

# Human Temporal Lobe Activation by Speech and Nonspeech Sounds

J.R. Binder, J.A. Frost, T.A. Hammeke, P.S.F. Bellgowan, J.A. Springer, J.N. Kaufman and E.T. Possing

Department of Neurology and Department of Cellular Biology, Neurobiology, and Anatomy, Medical College of Wisconsin, Milwaukee, WI, USA

**Functional organization of the lateral temporal cortex in humans is not well understood. We recorded blood oxygenation signals from the temporal lobes of normal volunteers using functional magnetic resonance imaging during stimulation with unstructured noise, frequency-modulated (FM) tones, reversed speech, pseudowords and words. For all conditions, subjects performed a material-nonspecific detection response when a train of stimuli began or ceased. Dorsal areas surrounding Heschl's gyrus bilaterally, particularly the planum temporale and dorsolateral superior temporal gyrus, were more strongly activated by FM tones than by noise, suggesting a role in processing simple temporally encoded auditory information. Distinct from these dorsolateral areas, regions centered in the superior temporal sulcus bilaterally were more activated by speech stimuli than by FM tones. Identical results were obtained in this region using words, pseudowords and reversed speech, suggesting that the speech-tones activation difference is due to acoustic rather than linguistic factors. In contrast, previous comparisons between word and nonword speech sounds showed left-lateralized activation differences in more ventral temporal and temporoparietal regions that are likely involved in processing lexical-semantic or syntactic information associated with words. The results indicate functional subdivision of the human lateral temporal cortex and provide a preliminary framework for understanding the cortical processing of speech sounds.**

## Introduction

Several lines of evidence suggest that the superior temporal cortex in humans plays a vital role in speech sound processing. This region contains the cortical representation of the auditory sensory system (Wernicke, 1874; Flechsig, 1908). Bilateral lesions restricted to the superior temporal lobes result in the syndrome of 'pure word deafness', characterized by a relatively isolated impairment of speech sound discrimination (Barrett, 1910; Henschen, 1918-1919; Buchman *et al.*, 1986; Tanaka *et al.*, 1987). More recently, a large number of studies using positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have demonstrated activation of the superior temporal cortex bilaterally when subjects are presented with speech sounds in contrast to no sounds (Petersen *et al.*, 1988; Wise *et al.*, 1991; Howard *et al.*, 1992; Mazoyer *et al.*, 1993; Binder *et al.*, 1994b; Fiez *et al.*, 1995, 1996; O'Leary *et al.*, 1996; Price *et al.*, 1996b; Warburton *et al.*, 1996; Dhankhar *et al.*, 1997; Grady *et al.*, 1997; Hirano *et al.*, 1997; Jäncke *et al.*, 1998).

What has not emerged from these observations is a more detailed picture of the functional organization of the human superior temporal region. From an information processing perspective, the perception of speech involves several stages of auditory sensory analysis and pattern extraction as well as interactions between sensory and stored linguistic information (McClelland and Elman, 1986; Klatt, 1989). These analysis stages imply a degree of hierarchical functional organization in the

speech perceptual system analogous to the levels of processing observed in ventral visual recognition pathways. From an anatomical and physiological standpoint, there is ample evidence in nonhuman primates for subdivision of the superior temporal region into a large number of distinct areas (Galaburda and Pandya, 1983; Morel *et al.*, 1993; Rauschecker *et al.*, 1995; Hackett *et al.*, 1998). Although available anatomical and physiological data are consistent with some degree of functional subdivision of the human superior temporal cortex (Celesia, 1976; Galaburda and Sanides, 1980; Liegeois-Chauvel *et al.*, 1991; Rademacher *et al.*, 1993; Howard *et al.*, 1996), the precise correspondence between human and nonhuman auditory cortical areas remains unclear. The human equivalent of primary auditory cortex (area AI) is largely confined to the caudal one-third of the transverse temporal or Heschl's gyrus (HG) (Celesia, 1976; Galaburda and Sanides, 1980; Liegeois-Chauvel *et al.*, 1991; Rademacher *et al.*, 1993), suggesting a role for this region in early sensory analysis of speech sounds. Little else is currently known regarding the relationships between putative subdivisions of the human temporal lobe and the multiple processing stages proposed to underlie speech perception.

A first step would be to identify cortical regions that are selectively responsive to speech sounds in contrast to simpler sounds. Just as certain visual regions are activated by specific stimulus attributes like motion or color, it is possible that certain auditory areas become active only in the presence of speech sounds, which contain a variety of characteristic physical attributes (Lieberman *et al.*, 1967). Four functional imaging studies comparing processing of speech and simpler nonspeech sounds support this concept. In these studies, speech sounds elicited significantly stronger responses than amplitude-modulated noise (Zatorre *et al.*, 1992) or tone sequences (Démonet *et al.*, 1992; Binder *et al.*, 1996, 1997) in a region of the superior temporal gyrus (STG) and adjacent superior temporal sulcus (STS) located lateral and ventral to HG and clearly distinct from human AI. While consistent with a functional subdivision of the superior temporal region, these findings raise other, as yet unresolved, issues.

One issue concerns the nature of the activation response elicited in these studies. Because the speech sounds differed from the nonspeech sounds in terms of both auditory complexity and phonetic content, it is not clear whether the differential activation by speech sounds is due to complex auditory processing, phoneme processing or both. Moreover, the speech stimuli in most of these studies included lexical items, raising the additional possibility that the speech activation could reflect processes pertaining to lexical access or semantic retrieval. To address this problem, several investigators contrasted listening to words with listening to meaningless speech sounds (pseudowords, syllables or reversed speech) that were closely matched to the words in terms of physical attributes. By matching on

auditory processing demands, these comparisons should identify brain regions involved in processing post-sensory, linguistic representations (i.e. lexical and semantic knowledge) associated with the word stimuli. Unfortunately, no clear result has emerged from these experiments, some of which showed no superior temporal lobe differences between conditions (Wise *et al.*, 1991; Démonet *et al.*, 1992, 1994; Hirano *et al.*, 1997; Binder *et al.*, 1999), and others of which identified superior temporal foci but in inconsistent locations (Howard *et al.*, 1992; Mazoyer *et al.*, 1993; Perani *et al.*, 1996; Price *et al.*, 1996b).

A second issue concerns the relationship of this 'speech-specific' region in lateral STG and STS to conventional neuroanatomical models of auditory word processing. The conventional view, based on lesion-deficit correlation and anatomical studies in humans, emphasizes the importance of the posterior left STG ('Wernicke's area') and posterior temporal plane, or planum temporale (PT), in auditory word perception (Wernicke, 1874; Geschwind and Levitsky, 1968; Geschwind, 1971; Bogen and Bogen, 1976; Braak, 1978; Galaburda *et al.*, 1978; Mesulam, 1990; Foundas *et al.*, 1994). Relative to these classical areas, the 'speech-specific' region identified by functional imaging is located somewhat anteriorly and ventrally, suggesting the possibility of an auditory processing stream that projects ventrolaterally from AI rather than posteriorly (Binder *et al.*, 1996). The functional imaging data suggest, furthermore, that speech-responsive cortex exists in both left and right superior temporal regions (Démonet *et al.*, 1992; Zatorre *et al.*, 1992; Binder *et al.*, 1997), in contrast to the conventional emphasis on functional lateralization to the language-dominant hemisphere. Taken as a whole, the functional imaging data are reminiscent of observations in nonhuman primate auditory cortex, which suggest a hierarchical, ventrolaterally directed flow of information from AI to STS in the superior temporal region bilaterally (Galaburda and Pandya, 1983; Hackett *et al.*, 1998; Rauschecker, 1998).

We addressed these issues in the following experiments by using blood oxygenation level dependent (BOLD) fMRI to measure superior temporal responses to speech and non-speech sounds. Our first aim was to identify cortical regions responsive to simple acoustic temporal structure in comparison to incoherent noise. BOLD responses to frequency-modulated (FM) tones were compared with responses elicited by unmodulated noise in 28 normal, right-handed subjects. The second aim was to identify regions that are selectively activated by speech sounds relative to these less complex nonspeech sounds. Accordingly, BOLD responses to speech sounds were measured concurrently in the same group of subjects and compared with responses elicited by the FM tones. Our third aim was to determine the sensitivity of these 'speech-specific' regions to post-sensory linguistic variables. BOLD responses to words were compared with responses elicited by phonetically familiar nonwords ('pseudowords') and to responses elicited by phonetically unfamiliar nonwords created by temporal reversal of speech sounds ('reversed speech'). Our rationale was that the word-pseudoword contrast would test whether the regions characterized as 'speech-specific' are sensitive to lexical-semantic attributes of input stimuli, and that the speech-reversed speech contrast would test sensitivity of this region to phonetic attributes of the input stimuli. As noted above, previous word-nonword comparisons have not demonstrated consistent results with respect to the superior temporal region. Many of these studies incorporated explicit language processing tasks that may have caused activation differences due to differences in

attentional state (Howard *et al.*, 1992; Perani *et al.*, 1996; Price *et al.*, 1996b). To avoid this potential confound, our subjects performed a simple, non-linguistic, detection response for all stimuli, such that explicit attentional requirements were uniform across conditions. Quantitative comparisons were made between activated regions in the left and right temporal lobe. Through meta-analysis of standard stereotaxic coordinate data, activation foci resulting from each contrast were compared with those identified in previous PET and fMRI studies.

## Materials and Methods

### *Pilot Study: Reversed Speech Transcription*

Reversed speech (i.e. backward playback of recorded speech waveforms) has been used frequently in functional imaging research (Howard *et al.*, 1992; Perani *et al.*, 1996; Price *et al.*, 1996b; Dehaene *et al.*, 1997; Hirano *et al.*, 1997). Reversal produces nonword stimuli that match the originals in terms of physical complexity and acoustic characteristics and are therefore ideal as controls for acoustic input. While it seems unlikely that such stimuli evoke specific lexical or semantic associations, the amount of phonetic information conveyed by reversed speech sounds is unclear. Many phonemes are temporally symmetrical or relatively symmetrical (fricatives and long-duration vowel sounds), or show approximate mirror reversal of formant transition structure before and after a vowel, and so might convey considerable phonetic information even in reversed form. On the other hand, reversal produces other sound sequences that do not occur in naturally uttered speech and so strike the listener as unfamiliar. The purpose of this pilot study was to explore the degree to which listeners perceive and reliably report familiar phonemes on hearing reversed speech sounds.

Participants were six men and four women (age range 24–48 years, mean 31 years) who were monolingual English speakers with no training in phonetics or audiology and no history of neurologic or audiologic symptoms. Testing was performed in a quiet room. Digitally recorded stimuli were presented at clearly audible levels through standard headphones. The duration of all stimuli was edited to 666 ms using cutting and duplication techniques (for fricative and pause segments) together with digital time compression and expansion without altering pitch (SoundDesigner II software, Digidesign Inc., Menlo Park, CA). The stimuli included:

- Words: monosyllabic, medium frequency, concrete English nouns (e.g. desk, fork, stream) spoken by a male.
- Pseudowords: monosyllabic, pronounceable nonwords, derived by rearranging the phoneme order of the word stimuli (e.g. /sked/, /korf/, /rimst/). These were recorded concurrently with the word stimuli by the same talker and were matched to the words in terms of duration, overall spectral content, overall intensity, and phoneme content.
- Reversed speech ('Reversed'): temporally reversed versions of the word stimuli.

There were 48 stimuli of each type, and each stimulus was presented once, for a total of 144 trials. Stimuli of a given type were presented in blocks of six trials (thus eight blocks per condition), and these blocks were presented in a randomized condition order.

