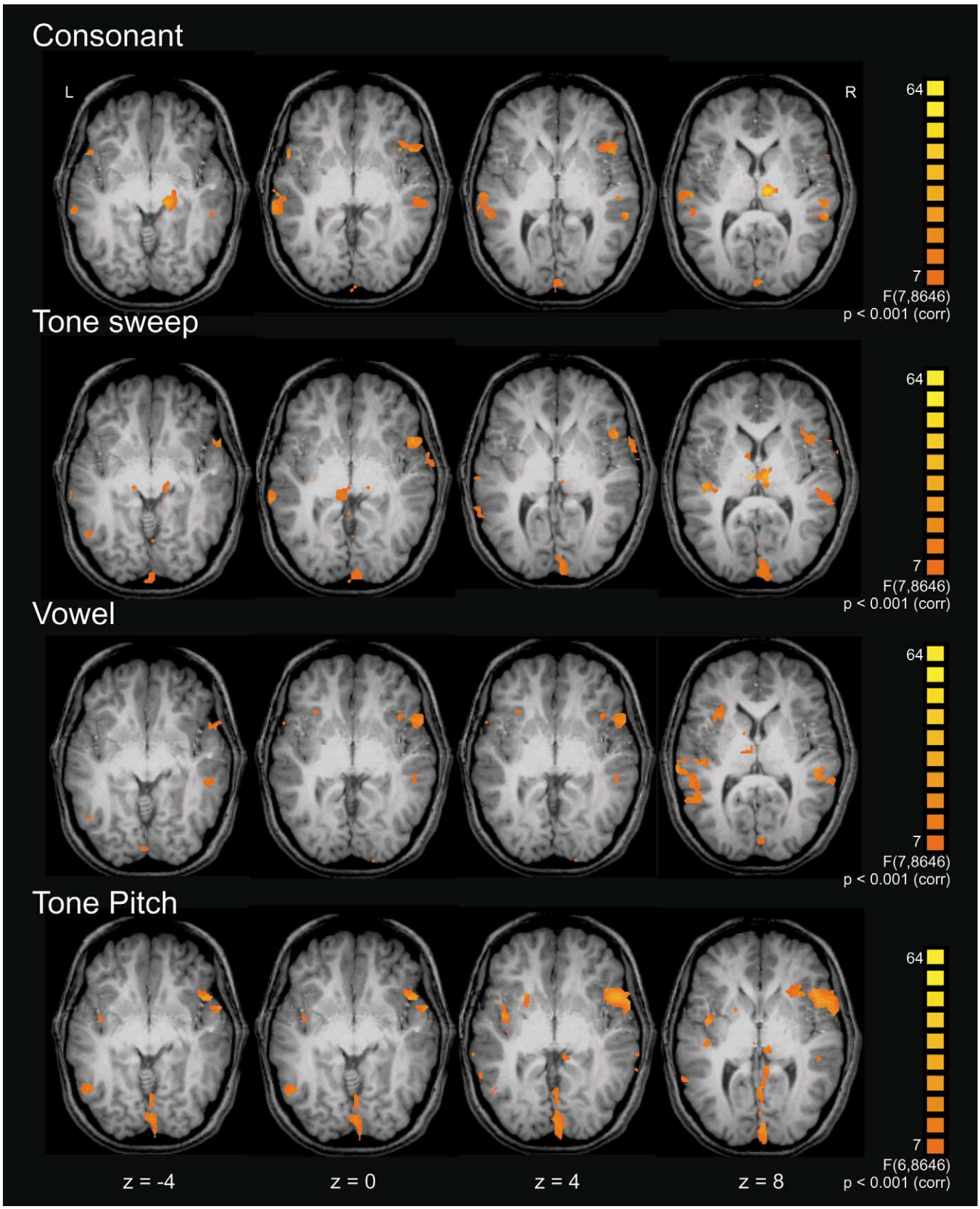

Fig. 3. Clusters of significantly activated voxels during the auditory discrimination task, for the four conditions. *t* value is indicated by color intensity.

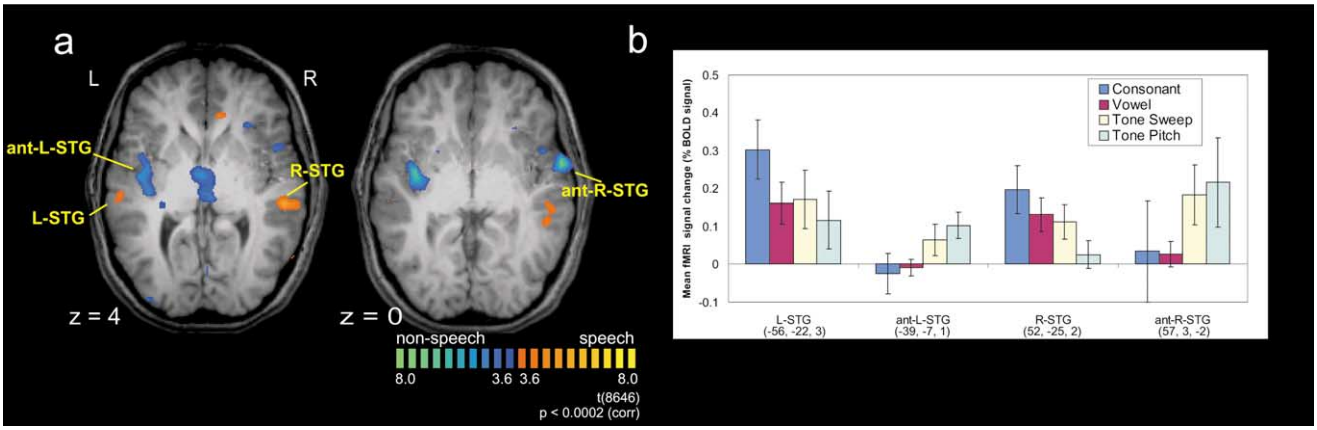

Fig. 4. (a) Clusters of significantly activated voxels for the speech (Consonant and Vowel conditions) vs nonspeech (Sine Sweep and Sine Tone conditions) contrast. *t* value is indicated by color intensity. (b) Mean event-related responses for indicated activated clusters across all conditions. Talairach coordinates of cluster centers of gravity are indicated in parentheses. Vertical bars indicate standard error of mean.

tion. Stimulus presentation was controlled using a Macintosh G4 computer, which was also used to record subjects' responses.