After each stimulus, the subject attempted to transcribe what had been presented by typing into a computer keyboard. No time constraints were used. The typed string was visible at all times, but the subject was instructed not to change a letter that had been typed except in the event of an accidental key press. Subjects were instructed to make the most accurate and complete transcription possible, and to guess or approximate when necessary. When satisfied with a response, the subject advanced to the next trial using the enter key.

The aim of data analysis was to estimate the degree of consistency of transcription responses across subjects for each condition. Our assumption was that if the sounds were readily assignable to phoneme categories (i.e. phonetically familiar), then both the number of letters and the choice of letters used in transcription should be relatively consistent across subjects. For each stimulus, an automated program determined the

most common number of letters used for transcription, as well as the most commonly selected letter for each letter position. Using these 'canonical' responses, the proportion of subjects in agreement with the canonical response was then computed. Most English phonemes have several acceptable alternative spellings, which may vary in length (e.g. c, k or ck for /k/). We therefore computed the percent agreement based on criteria that did not require a perfect match with the canonical response. For the letter count measurement ('Length'), any response with a letter count of  $\pm 1$  relative to the canonical was accepted as a match. For the letter choice measurement ('Letter'), a letter response could vary by  $\pm 1$  letter position relative to the canonical. Although only rough approximations, these measures were assumed to reflect the degree of transcription consistency across subjects, and thus the degree of consistency of phonetic categorization of the presented sounds. It was anticipated that performance on the words, which have known standard spellings, would be virtually perfect. Of interest was any difference that would arise between pseudowords and reversed speech, which would be most readily attributable to differences in phonetic familiarity.

### **fMRI Subjects**

Subjects in the fMRI study were 13 men and 15 women, ranging in age from 19 to 41 years (mean 26.0 years), with no history of neurologic or audiological symptoms. All subjects were monolingual English speakers, and all indicated right hand dominance (laterality quotient  $> 33$ ) on the Edinburgh Handedness Inventory (Oldfield, 1971). After a full explanation of the nature and risks of the research, subjects gave written informed consent for all studies according to a protocol approved by the institutional research review board.

### **Image Acquisition**

Imaging was performed on a 1.5 T General Electric Signa scanner (GE Medical Systems, Milwaukee, WI) using three-axis local gradient and insertable transmit-receive radiofrequency coils (Medical Advances, Milwaukee, WI) designed for echo-planar imaging (EPI) (Wong *et al.*, 1992). Sagittally oriented, gradient-echo EPI images were acquired for functional studies using the following conditions: echo time = 40 ms, repetition time = 4000 ms, field of view = 240 mm and in-plane voxel dimensions =  $3.75 \times 3.75$  mm. Sequential images were collected concurrently at contiguous sagittal locations in the lateral portion of each hemisphere using spatially interleaved acquisitions. Slice thickness (7–8 mm) and number of slices (3–6 per hemisphere) varied across subjects, although for a given subject the number of slices was the same in the left and right hemispheres. Care was taken to position the slices at the same location relative to the lateral brain edge in both hemispheres. To minimize overall scanner noise levels, the minimal number of slices needed to cover the temporal lobes was acquired, using a relatively long interscan interval (i.e. repetition time/number of slices). While slice coverage varied by a few millimeters from subject to subject, brain tissue located between stereotaxic *x* coordinates 42 and 62 was imaged in both hemispheres of all subjects. This volume includes HG, superior temporal plane, STS and lateral temporal lobe, excluding part of the lateral temporal surface lying midway along the length of the STS, and the medial tip of HG (Talairach and Tournoux, 1988). Also included in this slice selection are the middle and inferior temporal gyri, the lateral half of the fusiform gyrus, inferior frontal gyrus, supramarginal and angular gyri, and lateral occipital cortex. Two EPI series with 168 repetitions each were acquired and subsequently concatenated after removing initial signal equilibration images. The two final images in each series were obtained using slightly modified echo time to provide phase data for the generation of B-field maps, which were used for automated image unwarping. High-resolution anatomical images of the entire brain were obtained during the same session using either a fast-spin-echo sequence yielding contiguous 4 mm sagittal slices (six subjects) or a 3-D spoiled-gradient-echo sequence ('SPGR', GE Medical Systems, Milwaukee, WI) yielding contiguous 1.2 mm sagittal slices (22 subjects).

### **Stimuli and Task**

Room lights were dimmed during scanning, and subjects were instructed to keep their eyes closed. Stimuli were 16-bit digital sound files sampled at 44.1 kHz and low-pass filtered at 11 kHz. These sounds were presented binaurally using a computer playback system, a magnetically shielded

transducer system and plastic sound conduction tubes (Binder *et al.*, 1995). The conducting tubes were threaded through tightly occlusive ear inserts that attenuated the average sound pressure level (SPL) of the continuous scanner noise by  $\sim 20$  dB. Average intensity of the experimental stimulus was 100 dB SPL and was matched across all conditions. Average intensity of the scanner noise was  $\sim 75$  dB SPL after attenuation by the ear inserts.

Five types of auditory stimulus were compared (Fig. 1). Stimuli of a single type were presented during activation periods lasting 12 s. Activation periods were always preceded and followed by baseline periods each lasting 12 s, during which only the background scanner noise was heard. There were eight activation periods for each of the five conditions (40 activation periods in all). The order of these conditions was randomized, and all conditions were presented before a condition was repeated. The duration of all stimuli was 666 ms. All stimuli had 10 ms linear on and off intensity envelopes, and the inter-stimulus interval was 0 ms. There were thus 18 individual stimuli presented during each 12 s activation period and a total of 144 stimuli in each condition over the course of the experiment.

The stimuli included:

- Noise: a 'white' noise burst containing all frequencies from 20 to 11 000 Hz. Except for the on and off envelopes every 666 ms, this repeating stimulus contained no coherent frequency or amplitude modulation.
- Tones: pure sine wave tones having a constant frequency within the range 50–2400 Hz. Forty-eight different, harmonically unrelated, frequencies were used within this range; thus each frequency was presented three times during the course of the experiment. Because a new tone was presented every 666 ms, the frequency changed randomly and by at least 10 Hz in a stepwise manner every 666 ms.
- Words: monosyllabic, medium frequency, concrete English nouns (e.g. desk, fork, stream) spoken by a male. These were the same word stimuli used in the pilot study. All of the acoustic energy of these sounds was below 11 000 Hz, with the majority between 50 and 3000 Hz. All stimuli were edited to a duration of 666 ms. Forty-eight different words were used; thus each was presented three times during the course of the experiment.
- Pseudowords: monosyllabic, pronounceable nonwords, derived by rearranging the phoneme order of the word stimuli (e.g. /sked/, /korf/, /rimst/). These were the same pseudoword stimuli used in the pilot study. They were recorded concurrently with the word stimuli by the same talker. These stimuli were matched to the words in terms of duration, overall spectral content, overall intensity and phoneme content. Forty-eight different stimuli were used; thus each was presented three times during the course of the experiment.
- Reversed speech ('Reversed'): temporally reversed versions of the word stimuli. These were the same reversed stimuli used in the pilot study. Temporal reversal results in sounds that are, on average, less readily categorized phonetically (see Pilot Data, below). These stimuli were thus matched to the word and pseudoword stimuli on overall spectral content, intensity, duration and acoustic complexity, but conveyed less phonetic information and presumably very little lexical or semantic information. Forty-eight different stimuli were used; thus each was presented three times during the course of the experiment.

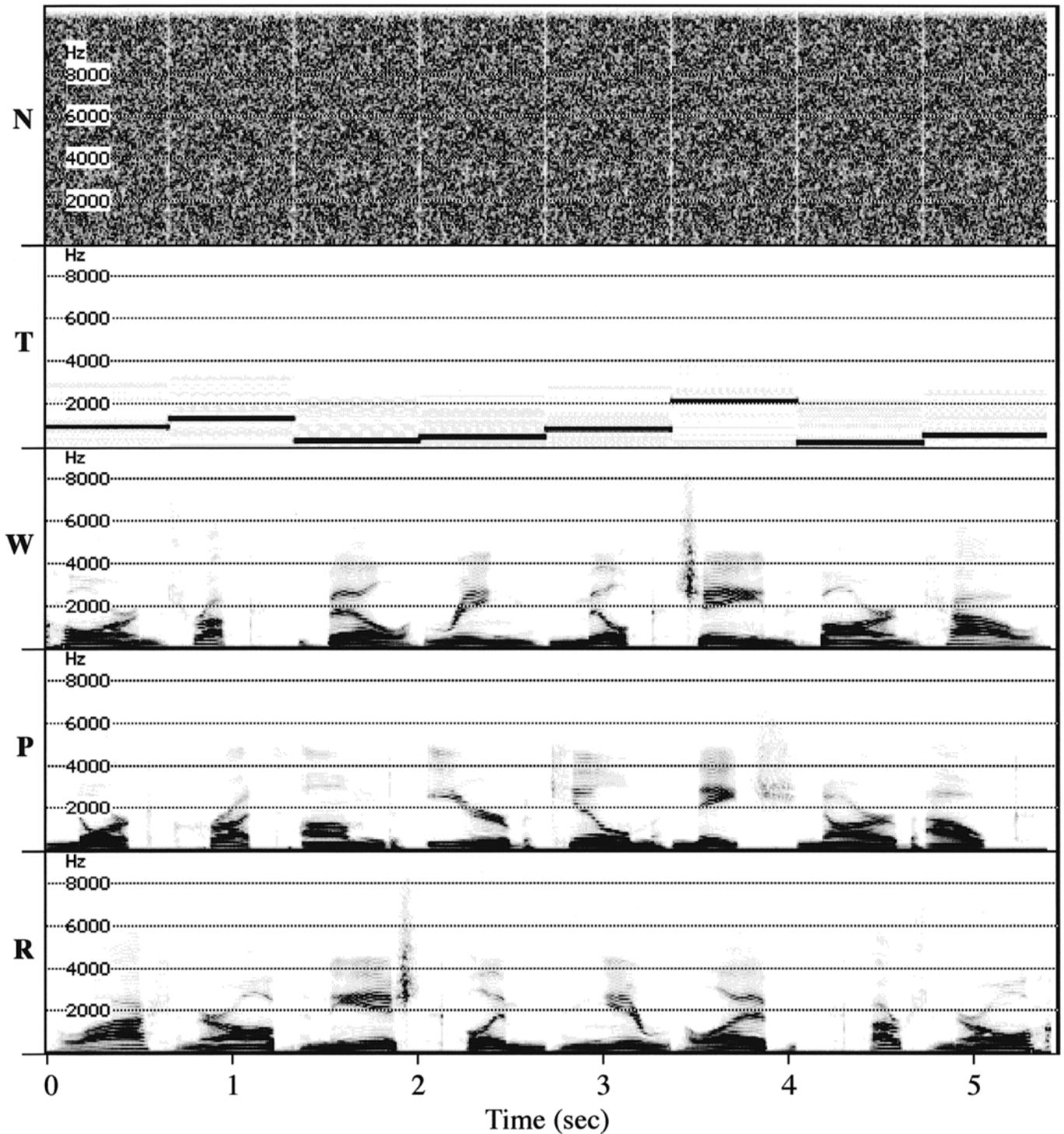
In summary, the stimulus types differed along several dimensions, as shown in Table 1.

Subjects were instructed to listen to the sounds being played through the eartubes and to briefly press a button with the left hand whenever a 12 s block of sounds began or ceased.

### **Within-subject Data Analysis**

All image analysis was performed with the AFNI software package (Cox, 1996a). Automated image unwarping was performed using field maps obtained from the final two images in each EPI series (Weisskoff and Davis, 1992). All images were then spatially co-registered (to the first steady-state EPI image acquired after the anatomical scan) using an iterative procedure that minimizes variance in voxel intensity differences between images (Cox, 1996b).

Activation–baseline difference (Diff) maps were created for each of the 40 activation periods as follows. First, the final two images from each



**Figure 1.** Example narrow-band spectrograms (plots of frequency spectra as a function of time) of the five stimulus types. Each row includes eight stimuli of each type. N = Noise, P = Pseudowords, R = Reversed, T = Tones, W = Words.

activation period were averaged to produce an image of average signal intensity values during the last 8 s of each period. Next, the final two images from the baseline periods immediately preceding and following each activation period were averaged, and a Diff image was created for each activation period by subtracting the average baseline image from the corresponding average activation image (Binder *et al.*, 1994a).

Pooled-variance Student's *t*-tests of the condition-specific Diff maps

were then used to create three-dimensional statistical parametric maps (SPMs) reflecting differences between conditions at each voxel location for each subject. SPMs were created for each of the five conditions compared with baseline by testing the Diff map values at each voxel against a hypothetical mean of zero. SPMs were also created for seven paired activation condition contrasts: Tones-Noise, Reversed-Tones, Pseudowords-Tones, Words-Tones, Pseudowords-Reversed, Words-

**Table 1**

Stimulus characteristics

Stimulus	Physical attributes			Associated representations			
	Frequency bandwidth	AM and FM	Diversity	Pitch	Phonemic	Lexical	Semantic
Noise	20–11000	–	–	–	–	–	–
Tones	50–2400	+	+	+	–	–	–
Reversed	50–3000	+	+	+	+	–	–
Pseudowords	50–3000	+	+	+	+	+	–
Words	50–3000	+	+	+	+	+	+

The degree to which each stimulus is characterized by a particular physical attribute or set of associated representations is indicated by '–' (not characteristic), '+' (characteristic) or '+ +' (highly characteristic).

Reversed and Words–Pseudowords. No assumptions were made about the direction of effects in the condition contrasts, thus all contrasts were tested and are reported for both directions.

The volume of significantly activated tissue in left and right superior temporal regions was determined for each condition contrast in each subject as follows. First, each SPM was thresholded at a  $t$  value of 3.5 (uncorrected  $P < 0.01$ ), and clusters of activation containing fewer than four raw data voxels were removed. These thresholds reduce the voxelwise probability of false positives to approximately  $P < 0.0001$ , as determined by Monte Carlo simulation using the 'AlphaSim' module of AFNI. Next, these thresholded SPMs were superimposed on anatomical images from the same subject, and the activated voxels were classified anatomically by reference to sulcal landmarks. Voxels were considered to lie within the superior temporal region if they fell within the STG (including HG and PT), STS (including lower bank of STS) or planum parietale (the lower bank of the ascending ramus of the sylvian fissure). The number of such voxels was determined for each hemisphere and multiplied by the voxel volume to estimate the total volume of activated superior temporal tissue in each hemisphere.

#### Stereotaxic Averaging

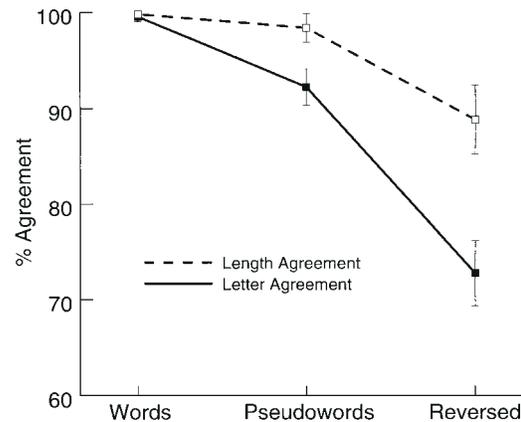
Individual anatomical scans and unthresholded SPMs were also projected into the standard stereotaxic space of Talairach and Tournoux (Talairach and Tournoux, 1988). To compensate for normal variation in anatomy across subjects (Toga *et al.*, 1993), the unthresholded, stereotaxically resampled SPMs were smoothed slightly with a Gaussian filter measuring 2 mm full-width-half-maximum. These datasets were then merged across subjects by averaging the  $t$ -deviates at each voxel (Binder *et al.*, 1997). The procedure of averaging statistics was chosen to guard against heteroscedasticity of MR signal variance among subjects that could arise, for example, from differing degrees of subject motion or tissue pulsatility, variability in global blood flow or reactivity, or scanner variability between sessions.

The maps of averaged  $t$ -statistics were thresholded to identify voxels in which the mean change in MR signal between comparison conditions was unlikely to be zero. Because the average of a set of  $t$ -deviates is not a tabulated distribution, the Cornish–Fisher expansion of the inverse distribution of a sum of random deviates was used to select a threshold for rejection of the null hypothesis (Fisher and Cornish, 1960). Only average  $t$ -scores of 0.909 or greater were considered significant (voxelwise  $P < 0.0001$ ). To further guard against false-positive findings, activation foci had to attain a contiguous volume of at least 15  $\mu$ l. These thresholds reduced the voxelwise probability of false positives to  $\sim P < 10^{-8}$  and the overall probability of false-positive voxels in each 3-D image to  $P < 0.001$ .

## Results

### Pilot Study: Speech Sound Transcription

Subjects in the transcription study uniformly described a sense of uncertainty and unfamiliarity regarding the reversed speech sounds. Transcriptions of these stimuli varied considerably across subjects. For example, responses to the reversed version of 'desk' included 'excit', 'hisech' and 'schah', and responses to the reversed version of 'milk' included 'heckloineb', 'lem',



**Figure 2.** Results of the pilot study on speech sound transcription. Error bars indicate 95% confidence intervals.

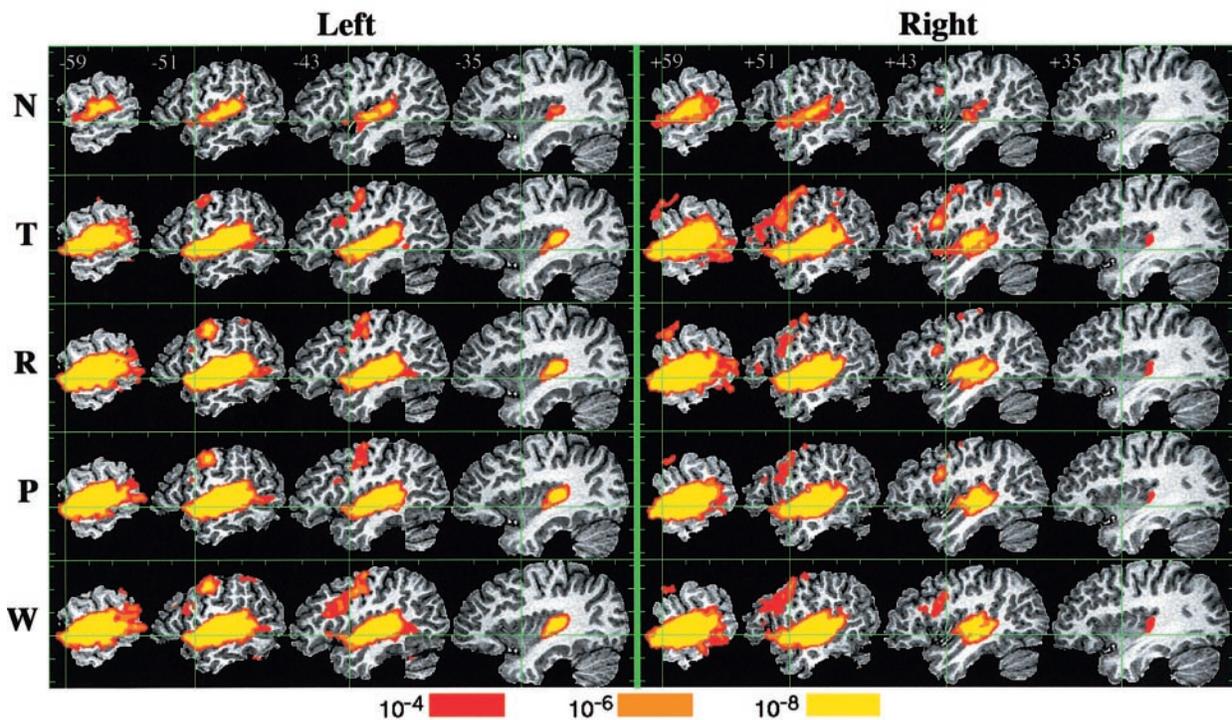
'koim' and 'loycoym'. Time taken to transcribe the sounds varied significantly as a function of stimulus condition (ANOVA with  $df$  2,27;  $F = 22.67$ ,  $P < 0.0001$ ), with Reversed transcription taking significantly longer (mean for all subjects = 8344 ms) than Pseudoword transcription (mean = 6048 ms;  $P = 0.0028$ ) and also significantly longer than Word transcription (mean = 3653 ms;  $P < 0.0001$ ).

Response consistency analyses are shown in Figure 2. As expected, both Length and Letter consistency were close to 100% for Word stimuli. Consistency on the Length measure (number of letters  $\pm 1$ ) varied significantly across conditions (ANOVA with  $df$  2,141;  $F = 28.684$ ,  $P < 0.0001$ ). Length consistency for Reversed stimuli was significantly less than for both Pseudoword and Word conditions ( $P < 0.0001$  for both comparisons). Pseudoword and Word conditions did not differ on this measure. Consistency on the Letter measure (position-specific letter choice  $\pm 1$  position) also varied significantly across conditions (ANOVA with  $df$  2,141;  $F = 149.508$ ,  $P < 0.0001$ ). Letter consistency for Reversed stimuli was significantly less than for both Pseudoword and Word conditions, and Pseudoword consistency was significantly less than Words consistency ( $P < 0.0001$  for all comparisons).

Despite these differences in transcription consistency, it is notable that the percent agreement on the Reversed stimuli was clearly better than chance for the Letter measure (72.7%). Thus, considerable phonetic information was conveyed by the stimuli even if subjects, on average, categorized these sounds less reliably.

### fMRI Task Performance

The listening task used during fMRI, which required a key press



**Figure 3.** Contrasts between stimulus and baseline. Sagittal sections through stereotaxic space show thresholded activation data superimposed on a representative brain. The stereotaxic  $x$  coordinate (left–right location) of each section is given in the top row. Green axis lines on each section indicate the AC–PC plane (stereotaxic  $y$  axis, horizontal line) and the vertical AC plane (stereotaxic  $z$  axis, vertical line). Tick marks appear at 20 mm intervals. Condition abbreviations are as in Figure 1.

after the beginning and after the end of each block of sounds, was performed well by all subjects. A response was correctly produced on 99.4% of trials (range 94% to 100%). The number of missed responses did not differ across the five stimulus conditions (ANOVA with  $df$  4,135;  $F$  = 1.066, *n.s.*).

#### **Averaged Activation Maps: Stimulus versus Baseline**

Averaged stimulus–baseline SPMs (Fig. 3) represent the activation induced by presentation of an experimental stimulus relative to ongoing activation from the background scanner sound. All stimuli produced extensive activation of the superior temporal region bilaterally. The dorsomedial boundary of this activated region was clear: while typically involving most of the dorsal temporal plane, activation did not spread medially into insular cortex. Posterior and ventral boundaries were less definite and were more dependent on stimulus type. Activated regions in the right temporal lobe were shifted somewhat anteriorly and laterally relative to left temporal activations. For example, in all conditions the activation along HG and PT extended farther medially in the left hemisphere, while activation of lateral STG and planum polare was more prominent in the right hemisphere. The stimuli also produced variable degrees of frontal lobe activation bilaterally.

#### **Noise**

Activated areas primarily included HG and surrounding dorsal plane bilaterally, with some spread to lateral STG. There were small right frontal activation foci near the ventral precentral sulcus (BA 6/44) and in pars triangularis of the inferior frontal gyrus (BA 45).