During each trial, a pair of stimuli from a single set was presented with an interstimulus interval of 250 ms. An intertrial interval of 12 s (10 s for Subjects 1 and 2) allowed event-related hemodynamic responses to be recorded relative to each trial (Dale and Buckner, 1997). Three types of stimulus pairs were randomly presented: the same item repeated, pairs that differed by only one step in the continuum, and pairs that differed by three steps (this third condition consisted of pairs differing by six steps in the nonspeech trials.) An equal number of each trial type was used in each condition.

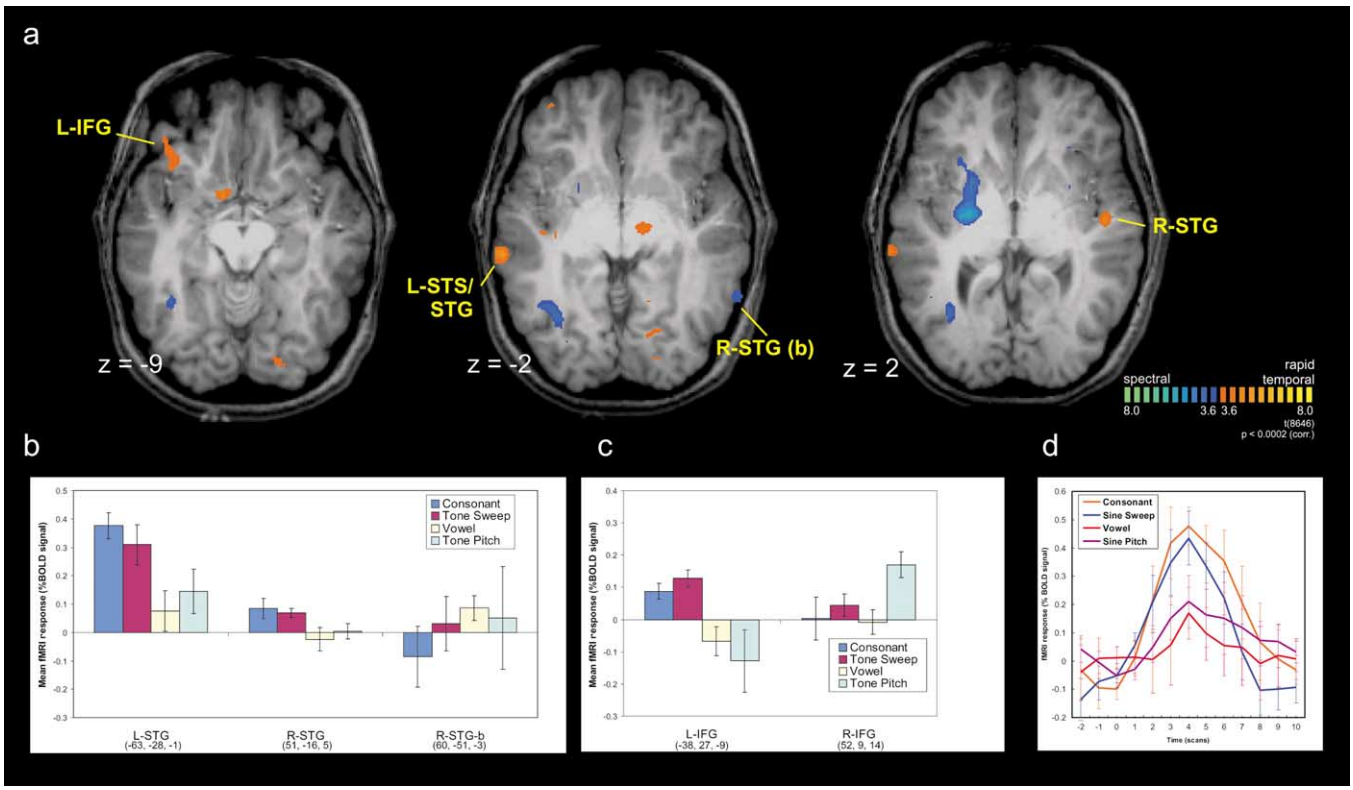

Fig. 5. (a) Clusters of significantly activated voxels for the rapid temporal (Consonant and Sine Sweep conditions) vs spectral (Vowel and Sine Tone conditions) contrast. (b,c) Mean event-related BOLD signal levels for indicated activated clusters, across all conditions. Talairach coordinates of cluster centers of gravity are indicated in parentheses. (d) Mean event-related BOLD signal response curve for the cluster of activation in left STG (vertical bars indicate standard error of the mean).

Scanning procedure

Functional magnetic resonance imaging was performed using a 4-T MRI system (Varian/Siemens) equipped with a hybrid quadrature head coil for signal transmission and reception. Foam padding was used to minimize head movement. Functional images were acquired during behavioral testing using a T2*-weighted, navigator echo corrected, interleaved two-shot gradient EPI pulse sequence for blood oxygen level dependent (BOLD) imaging (Ogawa, 1990). Eleven 64 × 64 transverse-oriented 5-mm slices were acquired in each volume (12 slices for Subjects 5–7), with an inplane resolution of 3 × 3 mm and no interslice gap. A T1 saggital scout image was used to orient slices parallel to the bicommissural plane and to align the topmost slice to the area superior to the top of the corpus callosum (Fig. 2). This assured that the functional volume coincided with the classical language regions that were of interest in the present study. Scans also covered the entire occipital lobe, the posterior portion of the inferior temporal lobe, and a significant portion of the parietal lobe. Frontal regions superior to the middle frontal gyrus, the superior parietal gyrus, and the anterior portion of the inferior temporal gyrus were not acquired during these scans. Additional scanning parameters were as follows: volume acquisition time, 1 s (1.2 s for Subjects 5–7); TE, 15 ms; field of view, 192 × 192 mm; flip angle, 30°; bandwidth, 140.35 kHz. To compensate for magnetic saturation effects, the first three volumes of every scanning session were discarded; behavioral testing did not begin until after these three volumes had been obtained. Functional images were aligned to a high-resolution whole-brain 256 × 256 × 128 T1-weighted anatomical volume acquired during the same session using a 3D segmented driven equilibrium FLASH sequence (FOV, 192 × 192 × 160 mm; in-plane resolution, 0.8 mm; slice thickness, 1.25 mm; inversion time, 600 ms; TE, 5.5 ms; TR, 10.0 ms; flip angle, 15°; bandwidth, 62.50 kHz).