#### **Tones**

Dorsal plane and HG activation was stronger and more extensive than with Noise, particularly along the anterior dorsal plane (planum polare) in the right hemisphere. Activation also extended ventrolaterally into lateral STG and STS along most of the length of the STG. Activation was observed extending, finger-like, into the posterior MTG bilaterally; this was more prominent in the right hemisphere. Frontal foci were observed in the precentral gyrus (BA 6), ventral precentral sulcus (BA 6/44) and pars opercularis (BA 44) of the left hemisphere, and in the precentral gyrus, ventral precentral sulcus, pars opercularis and pars triangularis of the right hemisphere.

#### **Reversed**

Activation extended more ventrally than with Tones, more clearly involving the middle portion of STS bilaterally. Caudal projections into posterior left MTG were more prominent. Frontal lobe foci were similar to those observed with Tones.

#### **Pseudowords**

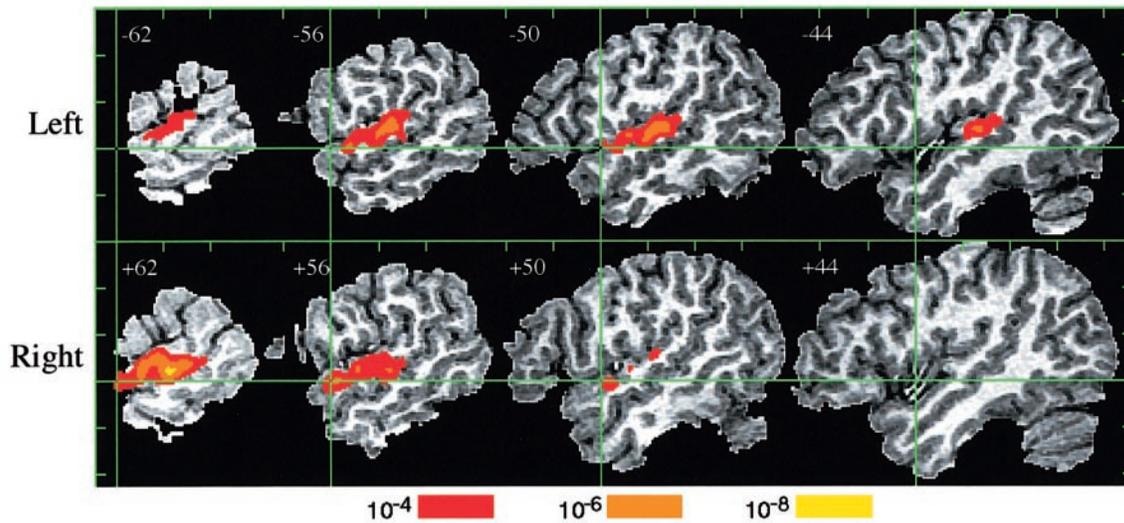
The activation pattern was approximately identical to that of Reversed.

#### **Words**

The activation pattern was approximately identical to Reversed and Pseudowords, except that activation in the ventral precentral sulcus and pars opercularis (BA 44 and 6) appeared somewhat more extensive bilaterally.

#### **Simple Temporal Structure: Tones versus Noise**

Direct comparisons between stimulus conditions were



**Figure 4.** Contrast between Tones and Noise. Left and right hemisphere slices appear in the top and bottom row, respectively. Axis and labeling conventions are otherwise as in Figure 3. Areas responding more to Tones than to Noise are located on the dorsal and dorsolateral surface of the STG and lie predominantly above the AC–PC plane.

**Table 2**  
Location of activation peaks for stimulus contrasts

Anatomical structure	Side	Contrast											
		T > N			W > T			P > T			R > T		
		x	y	z	x	y	z	x	y	z	x	y	z
Posterolateral STG	L	-53	-39	+15									
Planum temporale	L	-45	-33	+12									
	L	-59	-32	+14									
	L	-53	-26	+9									
	R	+48	-23	+12									
Mid-STS	L				-52	-42	+6	-53	-43	+6			
	L				-59	-33	+3	-57	-33	+4	-61	-33	+4
	L				-58	-27	0	-54	-26	0	-56	-24	+1
	L				-53	-14	-1	-54	-13	0	-53	-12	0
	L				-58	-7	-1	-56	-6	-1	-56	-6	-1
	R				+58	-32	+3	+56	-30	+4	+58	-30	+4
	R	+62	-23	+4									
	R				+58	-16	-2	+58	-13	+1	+59	-12	+2
	R				+60	-7	-4	+60	-7	-4	+60	-5	-5
Dorsolateral STG	L	-52	-12	+5									
	L	-54	-8	0									
	R	+62	-17	+10									
	R	+60	-11	+6									
	R	+58	-3	+1									

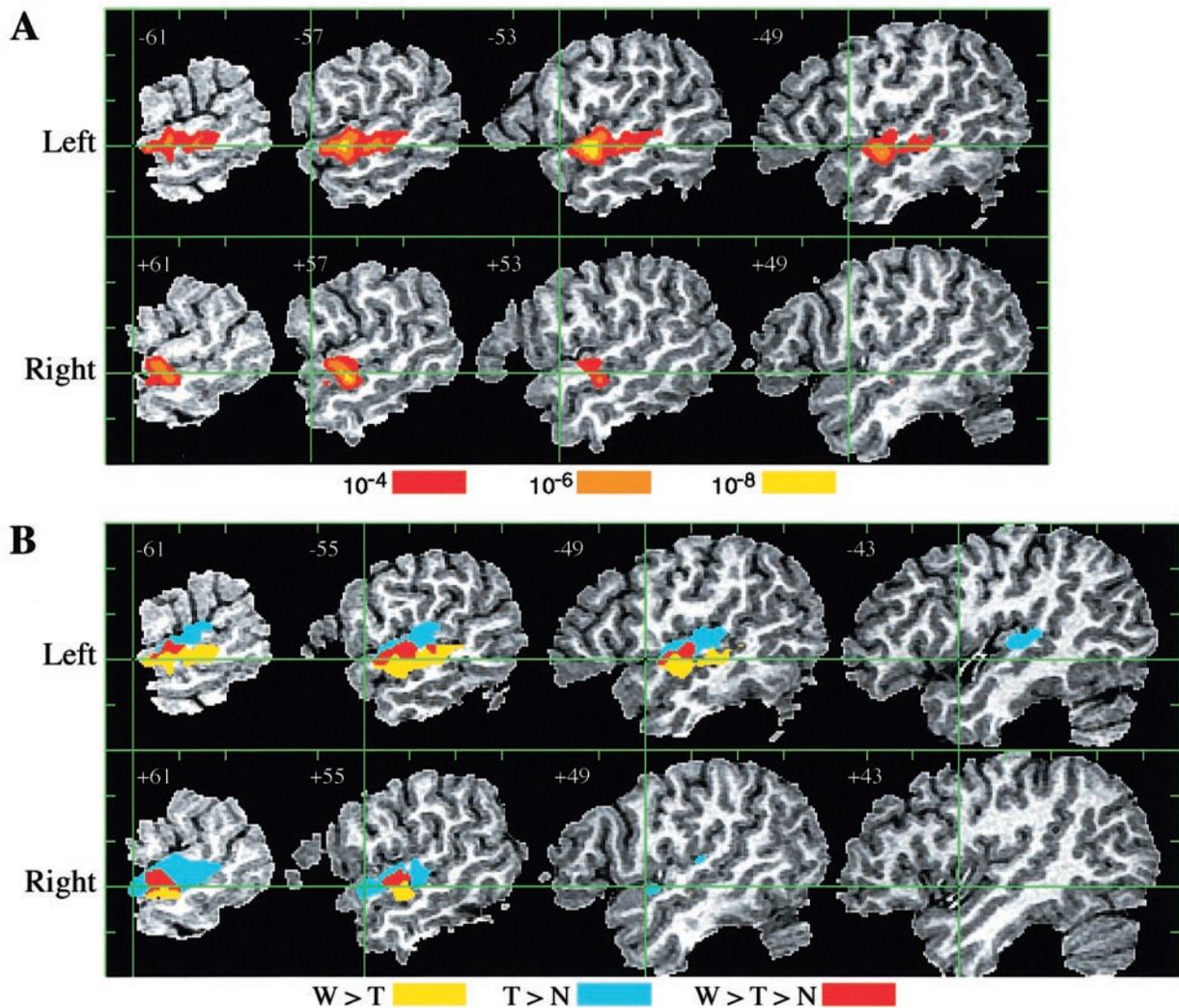
Coordinates and anatomical labels refer to the stereotaxic atlas of Talairach and Tournoux (Talairach and Tournoux, 1988). N = Noise, P = Pseudowords, R = Reversed, T = Tones, W = Words.

performed to more clearly delineate areas responding differently to different stimuli. Representative SPMs are displayed in Figures 4–6, and the location of activation peaks for each comparison are given in Table 2.

The Noise and Tones stimuli differ in two important respects (see Table 1). First, the frequency bandwidth of the Noise stimuli is broader and includes much higher frequencies than the Tones stimuli. Second, a prominent feature of the Tones stimuli is a stepwise frequency modulation every 666 ms, whereas the Noise stimuli exhibit no dominant frequency and no coherent frequency modulation. If frequency bandwidth is a major factor controlling auditory cortex activation, then the Tones–Noise contrast should reveal areas that are more responsive to Noise than Tones. Alternatively, if auditory cortex is

very sensitive to temporal structure in the acoustic signal, then the Tones–Noise contrast should reveal areas that are more responsive to Tones.

No areas showed significantly stronger responses to Noise than Tones. Areas responding more strongly to Tones included much of the dorsal plane and lateral surface of STG bilaterally and a portion of the right STS (Fig. 4). Peak foci for this contrast surrounded the medial HG (primary auditory area) posteriorly, laterally and anteriorly (Table 2), but the primary auditory area itself was not greatly involved. The preference for Tones was strongest posterior to HG (in planum temporale) in the left hemisphere and strongest anterolateral to HG (in anterolateral STG and planum polare) in the right hemisphere. No frontal areas showed significant differences between Tones and Noise.



**Figure 5.** Contrast between Words and Tones. Style and labeling conventions are as in Figure 4. (A) Areas responding more to Words than to Tones are located primarily in or near the STS and lie on or below the AC-PC plane. (B) Comparison between areas identified in the Words-Tones (yellow) and Tones-Noise (light blue) contrasts.

#### **Speech-sensitive Areas: Words versus Tones**

The Words-Tones contrast was intended to reveal brain regions that are more responsive to speech than to nonspeech sounds with simple temporal structure. This comparison provides a replication test of previously reported findings using similar stimulus contrasts (Démonet *et al.*, 1992; Zatorre *et al.*, 1992; Binder *et al.*, 1997), and allows the location of these regions to be compared with those associated with simple temporal processing (Tones-Noise) in the same subjects.

No areas showed significantly stronger responses to Tones than Words. Areas responding more strongly to Words included the mid-portion of the lateral STG (lateral to the anterior aspect of HG) and adjacent STS bilaterally (Fig. 5A). There were at least two neighboring but distinct peak foci in each hemisphere. These were located, on average, 7–8 mm ventral to the Tones-Noise foci (Table 2). No frontal areas showed significant differences between Words and Tones.

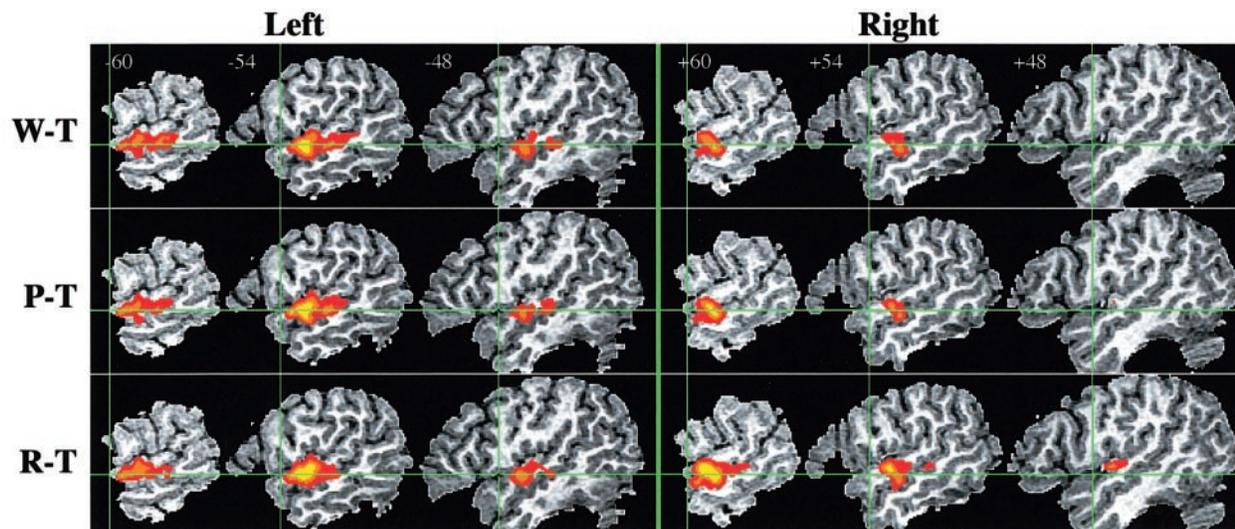
A direct comparison between speech-sensitive regions, identified in the Words-Tones contrast, and simple temporal processing regions, identified in the Tones-Noise contrast, is shown in Figure 5B. A dorsal-to-ventral organization is apparent, with simple temporal processing regions located dorsally on the

STG and speech-sensitive regions located more ventrolateral, within or near the STS. An intermediate zone responded more to Tones than Noise and also more to Words than Tones.

#### **Linguistic Factors: Words versus Nonword Speech**

Contrasts between word and nonword speech sounds were conducted to test whether the regions selectively activated by word stimuli are sensitive to phonetic or lexical-semantic attributes of the stimuli. As discussed earlier, the Pseudoword stimuli differ from the Word stimuli primarily in terms of eliciting fewer explicit lexical-semantic associations. The Reversed stimuli provide a stronger contrast, differing from the Word stimuli in terms of both phonetic intelligibility and lexical-semantic content. Pseudoword and Reversed conditions were each contrasted with Tones in order to compare these results with the Words-Tones data, and both of these nonword conditions were contrasted directly with the Words condition.

The Pseudowords-Tones and Reversed-Tones contrasts yielded results indistinguishable from Words-Tones (Fig. 6). In each case, bilateral regions at the mid-lateral STG and adjacent STS demonstrated stronger responses to speech sounds than Tones. The location of peak foci were strikingly similar (Table 2)



**Figure 6.** Comparison between three speech–Tones contrasts (Words–Tones, Pseudowords–Tones and Reversed–Tones). Areas responding more to speech than to Tones are very similar for all contrasts.

and differed across the three contrasts by an average of only 3.0 mm (SD = 1.8). No areas showed stronger responses to Tones in either contrast.

The direct contrasts Words–Pseudowords and Words–Reversed revealed no areas of significant activation difference in either direction in either hemisphere, nor did a direct contrast between Pseudowords and Reversed conditions. To investigate the possibility that small differences in activation between word and nonword conditions might have been missed by overly conservative thresholding, a second set of SPMs was produced using a much more lenient voxelwise threshold for significance (uncorrected  $P < 0.05$ ) and no cluster size threshold. Of particular interest was whether the set of voxels originally identified in the Words–Tones contrast (Fig. 5A) would show any activation differences between the three speech conditions (Words, Pseudowords, Reversed). Several small left hemisphere foci showed stronger activation to Words than to Pseudowords using this threshold. These were located in the posterior inferior temporal gyrus (BA 20/37, peak coordinates =  $-46, -40, -17$ ), rostral middle frontal gyrus (BA 46/10, peak coordinates =  $-40, 40, 0$ ), angular gyrus (BA 39, peak coordinates =  $-44, -54, 39$ ), and the border between posterior middle and inferior temporal gyri (BA 21/37, peak coordinates =  $-53, -46, -5$ ). None of these clusters overlapped with the clusters identified in the speech–nonspeech contrasts (Figs 5A, 6). No activation differences favoring Words were observed in the Words–Reversed contrast even using this more lenient threshold. Of note, however, several small regions in the *right* anterior STG (peak coordinates =  $-58, -9, 10$ ) and STS (peak coordinates =  $63, -13, 2$ ) showed stronger activation to Reversed than to Words at these thresholds. In the contrast between Pseudowords and Reversed, the only differences observed were several very small areas in the right anterior STG that favored Reversed stimuli. These regions were smaller but otherwise identical to those observed in the Words–Reversed contrast.

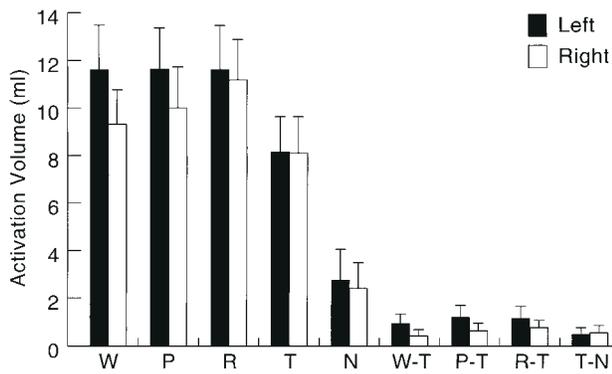
#### **Hemispheric Asymmetry of Superior Temporal Activation**

The volume of activated superior temporal lobe tissue was estimated for each condition contrast and each hemisphere by

multiplying the volume of each voxel by the number of voxels within the superior temporal region that survived a statistical threshold of  $P < 0.0001$  (Fig. 7). Activation volumes in individual subjects ranged from 0 to 27.76 ml. No voxels survived the statistical threshold in any subject in any of the word–nonword contrasts (i.e. contrasts between Words, Pseudowords, and Reversed), so these contrasts were not analyzed further. As expected, the largest activation volumes were obtained in the speech–baseline contrasts (mean for Words, Pseudowords and Reversed averaged across left and right hemispheres = 10.902 ml). Tones–baseline also resulted in extensive activation (mean averaged across hemispheres = 8.134 ml). Smaller activation volumes were obtained for Noise–baseline (mean averaged across hemispheres = 2.589 ml), speech–Tones (mean for Words–Tones, Pseudowords–Tones and Reversed–Tones averaged across hemispheres = 0.860 ml) and Tones–Noise (mean averaged across hemispheres = 0.524 ml). All of these means were significantly greater than zero at  $P < 0.0001$ .

Collapsing across all nine condition contrasts, activation volume was significantly greater in the left hemisphere than in the right hemisphere (mean left–right difference = 0.691 ml;  $t = 4.669, P < 0.0001$ ). However, this left–right difference varied significantly by condition contrast (ANOVA with df 8,243;  $F = 3.290, P = 0.0014$ ). Of the nine condition contrasts, only Words–baseline showed a significant left–right volume difference after Bonferroni correction for the number of comparisons (mean left–right difference = 2.315 ml;  $t = 4.312, P = 0.0002$ ), although Pseudowords–baseline ( $P = 0.027$ ), Words–Tones ( $P = 0.01$ ), Pseudowords–Tones ( $P = 0.038$ ) and Reversed–Tones ( $P = 0.089$ ) all showed trends in the same direction of greater volume on the left side. When the three speech–Tones condition contrasts were collapsed together, activation volume in the left hemisphere was significantly greater than the right ( $t = 3.831, P = 0.0002$ ), although the absolute mean left–right difference was comparatively small (0.493 ml). There was no significant lateralization for Tones (mean left–right difference = 0.057 ml), Noise (mean left–right difference = 0.353 ml) or Tones–Noise (mean left–right difference =  $-0.070$  ml).

Thus, left–right differences in activation volume were comparatively small except for Words–baseline. It is apparent from



**Figure 7.** Estimated activation volumes in each hemisphere for nine condition contrasts. Error bars indicate 95% confidence intervals. The left–right difference is significant ( $P = 0.0002$ ) for Words–baseline (W), and strong trends toward leftward asymmetry are observed in the speech–Tones contrasts.

Figure 7 that the hemispheric asymmetry resulting from the speech–baseline contrasts was greatest for Words, less for Pseudowords and least for Reversed speech. However, this difference is not due to differential left hemisphere activation but rather to differential right hemisphere activation. That is, Reversed–baseline and Words–baseline produced almost identical activation volumes in the left hemisphere (mean = 11.625 and 11.636 ml, respectively), whereas Reversed–baseline tended to activate a larger volume of the right hemisphere (mean = 11.182 ml) than did Words–baseline (mean = 9.320 ml).

## Discussion

BOLD fMRI responses were obtained from superior temporal cortex in the human brain using a wide range of auditory stimuli, including white noise, FM tones and speech. HG and PT, on the dorsal temporal surface, were activated by all stimuli, whereas activation of the lateral surface of the STG and the STS was more variable across stimuli. The spatial extent of activation was greatest for speech sounds, less for FM tones and least for noise. Although primary auditory cortex on the medial half of HG responded equivalently to all stimuli, much of the surrounding dorsal temporal surface and lateral STG showed stronger activation to FM tones than to noise, despite the greater bandwidth of the noise stimuli. More ventrolateral regions, located primarily in the STS, showed stronger activation to speech sounds than to FM tones. Activation of these areas was equivalent whether the speech sounds were words, pseudowords or reversed speech.

Superior temporal activation occurred bilaterally for all stimuli, as in previous auditory studies (Petersen *et al.*, 1988; Wise *et al.*, 1991; Howard *et al.*, 1992; Mazoyer *et al.*, 1993; Binder *et al.*, 1994b; Fiez *et al.*, 1995, 1996; O’Leary *et al.*, 1996; Price *et al.*, 1996b; Warburton *et al.*, 1996; Dhankhar *et al.*, 1997; Grady *et al.*, 1997; Hirano *et al.*, 1997; Jäncke *et al.*, 1998). Right temporal areas of activation were shifted anterolaterally relative to homologous activations in the left hemisphere. This asymmetry is consistent with previous morphometric studies showing a relative anterolateral shift of HG in the right hemisphere (Penhune *et al.*, 1996). Right auditory cortex activation included more of the anterolateral STG and planum polare than did left auditory cortex activation. This arrangement may be compensation for the fact that the horizontal portion of the PT is generally larger in the left hemisphere (Geschwind and Levitsky, 1968; Steinmetz *et al.*, 1991; Witelson and Kigar, 1992; Leonard