Data analysis

Image analysis was performed using Brain Voyager 4.6 software (Brain Innovation B.V., Maastricht NL, Max Planck Society). Functional images underwent three-dimensional motion correction, Gaussian filtering in the temporal domain (full-width, half-minimum of two volumes) and spatial domain (full-width, half-minimum of two voxels in three dimensions), and mean intensity adjustment to compensate for temporal signal drift. Subjects' functional images were coregistered with their anatomical scans and transformed into the three-dimensional stereotaxic space of Talairach and Tournoux (1988). Statistical analyses were performed on the transformed functional images using a General Linear Model (GLM) with an independent predictor for each condition (Consonant, Vowel, Tone Sweep, and Tone Pitch). Predictor functions consisted of canonical BOLD response estimates using a gamma wave function (delta, 2.5; tau, 1.25) and were based on a task duration of one volume. The GLM measured the extent to which individual voxels fit the predicted time course for one or more of these predictors. Condition-wise regions of activation were obtained for each GLM predictor by locating contiguous clusters of 25 or more voxels that reached a significance threshold of $P < 0.001$ (corrected for multiple comparisons across the entire acquisition volume; Fig. 3).

Statistical contrasts were next performed to identify regions in which activation was statistically greater for speech conditions than for nonspeech conditions and vice versa. This was done using a GLM with the Consonant and Vowel conditions entered as positive predictors and the Tone Sweep and Tone Pitch conditions entered as negative predictors. Clusters of voxels yielding positive and negative regression coefficients represented regions that were differentially active for the speech and nonspeech tasks, respectively (Fig. 4). A second GLM contrast was performed to identify regions differentially activated during the dynamic or spectral discrimination conditions. In this case, the Consonant and Tone Sweep conditions were entered into the GLM as positive predictors, and the Vowel and Tone Pitch conditions were entered as negative predictors (Fig. 5). For both contrasts, a corrected significance threshold was set at $t(8646) = 3.6$, $P < 0.0002$ across a minimum cluster size of 11 contiguous voxels. Correction for multiple comparisons across the acquisition volume was obtained using AlphaSim (http://afni.nimh.nih.gov/afni/AFNI_Help/AlphaSim.html; 10,000 simulations with a single-voxel significance level of $P < 0.00032$ and 6-mm FWHM spatial smoothing; Forman et al., 1995; Xiong et al., 1995).

Anatomical regions and Brodmann Area (BA) estimates were automatically obtained for the center of gravity coordinates of activated clusters using the Talairach Daemon v. 1.1 software (University of Texas Health Science Center at San Antonio). Anatomical regions of interest considered in the present study were areas along the superior temporal gyrus and sulcus (BA 21 and 22) and the inferior frontal gyrus (IFG BA 46/45), in both hemispheres. Activation clusters falling outside these regions were not considered in subsequent analyses; the Appendix lists all significant clusters falling within all cortical regions (criterion: 10 steps in Talairach space from gray matter).

Post hoc analyses of event-related (ER) BOLD signals were performed for activated clusters as follows: each subject's mean ER BOLD signal level was obtained for the activated voxels at each poststimulus scan, across all trials of each condition. Signal levels were independently normalized condition-wise to an estimated relaxation level based on the mean signal level for the two volumes immediately preceding and following the start of each trial in the condition. Post hoc comparisons of activation were based on mean BOLD signal levels of the second to sixth poststimulus volumes for each condition, which captured the time points within which ER BOLD signals typically peak for trials of this duration (Dale and Buckner, 1997). Planned

Table 1
Mean (SD) accuracy rates and reaction times for the four experimental conditions

Task	Discrimination trial type		
	Different–hard	Different–easy	Same
Speech			
Consonant	31.6 (10.8)	65.7 (20.7)	64.3 (11.3)
	990.3 (229)	964.6 (384)	998.0 (340)
Vowel	37.8 (23.9)	70.0 (20.8)	72.9 (16.0)
	1021.5 (190)	919.7 (249)	877.5 (276)
Nonspeech			
Tone Sweep	52.9 (11.1)	91.4 (9.0)	91.4 (9.0)
	971.4 (305)	702.7 (247)	840.0 (190)
Tone Pitch	82.5 (16.4)	88.7 (11.1)	73.8 (18.6)
	801.5 (245)	661.0 (148)	1022.7 (381)

Note. Accuracy rates are percent correct "same" or "different" responses; reaction times are the means for each condition, in milliseconds.

comparisons were then used to compare these signal means using a repeated-measures *t* statistic, one-tailed. Since the number of subjects in the present study precluded a random effects model for locating regions of activation, the use of a repeated measures statistic for comparing obtained signal levels within these regions of interests helped to assure that obtained differences between factors in the two GLM contrasts were statistically robust.

Results

Behavioral results

Table 1 lists participants' response rates and reaction times (RTs) for each trial type of each condition. Response rates were calculated as the number correct discriminations in the one-step ("hard"), three/six-step ("easy") trials, and the mean correct rejections on the no-change trials. RTs were calculated as the mean millisecond lag between the offset of the second sound in the pair and subjects' responses. Two $2 \times 2 \times 3$ repeated measures ANOVAs were performed to measure the effects of three factors of interest on accuracy and RT: linguistic status (speech vs nonspeech), acoustic cue type (rapid temporal vs spectral), and difficulty (same, different–easy, and different–hard). The analysis of accuracy rates revealed a significant effect for linguistic status (F (1,6) = 37.33, $P < 0.001$) and difficulty (F (2,12) = 15.85, $P < 0.05$) and for the three-way interaction (F(2,5) = 13.0, $P < 0.01$). The acoustic cue type main effect and two-way interactions were not significant ($P > 0.05$). The repeated measures ANOVA of the RT data failed to reach significance for both linguistic status and acoustic cue type ($P > 0.05$), but was significant for difficulty (F (2,12) = 5.99, $P < 0.05$). None of the interactions was significant ($P > 0.05$).

Differences in the behavioral measures are of some concern since accuracy and RT differences might be influencing the fMRI results. Of primary concern was the finding that participants' accuracy and RT differed across acoustic difference conditions, such that accuracy was higher and RTs were shorter for the different–easy items than for the different–hard and same items. This difference confirms that subjects found small differences between stimuli more difficult to discriminate, compared to larger differences. In addition, same trials were also difficult to identify, possibly due to the proportion of different trials involving very small acoustic differences. RTs also differed as a function of the difficulty manipulation in a similar way. Since these differences might impact our ability to compare fMRI signals, the different–easy, different–hard, and same trials were conflated for the imaging analyses presented below.

We also found lower accuracy for the speech conditions compared to the nonspeech conditions, which was due at least in part to the fact that many of the acoustically different stimuli in the speech conditions involved within-category judgments. This was especially the case in the hard conditions, in which most of the acoustically different syllable pairs nevertheless represented two instances of the same token (e.g., two acoustically different syllables both falling within the [ga] category). Categorical perception effects typically lead to listeners identifying minimally different pairs of speech tokens as the same sound (Liberman et al., 1957). The finding that participants' accuracy differed between speech and nonspeech conditions might thus be reflecting a greater tendency toward categorical judgments on the speech task. The results suggest that subjects were better at discriminating the nonspeech items because their responses were not categorical for these items. This is reflected in their higher accuracy on the different–hard nonspeech items compared to the different–hard speech items and might also explain the three-way interaction for accuracy. These performance differences must be taken into consideration when interpreting the imaging results, as we discuss further below.

Imaging results

Mean fMRI data for each condition were analyzed to determine cortical areas activated while listeners discriminated stimuli, compared to the baseline intertrial relaxation period. Fig. 3 presents averaged functional activation maps on the four stimulus types. Activation patterns across stimulus types suggest several interesting patterns. As expected, all conditions activated regions of posterior STG/STS in BA 22, adjacent to primary auditory cortex. More interesting was the finding that the Consonant and Tone Sweep conditions yielded activation in the left STG (slices $z = -4$ and $z = 0$) that was not present in the Vowel and Tone Pitch conditions. Activation was observed across all conditions in more superior regions of STS ($z = 8$) and to the greatest extent in the Vowel condition. All four conditions were also found to activate ventral regions of IFG (BA 43/44) in the right hemisphere. Activation appeared more sparse in sim-

ilar regions of left IFG and seemed to be most apparent for the two speech conditions (slices $z = 0$ and $z = 4$). The two speech conditions also appeared to differ with respect to the lateralization of activation in this region, with a slightly larger extent of activation in the left IFG for the Consonant condition and the opposite observation in the right IFG.

To better quantify the extent to which activation in these frontal and temporal regions was influenced by the acoustic properties of the stimuli, two statistical contrasts were performed. Both contrasts sought to identify regions of activation that significantly differentiated two of the conditions from the other two. A drawback of this methodology is the potential for identifying regions that do not represent purely contrastive areas of activation, such as regions in which the speech/nonspeech contrast might have been due to differences in only two of the conditions. This was investigated using a set of planned comparisons of event-related BOLD signals across conditions.

Speech vs nonspeech discrimination

Previous studies have indicated speech-specific activation in superior temporal and inferior frontal cortex in humans. We investigated the extent to which this was true in the present study by identifying clusters of activated voxels that significantly contrasted the speech conditions (Consonant and Vowel) with the nonspeech conditions (Tone Sweep and Tone Pitch) and vice versa (see Appendix). Clusters of activation reaching significance located in superior temporal and inferior frontal gyri are illustrated in Fig. 4a.

Results of this contrast indicated significant clusters in the posterior portion of left and right STG/STS for the speech conditions. Significant clusters were also found in a more anterior portion of the superior temporal gyrus for the nonspeech > speech contrast. These results might indicate the existence of both speech-specific and nonspeech-specific areas of cortex in the superior temporal lobe. To further assess whether these differences represent a true dissociation between the speech and nonspeech conditions in this study, event-related BOLD signal levels for each condition were compared for each significant cluster of voxels (Fig. 4b). These comparisons confirmed the finding that the anterior superior temporal regions tended to show significantly greater activation for the two nonspeech conditions than for the two speech conditions (LH, Consonant–Tone Pitch: $t = 2.2$, Vowel–Tone Sweep: $t = 2.2$, Vowel–Tone Pitch: $t = 2.9$, Ps < 0.05; RH, Consonant–Tone Pitch: $t = 5.9$, Vowel–Tone Sweep: $t = 3.2$, Vowel–Tone Pitch: $t = 2.2$, Ps < 0.05). The Consonant–Tone Sweep comparison was marginal in both hemispheres (LH, $t = 1.9$, $P = 0.054$; RH, $t = 1.9$, $P = 0.053$). Differences within the speech and nonspeech conditions were not significant.

Analyses were also conducted for clusters of significantly activated voxels in more posterior regions of the right and left STG yielding the opposite contrast (speech > nonspeech). However, comparison of activation levels for each condition suggested that these regions were not in fact clearly dissociating the speech and nonspeech conditions (Fig. 4b). The Consonant condition showed significantly greater activation than the two nonspeech tasks in the left and right superior temporal regions (LH, Consonant–Tone Sweep: $t = 2.2$, Consonant–Tone Pitch: $t = 3.4$, Ps < 0.05; RH, Consonant–Tone Sweep, $t = 2.4$, Consonant–Tone Pitch: $t = 2.8$, Ps < 0.05). However, the Vowel condition did not differ significantly from either nonspeech condition in either hemisphere. This suggests that the apparently speech-specific clusters of activation in the posterior superior temporal gyrus were due to large differences between the Consonant and Tone Pitch conditions, rather than an overall tendency for these regions to exclusively respond to speech signals. No significant clusters of activation were observed in either the left or the right IFG.