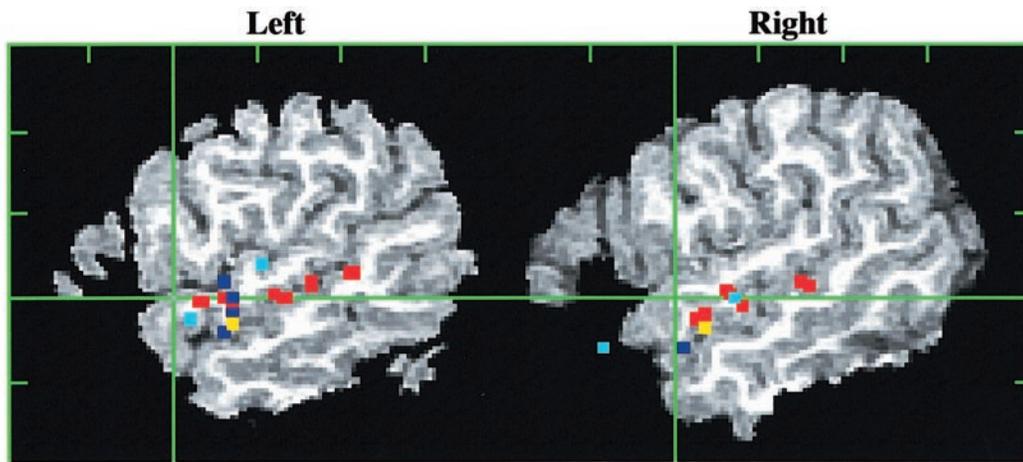
*et al.*, 1993; Foundas *et al.*, 1994). Quantitative measures indicated relative symmetry of activation volume for the superior temporal region as a whole for most stimulus conditions. A notable exception was the Words condition, which produced significantly less activation of the right temporal lobe than the left. However, because Words produced no more left hemisphere activation than the Reversed stimuli, this asymmetry is not likely due to additional (language specific) left hemisphere recruitment by the Words stimuli. Rather, the asymmetry during Words results from the fact that these stimuli were relatively less effective (compared with nonwords) at activating the right temporal lobe (Fig. 7). Using relaxed statistical criteria, regions in the anterior right STG were more strongly activated by Reversed stimuli than by Words or Pseudowords in direct voxel-wise comparisons. One possible explanation for this pattern is that the nonwords (particularly the Reversed stimuli) were relatively novel and may have engaged subjects’ attention more than the Words stimuli (Wood and Cowan, 1995). Several prior studies have demonstrated modulatory effects of attention on auditory cortex activation (O’Leary *et al.*, 1996; Pugh *et al.*, 1996; Grady *et al.*, 1997). If this account is correct, the data imply a greater susceptibility of the right auditory cortex than the left to such attentional modulation.

In the following discussion, we focus on the temporal lobe activation observed when speech sounds are contrasted with simpler nonspeech sounds. At issue are the following: (i) How reliably is this activation observed? (ii) Where is the activation located, and how variable is the location across different studies? (iii) Does the activation reflect auditory, phonetic, lexical or semantic processing? (iv) Why were no differences in activation observed using words and nonwords? Previously published results from studies contrasting speech to nonspeech and words to nonwords are reviewed. Finally, the available data are interpreted within a proposed hierarchical model of auditory word processing in the temporal lobe.

## Speech Activation Relative to Nonspeech

The results confirm previous reports describing areas within the superior temporal region that respond more strongly to speech sounds than to nonspeech sounds. Relative to areas identified in the Tones–Noise contrast, these speech processing areas are significantly more ventral (Fig. 5B). Figure 8 shows speech > nonspeech peak activation foci identified in the present study and in three previous studies (Démonet *et al.*, 1992; Zatorre *et al.*, 1992; Binder *et al.*, 1997). The findings are remarkably consistent, particularly in the left hemisphere, despite differences across studies with regard to the speech stimuli used (words, pseudowords, syllables, reversed speech), type of nonspeech stimuli (amplitude modulated noise, frequency modulated tones), task performed by the subject and imaging methodology. The activation peaks cluster closely near the mid-portion of the STS and lateral STG. The mean of these peaks is at  $x = -55.5$  (SD 2.3),  $y = -20.2$  (SD 10.9),  $z = 0.3$  (SD 4.1) in the left hemisphere and  $x = 57.0$  (SD 4.2),  $y = -15.0$  (SD 13.8),  $z = -2.2$  (SD 6.1) in the right hemisphere. According to the atlas of Talairach and Tournoux, these mean coordinates are in a portion of the STS located ventral to the lateral aspect of the transverse temporal sulcus (Talairach and Tournoux, 1988).

Activation of the right temporal lobe was less consistent across the four studies. In two studies there was a strong left–right asymmetry of the areas identified (Démonet *et al.*, 1992; Binder *et al.*, 1997). This discrepancy may be due to differences in task requirements. No material-specific (linguistic or pitch)



**Figure 8.** Summary of speech > nonspeech superior temporal activation foci in four functional imaging studies. For presentation purposes, peak foci are collapsed onto representative brain slices at  $\pm 55$  mm from midline. Red (present study), light blue (Zatorre *et al.*, 1992), dark blue (Démonet *et al.*, 1992), yellow (Binder *et al.*, 1997).

discrimination was required in the present study, which used a nonspecific detection response, or in the experiment by Zatorre *et al.* (Zatorre *et al.*, 1992), in which subjects listened passively and made no motor response. Superior temporal responses in these two studies were relatively symmetric. In contrast, Démonet *et al.* and Binder *et al.* (Démonet *et al.*, 1992; Binder *et al.*, 1997) used material-specific discrimination tasks for both the speech and nonspeech stimuli. These additional processing demands appear to have accentuated the response asymmetry in this region, either through additional activation of the left temporal lobe during the speech discrimination task or through additional activation of the right temporal lobe during the nonspeech discrimination task (or both).

#### **Effects of Linguistic and Acoustic Factors**

Despite what appear to be significant linguistic differences between the Word, Pseudoword and Reversed stimuli used in the present study, we detected no significant differences in brain activation state when subjects listened to these stimuli and performed a nonspecific observing response. The same voxels identified in the Words–Tones contrast were also identified by contrasting Pseudoword or Reversed stimuli with Tones, and there were no differences in activation level of these voxels across the Words, Pseudowords and Reversed conditions even using very lenient significance thresholds. Processing in these regions thus appears to be unrelated to activation of semantic associations of the stimuli, given that the nonwords (particularly the Reversed stimuli) are unlikely to have resulted in retrieval of specific semantic information. What remains at issue is whether the data suggest lexical, phonetic or auditory processes as the explanation for activation in the speech–nonspeech contrasts.

These findings do not exclude the possibility that some or all of the speech > nonspeech activation could be due to processing of lexical or phonetic representations. Some theorists have proposed computational models of word recognition that accomplish pseudoword recognition through activation of multiple lexical representations (Glushko, 1979; McClelland and Rumelhart, 1981), so it is possible that lexical information was activated during all three conditions (Wise *et al.*, 1991; Price *et al.*, 1996a). Similarly, the pilot data presented earlier demonstrate that subjects are capable of extracting phonetic information from pseudowords and reversed speech sounds,

suggesting that phonetic categorization processes were probably activated to some degree during all three conditions. On the other hand, previous investigators have generally assumed that reversed speech sounds do not elicit lexical processing (Howard *et al.*, 1992; Perani *et al.*, 1996; Price *et al.*, 1996b; Dehaene *et al.*, 1997; Hirano *et al.*, 1997). Although transcription reliability was better than chance for all stimuli in the pilot study, subjects had greater difficulty transcribing reversed speech than pseudoword speech, suggesting that reversed speech sounds are, in general, less easily classified within familiar phonetic categories. Phonetic processing of the reversed sounds may have been further suppressed during the imaging study by the competing scanner noise. Despite these important differences in the way subjects interpret reversed speech sounds, there were no corresponding differences in activation of STG and STS regions identified in the speech–nonspeech contrasts. The data thus provide no definite evidence that the speech > nonspeech activation difference in these regions is related to either lexical or phonetic processing.

A simpler explanation is that the activation observed during these speech–nonspeech contrasts is due largely to physical differences between the speech and nonspeech stimuli (Table 1, Fig. 1). Compared with the Tones stimuli, which contain only simple, stepwise frequency modulations at predictable, infrequent intervals, speech sounds are characterized by nearly continuous frequency and amplitude modulations with complex and highly variable temporal structure. Speech sounds typically include both noise and periodic elements, and these components may occur simultaneously or in rapid succession. Unlike the Tones, the periodic components in speech sounds are harmonically complex, with constantly changing, inhomogeneous distributions of energy across the harmonic series caused by vocal tract resonances (speech formants). Acoustic features distinguishing speech phonemes can include any of these phenomena alone or in combination with other elements. For example, information about the voicing of stop consonants (i.e. the distinction between /b/ and /p/ or between /d/ and /t/) is represented in the time latency between consonant–initial noise burst and onset of periodicity, the fundamental frequency of the periodicity, the onset frequency of the first formant resulting from periodicity, the relative amplitude of the aspiration noise preceding periodicity and the duration of the vowel (if any)

preceding the stop (Lisker and Abramson, 1964; Summerfield and Haggard, 1977; Repp, 1979, 1982; Kingston and Diehl, 1994). All of this acoustic information is typically presented within a few tens of milliseconds, necessitating continuous, rapid analysis not only of individual acoustic events but also of sequential or simultaneous combinations of events. While differing from the Tones condition, the three speech conditions were matched to each other in terms of this acoustic complexity. Thus the data, which demonstrate brain regions that respond more strongly to speech than to tones, but respond identically to word and nonword speech, are explainable on the basis of differences in acoustic content alone. This interpretation is also consistent with the fact that the speech > nonspeech activations were bilateral. If these activation differences are due primarily to differences in auditory sensory processing demands rather than to linguistic processing, they would be expected to occur in auditory cortex bilaterally.

### **Word versus Nonword Processing**

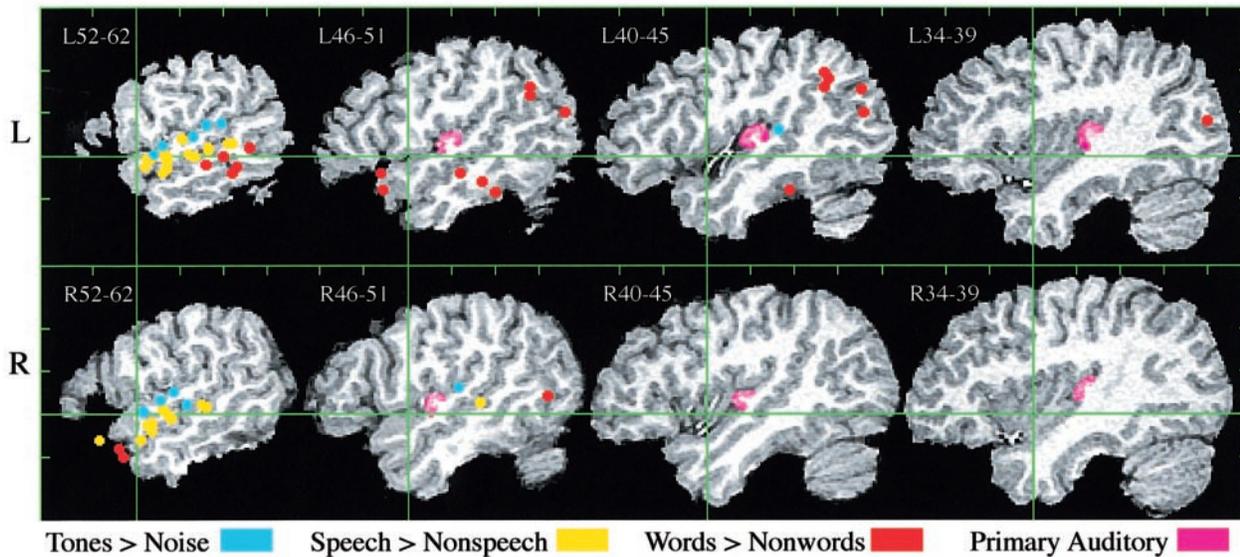
Even if this interpretation of the speech > nonspeech activation is correct, there remains the question of why no differences were observed between the Words, Pseudowords and Reversed conditions in other brain areas. For example, left hemisphere ventral temporal regions, angular gyrus and prefrontal cortex have all been implicated in semantic processing (Petersen *et al.*, 1989; Démonet *et al.*, 1992; Kapur *et al.*, 1994; Demb *et al.*, 1995; Vandenberghe *et al.*, 1996; Binder *et al.*, 1997, 1999; Price *et al.*, 1997b). Given the differences in semantic content across conditions outlined in Table 1, it is surprising that no significant activation differences were observed in these or other regions in the word–nonword contrasts. This lack of difference in brain activation by word and nonword speech has been reported by other investigators as well (Wise *et al.*, 1991; Hirano *et al.*, 1997). One possible explanation is that subjects simply did not process the semantic associations of the word stimuli because there was no explicit requirement to do so. Although straightforward, this account seems inadequate given the extensive body of evidence demonstrating ‘implicit’ semantic processing when subjects are presented with meaningful linguistic stimuli (Carr *et al.*, 1982; Marcel, 1983; Van Orden, 1987; Macleod, 1991; Price *et al.*, 1996a). A familiar example is the Stroop effect, in which semantic processing of printed words (i.e. retrieval of word meaning) occurs even when subjects are instructed to attend to the color of the print and even when this semantic processing interferes with task performance (Glaser and Glaser, 1989; Macleod, 1991). Price *et al.* demonstrated that such implicit processing alters brain activity sufficiently to be detected by functional imaging (Price *et al.*, 1996a). They showed that visual words and pseudowords produce activation in widespread, left-lateralized cortical areas relative to activation produced by nonsense characters and consonant strings, even when the explicit task (visual feature detection) is identical across all conditions.