The results appear to replicate previous findings of activation in superior temporal cortex during speech processing compared to a tone discrimination baseline (Burton et al., 2000; Jäncke et al., 2002; Demonet et al., 1992; Zatorre et al., 1992). However, they also suggest that the specific type of phonetic signal being processed will influence the character of these results, given weaker results for the Vowel condition than the Consonant condition in these regions. In addition, our failure to observe any regions of IFG that discriminated the speech and nonspeech conditions suggests that this was not the key factor in the observed activation in this region.

Dynamic vs spectral discrimination

A second statistical contrast was performed to differentiate regions engaged during the discrimination of dynamic acoustic features (i.e., the Consonant and Tone Sweep conditions) or spectral features (the Vowel and Tone Pitch condition). This analysis revealed several significant clusters in areas of right and left STG and left IFG (Fig. 5a). A large cluster of significantly activated voxels that demonstrated a rapid temporal > spectral contrast was found in a posterior region of left STG lateral to Heschl's gyrus extending into left middle temporal gyrus. Pairwise comparisons revealed significantly greater activation levels for the two rapid temporal discrimination tasks, compared to the two spectral conditions (Fig. 5b; Consonant–Vowel: $t = 3.0$, Consonant–Tone Pitch: $t = 5.2$, Tone Sweep–Tone Pitch: $t = 3.0$, Tone Sweep–Vowel: $t = 2.0$, Ps < 0.05). Additionally, no significant differences were found for comparisons within the two spectral and dynamic conditions. The average event-related response for voxels in this cluster of activation is plotted in Fig. 5c and demonstrates that this pattern of activation was similar across the entire poststimulation period.

A significant cluster was also found in a similar region of the right hemisphere and demonstrated a similar—albeit smaller—differentiation between the rapid temporal and

spectral conditions (Consonant–Vowel: $t = 2.2$, Consonant–Tone Pitch: $t = 6.2$, Tone Sweep–Vowel: $t = 2.0$, Tone Sweep–Tone Pitch: $t = 2.2$, Ps < 0.05). The pattern of activation found in this region is perhaps not surprising given that the region of interest is very similar to the one found in the speech/nonspeech contrast, which yielded a similar pattern of activation levels across the four conditions. A different region of right STG also showed a significant contrast in the opposite direction (Fig. 5b). However, a contrast of condition-wise BOLD signals failed to reveal any significant differences between conditions.

A cluster of significantly activated voxels along the inferior plane of the left IFG showed a significant dynamic > spectral contrast. Condition-wise activation levels showed the predicted differentiation between the dynamic and spectral conditions (Fig. 5c: Consonant–Vowel: $t = 2.9$, Consonant–Tone Pitch: $t = 2.2$, Tone Sweep–Vowel: $t = 6.3$, Tone Sweep–Tone Pitch: $t = 3.0$, Ps < 0.05). Differences within the two dynamic and spectral conditions were not significant.

A cluster in the right precentral gyrus (BA 44) yielded a significant contrast in the opposite direction (spectral > dynamic). However, post hoc analyses of condition-wise activation levels (Fig. 5c) indicated that this region did not demonstrate as clear a dissociation as the controlateral region; whereas the Tone Pitch did show greater activation compared to the Consonant condition ($t = 2.0$, $P < 0.05$), the Consonant–Vowel, Tone Pitch–Consonant, and Tone Sweep–Vowel contrasts failed to reach significance. An additional post hoc comparison indicated that activation for the Tone Pitch condition was also significantly greater than for the Vowel condition ($t = 4.7$, $P < 0.001$). This result suggests that activation in this cluster was due primarily to the Tone Pitch condition, compared to all other regions.

Discussion

Previous studies have identified regions of IFG and STG/STS that are engaged during phonetic processing. The present study addressed a key question raised by these findings, whether these brain regions are specialized for processing speech. Functional MRI was used to measure regions of neural activity while participants performed auditory discrimination tasks. Four types of stimuli were used, representing the crossing of two factors of interest: speech vs nonspeech signals and dynamic vs spectral auditory cues. Condition-wise analyses revealed significant regions of activation in the areas of left and right STG/STS surrounding primary auditory cortex and in left or right IFG, although the extent of activation in these regions tended to vary across all conditions. In order to better specify how these brain regions were differentially engaged during speech or nonspeech processing, statistical contrasts were performed for the speech vs nonspeech and dynamic vs spectral conditions. As discussed below, these contrasts addressed a number of important questions about the neural bases of auditory speech perception.

One objection that might be raised in comparing activation maps across the four discrimination conditions regards the differences in behavioral patterns across these conditions. Participants tended toward a greater proportion of errors on the two speech conditions, especially in the case of within-category discriminations. A consequence of this might have been artificially higher activation levels for the speech conditions compared to the nonspeech conditions, which could have obscured the effects of the acoustic characteristics of the stimuli that were of interest in this study. None of the regions of interest demonstrated an unequivocal speech > nonspeech contrast, which seems to suggest that participants' accuracy and reaction times were not influencing this aspect of the results. In addition, significant fMRI signal differences between the rapid temporal and spectral conditions were not accompanied by differences in the behavioral measures, which again suggests that the fMRI results in this direction were not a consequence of differences in participants' accuracy or reaction times for the relevant stimuli. Nevertheless, we emphasize some caution about the finding that no temporal or frontal regions reliably discriminated speech from nonspeech signals; it remains possible that differences in subjects' accuracy, or differences in the lexical status of the vowel and consonant stimuli, might be obscuring results to this effect. Further research will be necessary to assess whether these results are upheld if lexical status and subject response rates can be more closely controlled for.

GLM contrasts were performed in order to test two (nonorthogonal) hypotheses. The first sought to identify cortical regions that differentiated the two speech and nonspeech discrimination conditions. The second contrast identified regions that were differentially activated while participants discriminated either the dynamic or the spectral components of auditory signals. We were interested in two specific neural regions known to be implicated in speech processing, located in the superior temporal and inferior frontal gyri. Results of each region are discussed below.