Given the possibility of ‘implicit’ lexical–semantic retrieval, any explanation for the observed lack of word–nonword differences necessarily becomes more complex. One variable that may be critical is presentation rate. In the present study, words were presented at a relatively rapid rate (1.5 per s) with no inter-stimulus interval. Although this is not fast compared with a typical spoken sentence, the words in this case were all content nouns, which typically occur at much slower rates in natural speech. It may be that activation of lexical–semantic word associations requires considerably more time than activa-

tion of phonological representations (Collins and Loftus, 1975; Carr *et al.*, 1982; Van Orden *et al.*, 1988). Moreover, the presentation of a second word immediately after the first may act as a ‘semantic mask’ that disrupts retrieval processes related to the first word. In this way, rapid presentation of successive content words might result in relatively little activation of semantic information.

Another possibility is that processing of stored knowledge (including lexical and semantic information) is something of a ‘default’ state during normal consciousness (James, 1890; Pope and Singer, 1976; Binder *et al.*, 1999). When not otherwise occupied with processing sensory stimuli, we appear to continually retrieve and manipulate stored knowledge for the purposes of problem solving and planning (Antrobus *et al.*, 1966; Pope and Singer, 1976; Teasdale *et al.*, 1993; Binder *et al.*, 1999). Many of the same left hemisphere regions previously implicated in semantic processing of words (angular gyrus, prefrontal cortex, ventral temporal cortex) were shown recently to be active during the awake resting state relative to a non-linguistic discrimination task (Binder *et al.*, 1999). Relative to this ongoing ‘conceptual’ activity, activation evoked by processing specific semantic associations of the word stimuli may have been negligible. Thus, a possible reason that no activation difference was observed in semantic processing regions (e.g. angular gyrus, prefrontal cortex, ventral temporal cortex) between word and nonword listening conditions is that these regions were active during all conditions, including rest. This account also explains why there was little or no activation of these areas during any of the listening conditions relative to rest, a typical finding in auditory studies that use a resting baseline (Petersen *et al.*, 1988; Wise *et al.*, 1991; Mazoyer *et al.*, 1993; Binder *et al.*, 1994b; O’Leary *et al.*, 1996; Price *et al.*, 1996b; Grady *et al.*, 1997; Hirano *et al.*, 1997; Jäncke *et al.*, 1998).

Although several other studies also showed no differences in activation between word and nonword speech sounds (Wise *et al.*, 1991; Hirano *et al.*, 1997), some investigators have observed differences (Démonet *et al.*, 1992, 1994; Howard *et al.*, 1992; Perani *et al.*, 1996; Price *et al.*, 1996b; Binder *et al.*, 1999). This apparent discrepancy could relate to differences in task demands, stimulus characteristics, statistical power or random effects. In nearly all of these studies the subjects performed contrasting attentional tasks during the word and nonword conditions, which were intended to differentially emphasize or de-emphasize retrieval of lexical or semantic information (Démonet *et al.*, 1992, 1994; Howard *et al.*, 1992; Perani *et al.*, 1996; Price *et al.*, 1996b; Binder *et al.*, 1999). Thus, these attentional shifts probably contributed to activation differences between word and nonword conditions, unlike the neutral observing task used in the present study. In one study, differences between word and nonword activation were observed even during passive listening conditions [Experiment 3 in (Price *et al.*, 1996b)]. One possibility is that the relatively small number of subjects in this study (four) increased the likelihood that weak activation foci could exceed statistical thresholds by coincidence. Another possibility is that the longer activation periods used in this PET study (2–3 min, in contrast to the 12 s activation periods used in the present study) and the slower rate of word presentation (40 words per min, in contrast to the 90 words per min rate used in the present study) allowed more time for build up of activation in semantic networks during the word condition. In any case, the discrepancies in results are more likely to be quantitative than qualitative. At lower statistical



**Figure 9.** Summary of temporal and temporoparietal responses to speech and nonspeech auditory stimuli in nine functional imaging studies. HG, the site of primary auditory cortex, is highlighted in pink. Light blue foci on the dorsal STG indicate peaks identified in the Tones–Noise contrast of the present study. Yellow foci clustering in the STS bilaterally indicate peaks identified in various speech–nonspeech contrasts (see Fig. 8). Red foci indicate peaks identified in studies contrasting auditory words and nonwords (Démonet *et al.*, 1992, 1994; Perani *et al.*, 1996; Price *et al.*, 1996b; Binder *et al.*, 1999), and peaks from the Words–Pseudowords contrast of the present study. The results suggest a hierarchical processing stream from dorsal temporal cortex ventrally to STS, then ventrally to MTG and ITG, posteriorly to angular gyrus, and anteriorly to temporal pole. Only the word–nonword activations are strongly left-lateralized.

thresholds, differences between Words and Pseudowords conditions were detected in several left hemisphere areas in the present study, and the locations of these foci agree closely with those reported in other studies using word–nonword contrasts.

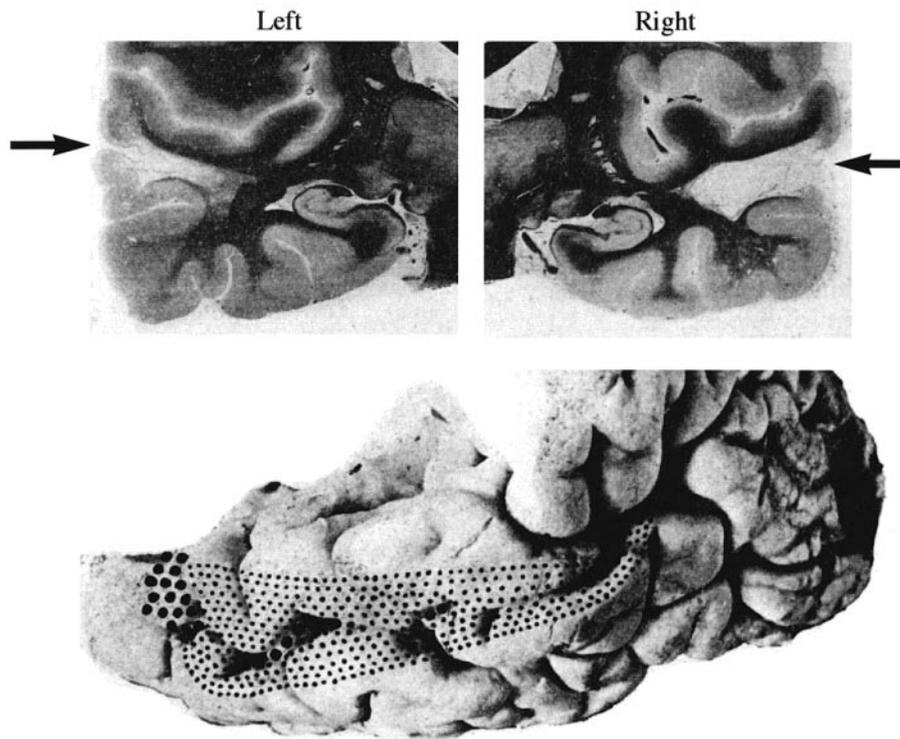
A final point is that the activation foci observed in all of these word–nonword contrasts are distinctly separate from those observed in the speech–nonspeech contrasts. As shown in Figure 9, the word > nonword foci lie in distributed temporal and temporoparietal areas well outside the superior temporal auditory cortex, including posterior middle and inferior temporal gyri, angular gyrus and temporal poles. Thus, these observations are consistent with the interpretation given earlier regarding speech–nonspeech activation differences. Differences in activation between word and nonword speech sounds, when observed, do not occur in areas that are differentially activated by speech and nonspeech. These latter differences are thus more likely explainable on the basis of acoustic differences between the speech and nonspeech sounds.

#### **A Hierarchical Model of Auditory Word Processing**

Together, these data provide evidence for a preliminary hierarchical model of auditory word processing in the temporal lobe. Cortex on the dorsal temporal plane bilaterally responded to all stimuli, including unstructured noise. Posterior and lateral areas within this region, particularly the PT and lateral STG, showed stronger responses to sounds with simple temporal structure (Tones) than to noise (light blue foci in Fig. 9). The arrangement of these ‘structure-sensitive’ areas at locations posterior and lateral to HG is comparable to the arrangement of ‘belt’ auditory regions around the AI ‘core’ in primate and cat superior temporal lobe (Pandya and Sanides, 1973; Galaburda and Pandya, 1983; Morel *et al.*, 1993; Rauschecker *et al.*, 1995; Rauschecker *et al.*, 1997; Hackett *et al.*, 1998; Kaas and Hackett, 1998). Neurons in the auditory belt of such animals are characterized by sensitivity to temporal and spectral structure, including direction and rate of frequency modulations (Tian and

Rauschecker, 1994, 1998; Rauschecker *et al.*, 1995). Consistent with these findings, the present data suggest that dorsolateral temporal areas surrounding HG in humans are likely to be involved in processing elemental, temporally encoded auditory information. Such information is the raw material on which consonant phoneme discriminations and many other ecologically salient auditory discriminations are based, supporting the idea that there may be specific cortical areas specialized for encoding and integrating such information. In the present study, areas responding more to sounds with simple temporal structure than to noise were observed bilaterally, and there was no inter-hemispheric difference in the size of these areas.

Encoding of individual temporal features, however, is insufficient for speech perception; the perceptual correlate of a single feature may vary considerably in different acoustic contexts (Lieberman *et al.*, 1967). It is thus clear that phonetic contrasts are almost never based on one feature, but rather on synchronous and sequential combinations of features that collectively contribute to phoneme identity (Lieberman *et al.*, 1967; Dorfman *et al.*, 1977; Massaro and Oden, 1980; Repp, 1982; McClelland and Elman, 1986; Lieberman and Blumstein, 1988; Klatt, 1989). These considerations suggest the possibility of a second processing level that gives rise to neural representations of temporal and spectral feature combinations that can be assigned to unique phoneme (or other sound) categories (Pisoni, 1973; Massaro and Oden, 1980; Samuel, 1982; McClelland and Elman, 1986; Kuhl, 1991). Areas identified in the speech–nonspeech contrasts (yellow foci in Fig. 9) could conceivably represent such a processing stage. One salient factor in all of these contrasts is that the nonspeech stimuli, although possessing simple frequency and/or amplitude modulations, lack complex feature combinations. In contrast, the word, pseudo-word and reversed speech stimuli used in the present study were matched on this variable, and all resulted in equivalent activation of this region. This region, which is located just ventral to the presumed belt areas, may correspond roughly to part of the



**Figure 10.** Anatomical findings in a patient with pure word deafness (Barrett, 1910). At the top are coronal sections through the planum temporale and Heschl's gyrus, showing infarction within the depths of the STS (arrows). PT and HG are spared bilaterally. At bottom is Barrett's drawing of the involved regions in the left hemisphere. Large dots indicate regions of cortical involvement, which is confined to anterolateral STG and depths of STS. Small dots indicate subcortical involvement.

auditory 'parabelt' in nonhuman primates, which is also ventral to the belt region and may show particular sensitivity to complex sounds such as vocalizations (Symmes, 1971; Rauschecker *et al.*, 1995). Areas responding more to speech than to nonspeech sounds were observed bilaterally but were slightly more extensive in the left hemisphere (Figs 5, 7). Thus, these areas display a greater degree of hemispheric specialization than was observed for the dorsolateral temporal plane.