Superior temporal regions

We identified two clusters of significant voxels that were more activated for the nonspeech conditions: one in the anterior portion of the right STG and another in the insula adjacent to the left STG. One possible explanation for these regions of activation is that they reflect the novelty of the nonspeech discrimination tasks and the additional resources that they necessitated, although further research is necessary to better understand these results. Analyses also identified posterior regions of left and right STG that showed the opposite effect (speech > nonspeech). This might support the theory of neural regions that are specialized for processing the phonetic form of speech. However, closer investigation of condition-wise activation levels in these regions

suggested that this dissociation was much more graded. The Consonant condition yielded the highest activation levels in this region, compared to the Vowel and Tone Sweep conditions that showed roughly similar levels of activation. It seems that the best interpretation of these results is that these regions were responding to both speech and nonspeech signals, although to different degrees. As such, Consonant discrimination engaged this region to a greater degree than the Tone Sweeps, but also to a greater degree than Vowel discrimination. Thus, this region was sensitive to both the acoustic characteristics of the signal (the dynamic and broadband nature of the speech stimuli) and the nature of the discrimination being performed (discriminating rapidly changing acoustic cues, as in the Consonant and Tone Sweep conditions).

The spectral/rapid temporal contrast identified a region in the posterior portion of left STG that showed significantly greater activation when participants discriminated rapidly changing acoustic cues of both speech and nonspeech conditions. Event-related BOLD signal levels for this region showed a clear differentiation along these lines, demonstrated by greater activation levels for the Consonant and Tone Sweep conditions compared to the Vowel and Tone Pitch conditions.

The results fit well with previous findings that this region of posterior left temporal cortex is differentially sensitive to rapidly changing auditory information (Zatorre and Belin, 2001). Clear differences were found in left STG/STS between tasks that involved actively discriminating these dynamic and spectral cues. These results differ from previous studies that have identified regions of STG/STS that showed greater activation when processing speech, compared to nonspeech. However, these studies have typically compared the processing of speech to nonspeech baseline conditions involving steady-state acoustic cues (e.g., Demonet et al., 1992). The present study did observe weaker activation for the Tone Pitch task than for the Consonant task in posterior regions of STG, although signal levels for the Tone Sweep task did not differ from those of the Consonant task in the same regions. The results indicate that acoustic signals capturing the rapid temporal characteristics of speech will tend to show similar activation in STG as is typically found during speech tasks.

Other studies have used nonspeech materials somewhat similar to those used in the present study but have not found the same overlapping results for speech and nonspeech. For example, Scott et al. (2000) observed increased activation in STG when participants listened to intelligible auditory sentences, compared to the same sentences that were modified to destroy intelligibility. One explanation for this is that the speech signals in their study were semantically intelligible, whereas the baseline stimuli were not. This raises the possibility that the obtained differences were due to the contribution of neural regions responsible for semantic processing similar to what was observed by Binder et al. (1997, 2000), rather than regions involved in phonetic processing proper.

The present results are in fact consistent with the finding by Scott et al. (2000) that superior temporal regions were activated when listening both to intelligible speech (intact, or noise-vocoded) and to spectrally rotated speech, but not speech that was both noise vocoded and spectrally rotated. The common denominator among these signals is the presence of rapidly changing acoustic cues. While Scott et al. suggested that this was due to the fact that rotated speech retains some actual *phonetic* content, the results of the present study point to an alternative interpretation, that overlapping regions of STG/STS are engaged during the processing of *any* type of acoustic signal that involves dynamic temporal information. The finding that regions of right STG showed greater activation for the Consonant, Vowel, and Tone Sweep conditions, compared to the Tone Pitch condition (Fig. 4), might in fact be supportive of this claim. These regions appeared to be activated by all acoustic stimulation that involved at least some rapid modulation of frequency information. Like most speech signals, the stimuli in the Vowel condition consisted of both rapid formant sweeps and more steady-state vowel information. The difference between the Vowel and Consonant conditions was the type of information that participants were actively seeking to discriminate. The fact that this region of right STG showed some amount of activation in both of these speech conditions, in addition to the Tone Sweep condition, might reflect its involvement in the automatic processing of all types of auditory stimuli containing rapidly changing constituents, regardless of what type of cue is being actively attended to. This is supported by recent work by Jäncke et al. (2002), which shows stronger activation in specific regions of STG/STS during passive listening to CV syllables, compared to vowels alone.

The present study sought to minimize the influence of lexical and semantic effects as much as possible in order to focus participants' attention on the acoustic and phonetic properties of the stimuli. A limitation of this design was that syllables in the vowel condition did correspond to familiar English words due to the paucity of useable stimuli of this type that were not real words. While further research is necessary to assess the degree to which this affected the present results, it does not appear to undermine the finding of similar activation for the Consonant and Tone Sweep conditions in superior temporal brain regions.

Inferior frontal gyrus

The second anatomical region of interest in this study was the inferior frontal gyrus, which is frequently observed to be active during auditory speech processing. The present study observed significant increases in left IFG activation when participants discriminated the initial consonant of a syllable pair, consistent with results from other studies finding activation in this region during speech discrimination, phoneme monitoring, and rhyme monitoring (Binder et al., 1997; Burton et al., 2000; Demonet et al.,

1992; Zatorre et al., 1992). Nevertheless, the exact role that left IFG plays in these types of tasks remains controversial (Poeppel, 1996). For instance, it has been suggested that left IFG is specifically engaged by higher-order phonological processing, such as what is involved in segmenting the speech stream into phonemes or syllables (Burton et al., 2001). The results of this study are inconsistent with this assertion, given that the stimuli in the Consonant discrimination task could be phonetically discriminated based only on the syllables' initial phonemes (e.g., [ba] and [ga]), and did not require phonological analysis such as segmentation.

The speech–nonspeech contrast did not yield any significant clusters of activation, suggesting that IFG does not represent a speech-specific area of cortex. Instead, this region showed significant activation for both the Consonant and Tone Sweep tasks compared to the Vowel and Tone Pitch tasks, suggesting it is engaged during the processing of any auditory signal that involves rapid temporal cues. This is borne out by previous results finding activation in this region during the processing of nonspeech auditory signals that involve rapidly changing acoustic cues (Johnsrude et al., 1997; Müller et al., 2001). It is also possible that this region's role in auditory processing is more contingent on quantitative effects such as the relative processing demands of a given task. The present study did not directly test the effects of task difficulty, and so it is difficult to fully assess this possibility. However, it is noteworthy that acoustic differences between stimuli in this experiment tended to be much smaller than what is typically used in speech processing tasks, thanks to the use of a synthetic speech continuum rather than natural speech tokens. As a result, participants were much less likely to correctly judge many of the stimuli as different than in other studies in which the observed performance was much more accurate (e.g., Burton et al., 2000, obtained 94–99% accuracy rates.) One hypothesis for why this would tend to result in greater IFG activation draws from previous findings that implicate this region in phonological working memory (Braver et al., 1997; Stowe et al., 1998). Small acoustic differences between stimuli might place a heavier load on working memory in order to maintain accurate representations of the signals for the purpose of discrimination. In contrast, the experiments reported in Burton et al. (2001) involved natural speech stimuli that might not have relied as heavily on working memory capacity when the word pairs differed on only a single phoneme. This might explain why their study only observed left IFG activation in tasks that involved phonological segmentation: this task arguably relied more heavily on working memory resources leading to greater degrees of left IFG activation. That said, this conclusion must be approached with caution given that Burton et al. did not observe accuracy or RT differences between the two tasks, suggesting that participants might have found both task types equally demanding.

Conclusion

The uniqueness and complexity of speech perception have lead some to argue that it is qualitatively different from other cognitive capacities (Liberman and Mattingly, 1985). However, studies to this effect are complicated by the fact that perceiving speech represents the confluence of two separate tasks: recognizing both the phonetic form and the semantic content encoded within an auditory signal. The key question that is addressed in this work regards the extent to which classical language areas in the human brain are uniquely specialized for phonetic processing. The results speak to this question by suggesting that both speech and other signals that share key acoustic characteristics with speech tend to activate similar brain regions. They are consistent with previous neuroimaging results showing that rapidly changing auditory signals will tend to activate classical language areas of the brain, such as the inferior frontal lobe and the posterior portion of the superior temporal gyrus and sulcus. The present work considered only one type of rapid temporal cue, namely the rapid changes in frequency that occur in stop consonants, and further research is needed to determine whether other types of dynamic speech cues will yield similar results.

The results are also consistent with studies of children with developmental speech–language impairments, who have been observed to have deficits in processing the phonetic form of speech (Joanisse et al., 2000; Tallal et al., 1980). This deficit might be related to difficulties processing rapid acoustic signals and suggests that a similar deficit could impair the processing of speech and nonspeech signals containing rapidly changing acoustic features (Tallal and Piercy, 1974; Tallal et al., 1993). The present study tends to support this assertion by indicating a shared neural substrate that is critically involved in processing temporal information in both speech and nonspeech signals.

Appendix

Cortical regions reaching significance for each contrast

XYZ coordinates are in the stereotaxic space of Talairach and Tournoux (1998). Statistical threshold, $t = 3.6$ ($P < 0.0002$, corrected); spatial threshold, 50 contiguous significant voxels.

Contrast/region	*X*	*Y*	*Z*	Cluster size (No. voxels)
Speech > nonspeech				
L STG (BA 22)	−56	−22	3	212
L STG (BA 22)	−51	−14	−4	88
R STG (BA 22)	54	−46	9	461
R STG (BA 22)	52	−25	2	1937
R middle temporal (BA 21)	62	1	−12	97
R fusiform (BA 20)	31	−40	−15	899

Contrast/region	X	Y	Z	Cluster size (No. voxels)
L medial frontal gyrus (BA 9)	−15	42	15	140
R occipital (BA 19)	20	−65	−9	126
R occipital (BA 18)	32	−71	−10	190
Nonspeech > speech				
R STG (BA 22)	57	3	−2	1105
R STG (BA 22)	60	−35	18	1785
R middle temporal (BA 21)	65	−40	−2	124
L precentral gyrus (BA 6)	−56	4	12	161
L angular gyrus (BA 13)	−48	−42	18	182
L STG/insula (BA 21/13)	−39	−7	1	1972
R insula (BA 13)	28	24	6	419
R precentral gyrus (BA 44)	40	7	9	1573
R precentral gyrus (BA 6)	58	4	18	66
R postcentral gyrus (BA 43)	53	−12	18	58
R occipital (BA 17)	6	−92	0	69
R occipital (BA 18)	−36	−89	4	65
R occipital (BA 18)	3	−73	18	194
R occipital (BA 19)	44	−78	−4	109
R occipital (BA 30)	1	−70	6	62
Rapid temporal > spectral				
L STG/middle temporal (BA 21/22)	−63	−28	−1	346
L STG (BA 22)	−48	−31	7	91
R STG (BA 22)	51	−16	5	665
R STG (BA 39)	55	−57	24	150
L IFG (BA 47)	−38	27	−9	145
L middle frontal gyrus (BA 10)	−36	48	−3	54
L occipital (BA 30)	−5	−70	9	93
R occipital (BA 18)	16	−75	−5	706
R occipital (BA 19)	14	−54	−3	72
R occipital (BA 18)	25	−90	19	69
Spectral > rapid temporal				
L STG (BA 22)	−38	−56	18	85
R middle temporal (BA 21)	60	−51	−3	107
L insula (temporal) (BA 13)	−41	−28	18	277
L IFG (BA 44)	−49	3	16	738
R IFG/precentral gyrus (BA 44)	52	9	14	557
L occipital (BA 18)	−27	−75	−5	135
L occipital (BA 18)	−14	−67	18	303
R fusiform (BA 19)	23	−54	−14	411

References

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

Berndt, R., Haedinges, A., Mitchum, C., Wayland, S., 1996. An investigation of nonlexical reading impairments. Cogn. Neuropsychol. 13, 763–801.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S.F., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10, 512–528.

Binder, J.R., Frost, J.A., Hammeke, T.A., Cox, R.W., Rao, S.M., Prieto, T., 1997. Human brain language areas identified by functional magnetic resonance imaging. J. Neurosci. 17, 353–362.

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

Blumstein, S.E., 1998. Phonological aspects of aphasia, in: Sarno, M. (Ed.), Acquired Aphasia, third ed. Academic Press, New York, pp. 157–185.

Blumstein, S.E., Milberg, W., Brown, T., Hutchison, A., Kurowski, K., Burton, M.W., 2000. The mapping from sound structure to the lexicon in aphasia: evidence from rhyme and repetition priming. Brain Language 72, 75–79.

Braver, T.S., Cohen, J.D., Nystrom, L.E., Jonides, J., Smith, E.E., Noll, D.C., 1997. A parametric study of prefrontal cortex involvement in human working memory. NeuroImage 5, 49–62.

Bregman, A.S., 1990. Auditory Scene Analysis. MIT Press, Cambridge, MA.

Burton, M.W., 2001. The role of inferior frontal cortex in phonological processing. Cogn. Sci. 25, 695–709.

Burton, M.W., Small, S.L., Blumstein, S.E., 2000. The role of segmentation in phonological processing: an fMRI investivation. J. Cogn. Neurosci. 12, 679–690.

Dale, A.M., Buckner, R.L., 1997. Selective averaging of rapidly presented individual trials using fMRI. Hum. Brain Mapp. 5, 329–340.

Demonet, J-F., Chollet, F., Ramsay, S., Cardebat, J-L., Wise, R., Rasol, A., Frackowiak, R., 1992. The anatomy of phonological and semantic processing in normal subjects. Brain 115, 1753–1768.

Fiez, J.A., Raichle, M.E., Miezin, F.M., Petersen, S.E., Tallal, P., Katz, W.F., 1995. PET studies of auditory and phonological processing — effects of stimulus characteristics and task demands. J. Cogn. Neurosci. 7, 357–375.

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant change in functional magnetic resonance imaging (fMRI): use of a cluster size threshold. Magn. Reson. Med. 33, 636–647.

Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends Neurosci. 15, 20–25.

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4, 131–138.

Jäncke, L., Wüstenberg, T., Scheich, H., Heinze, H.-J., 2002. Phonetic perception and the temporal cortex. NeuroImage 15, 733–746.

Joanisse, M.F., Manis, F.R., Keating, P., Seidenberg, M.S., 2000. Language deficits in dyslexic children: speech perception, phonology and morphology. J. Exp. Child Psychol. 77, 30–60.

Johnsrude, I.S., Zatorre, R.J., Milner, B.A., Evans, A.C., 1997. Left-hemisphere specialization for the processing of acoustic transients. NeuroReport 8, 1761–1765.

Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. J. Acoust. Soc. Am. 67, 971–995.

Liberman, A.M., Harris, K.S., Hoffman, H., Griffith, B., 1957. The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. 54, 358–368.

Liberman, A.M., Mattingly, I.G., 1985. The motor theory of speech revised. Cognition 21, 1–36.

Massaro, D., 1997. Categorical partition: a fuzzy-logical model of categorization behavior, in: Harnad, S. (Ed.), Categorical Perception: The Groundwork of Cognition, Cambridge Univ. Press, Cambridge, UK, pp. 254–283.

Müller, R-A., Kleinhans, N., Courchesne, E., 2001. Broca's area and the discrimination of frequency transitions: a functional MRI study. Brain Language 76, 70–76.

Ogawa, S., Lee, T-M., Kay, A.R., Tank, D.W., 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. USA 87, 9868–9872.

Poeppel, D., 1996. A critical review of PET studies of phonological processing. Brain Language 55, 317–351.

Poldrack, R.A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M.M., Gabrieli, J.D.E., 2001. Relations between the neural bases of dynamic auditory processing and phonological processing: evidence from fMRI. J. Cogn. Neurosci. 13, 687–697.

Poldrack, R.A., Wagner, A.D., Prull, M., Desmond, J.E., Glover, G.H., Gabrieli, J.D.E., 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. NeuroImage 10, 15–35.

Schlosser, M.J., Aoyagi, N., Fulbright, R.K., Gore, J.C., McCarthy, G., 1998. Functional MRI studies of auditory comprehension. Hum. Brain Mapp. 6, 1–13.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J.S., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406.

Shaywitz, B.A., Pugh, K.R., Constable, R.T., Shaywitz, S.E., Bronen, R.T., Fulbright, R.K., Shankweiler, D.P., Katz, L., Fletcher, J.M., Skudlarski, P., Gore, J.C., 1994. Localization of semantic processing using functional magnetic-resonance-imaging. Ann. Neurol. 36, 504–504.

Stowe, L.A., Broere, C.A.J., Paans, A.M.J., Wijers, A.A., Mulder, G., Vaalburg, W., Zwarts, F.G., Vaalburg, W., 1998. Localizing components of a complex task: sentence processing and working memory. NeuroReport 9, 2995–2999.

Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Thieme, New York.

Tallal, P., Miller, S., Fitch, R.H., 1993. Neuirobiological basis of speech: a case for the preeminence of temporal processing. J. N. Y. Acad. Sci. 682, 27–47.

Tallal, P., Piercy, M., 1974. Developmental aphasia: rate of auditory processing and selective impairment of consonant perception. Neuropsychologia 12, 83–94.

Tallal, P., Stark, R.E., Kallman, C., Mellits, D., 1980. Perceptual constancy for phonemic categories: a developmental study with normal and language impaired children. Appl. Psycholinguistics 1, 49–64.

Utman, J.A., Blumstein, S.E., Sullivan, K., 2001. Mapping from sound to meaning: reduced lexical activation in Broca's aphasics. Brain Language 79, 444–472.

Vouloumanos, A., Kiehl, K., Werker, J.F., Liddle, P., 2001. Neurological bases of speech and non-speech processing: an event-related functional magnetic resonance imaging study. J. Cogn. Neurosci. 13 (7), 994–1005.

Xiong, J., Gao, J., Lancaster, J.L., Fox, P.T., 1995. Clustered pixels analysis for functional MRI activation studies in the human brain. Hum. Brain Mapp. 3, 1–15.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11, 946–953.

Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. 6, 37–46.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Nature 256, 846–849.