Neuropsychological studies have long linked speech sound processing to the superior part of the temporal lobe, although no clear hierarchical organization has emerged from these data. Of greatest relevance is the 'pure word deafness' syndrome, in which patients are able grossly to detect sounds [although there may be interference with threshold testing by marked habituation and a variety of other subtle auditory deficits (Auerbach *et al.*, 1982; Tanaka *et al.*, 1987; Mendez and Geehan, 1988)], but show severely impaired phoneme discrimination. Because phoneme perception depends on early as well as later stages of auditory processing, such a deficit pattern could arise from lesions at almost any level in auditory cortex, provided that the lesions are bilateral. The available case material supports this formulation, in that nearly all cases with autopsy confirmation have had bilateral lesions, and the lesions have involved a variety of regions on the STG and STS (Barrett, 1910; Henschen, 1918–1919; Wohlfart *et al.*, 1952; Lhermitte *et al.*, 1972; Kanshepolsky *et al.*, 1973; Buchman *et al.*, 1986). In two well-documented cases, the left hemisphere lesion was centered on the STS and anterior STG, demonstrating that relatively isolated speech perception impairment can occur from a lesion sparing HG and PT (Barrett, 1910; Henschen, 1918–1919) (Fig. 10).

Finally, a third level of processing during auditory word recognition can be envisioned, involving activation of information in long-term storage that pertains to the stimulus, including such information as lexical status, syntactic function and meaning. This information is independent of the physical characteristics of the stimulus (e.g. there is nothing about the speech waveform for 'desk' that contains information about its meaning or lexical status). These areas are thus distinct from earlier sensory stages concerned with encoding and integrating physical features of sounds, although there is ample evidence that stored phonological, lexical, semantic and syntactic information exerts top-down influence on these earlier stages (Bagley, 1900; Miller *et al.*, 1951; Warren and Obusek, 1971; Pisoni, 1973; Marslen-Wilson and Welsh, 1978; Ganong, 1980; Massaro and Oden, 1980; Samuel, 1982; McClelland and Elman, 1986; Kuhl, 1991). The contrasts between word and nonword speech (red foci in Fig. 9) provide data regarding the location of these areas. As noted above, these areas surround the superior temporal auditory region posteriorly (angular gyrus), ventrally (middle and inferior temporal gyri) and anteriorly (temporal poles) but are spatially separate from the auditory region. With the exception of the temporal poles, where activation has been observed bilaterally, the temporal and parietal areas revealed by word–nonword contrasts are strikingly lateralized to the left hemisphere (Démonet *et al.*, 1992, 1994; Perani *et al.*, 1996; Price *et al.*, 1996b; Binder *et al.*, 1999). Consistent with the notion that these areas are involved in supramodal processing, several studies contrasting words to nonwords or semantic to nonsemantic tasks in the visual modality also demonstrated activation of these areas (Price *et al.*, 1994, 1996a, 1997a,b; Vandenberghe *et al.*, 1996; Herbster *et al.*, 1997).

## Conclusions

In summary, the human superior temporal region, as in nonhuman primates and other animals, consists primarily of auditory sensory cortex. As such, much of this region is activated by a variety of speech and nonspeech auditory stimuli, and this activation is bilateral and relatively symmetric. There is clear converging evidence for a specialized region in the mid-portion of the STS that responds preferentially to speech sounds when compared with simple nonspeech sounds. This response selectivity is unlikely to be due to activation of lexical-semantic associations of the speech sounds and may reflect processing of complex acoustic feature combinations. Although bilateral, these regions are somewhat more extensive in the left hemisphere. Areas involved in processing lexical-semantic information associated with speech sounds are spatially distinct from these auditory processors and show considerable dominant hemisphere lateralization. These observations provide a preliminary framework for functional subdivision of the dorsal and lateral temporal lobe in humans, to which future studies can be expected to add greater detail concerning intermediate stages in auditory processing, visual word processing and access to meaning.

## Notes

This research was supported by National Institute of Neurological Diseases and Stroke grants RO1 NS33576 and RO1 NS35929, and National Institute of Mental Health grant PO1 MH51358.

Address correspondence to Jeffrey Binder, MD, Department of Neurology, 9200 W. Wisconsin Ave, Milwaukee, WI 53226, USA. Email: jbinder@mcw.edu.

## References

Antrobus JS, Singer JL, Greenberg S (1966) Studies in the stream of consciousness: experimental enhancement and suppression of spontaneous cognitive processes. *Percept Motor Skills* 23:399-417.

Auerbach SH, Allard T, Naeser M, Alexander MP, Albert ML (1982) Pure word deafness. Analysis of a case with bilateral lesions and a defect at the prephonemic level. *Brain* 105:271-300.

Bagley WC (1900) The apperception of the spoken sentence: a study in the psychology of language. *Am J Psychol* 12:80-130.

Barrett AM (1910) A case of pure word-deafness with autopsy. *J Nerv Ment Dis* 37:73-92.

Binder JR, Rao SM, Hammeke TA, Frost JA, Bandettini PA, Hyde JS (1994a) Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Cogn Brain Res* 2:31-38.

Binder JR, Rao SM, Hammeke TA, Yetkin YZ, Jesmanowicz A, Bandettini PA, Wong EC, Estkowski LD, Goldstein MD, Haughton VM, Hyde JS (1994b) Functional magnetic resonance imaging of human auditory cortex. *Ann Neurol* 35:662-672.

Binder JR, Rao SM, Hammeke TA, Frost JA, Bandettini PA, Jesmanowicz A, Hyde JS (1995) Lateralized human brain language systems demonstrated by task subtraction functional magnetic resonance imaging. *Arch Neurol* 52:593-601.

Binder JR, Frost JA, Hammeke TA, Rao SM, Cox RW (1996) Function of the left planum temporale in auditory and linguistic processing. *Brain* 119:1239-1247.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional MRI. *J Neurosci* 17:353-362.

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Rao SM, Cox RW (1999) Conceptual processing during the conscious resting state: a functional MRI study. *J Cogn Neurosci* 11:80-93.

Bogen JE, Bogen GM (1976) Wernicke's region - where is it? *Ann NY Acad Sci* 290:834-843.

Braak H (1978) On magnopyramidal temporal fields in the human brain - probable morphological counterparts of Wernicke's sensory speech region. *Anat Embryol* 152:141-169.

Buchman AS, Garron DC, Trost-Cardamone JE, Wichter MD, Schwartz D

(1986) Word deafness: one hundred years later. *J Neurol Neurosurg Psychiatr* 49:489-499.

Carr TH, McCauley C, Sperber RD, Parmalee CM (1982) Words, pictures, and priming: on semantic activation, conscious identification, and the automaticity of information processing. *J Exp Psychol (Hum Percept Perform)* 8:757-777.

Celesia GG (1976) Organization of auditory cortical areas in man. *Brain* 99:403-414.

Collins AM, Loftus E (1975) A spreading-activation theory of semantic processing. *Psychol Rev* 82:407-428.

Cox RW (1996a) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162-173.

Cox RW (1996b) Algorithms for image registration, motion detection, and motion correction. In: *fMRI2DAY: a two day workshop on functional MRI* (Savoy R, ed.), pp. 25-43. Boston, MA: MGH-NMR Center.

Dehaene S, Dupoux E, Mehler J, Cohen L, Paulesu E, Perani D, van de Moortele P-F, Lehericy S, Le Bihan D (1997) Anatomical variability in the cortical representation of first and second language. *NeuroReport* 8:3809-3815.

Demb JB, Desmond JE, Wagner AD, Vaidya CJ, Glover GH, Gabrieli JDE (1995) Semantic encoding and retrieval in the left inferior prefrontal cortex: a functional MRI study of task difficulty and process specificity. *J Neurosci* 15:5870-5878.

Démonet J-F, Chollet F, Ramsay S, Cardebat D, Nespoulous J-L, Wise R, Rascol A, Frackowiak R (1992) The anatomy of phonological and semantic processing in normal subjects. *Brain* 115:1753-1768.

Démonet J-F, Price C, Wise R, Frackowiak RSJ (1994) Differential activation of right and left posterior sylvian regions by semantic and phonological tasks: a positron emission tomography study in normal human subjects. *Neurosci Lett* 182:25-28.

Dhankhar A, Wexler BE, Fullbright RK, Halwes T, Blamire AM, Shulman RG (1997) Functional magnetic resonance imaging assessment of the human brain auditory cortex response to increasing word presentation rates. *J Neurophysiol* 77:476-483.

Dorfman MF, Studdert-Kennedy M, Raphael LJ (1977) Stop-consonant recognition: release burst and formant transitions as functionally equivalent, context-dependent cues. *Percept Psychophys* 22:109-122.

Fiez JA, Raichle ME, Miezin FM, Petersen SE, Tallal P, Katz WF (1995) PET studies of auditory and phonological processing: effects of stimulus characteristics and task demands. *J Cogn Neurosci* 7:357-375.

Fiez JA, Raichle ME, Balota DA, Tallal P, Petersen SE (1996) PET activation of posterior temporal regions during auditory word presentation and verb generation. *Cereb Cortex* 6:1-10.

Fisher RA, Cornish EA (1960) The percentile points of distributions have known cumulants. *Technometrics* 2:209-226.

Flechsig P (1908) Bemerkungen über die Hörsphäre des menschlichen Gehirns. *Neurol Zentralbl* 27:2-7.

Foundas AL, Leonard CM, Gilmore R, Fennell E, Heilman KM (1994) Planum temporale asymmetry and language dominance. *Neuropsychologia* 32:1225-1231.

Galaburda AM, Pandya DN (1983) The intrinsic architectonic and connective organization of the superior temporal region of the rhesus monkey. *J Comp Neurol* 221:169-184.

Galaburda A, Sanides F (1980) Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol* 190:597-610.

Galaburda AM, LeMay M, Kemper T, Geschwind N (1978) Right-left asymmetries in the brain. Structural differences between the hemispheres may underlie cerebral dominance. *Science* 199:852-856.

Ganong WF (1980) Phonetic categorization in auditory word perception. *J Exp Psychol (Hum Percept Perform)* 6:110-115.

Geschwind N (1971) Aphasia. *N Engl J Med* 284:654-656.

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

Glaser WR, Glaser MO (1989) Context effects in Stroop-like word and picture processing. *J Exp Psychol (Gen)* 118:13-42.

Glushko RJ (1979) The organization and activation of orthographic knowledge in reading aloud. *J Exp Psychol (Hum Percept Perform)* 5:674-691.

Grady CL, Van Meter JW, Maisog JM, Pietrini P, Krasuski J, Rauschecker JP (1997) Attention-related modulation of activity in primary and secondary auditory cortex. *NeuroReport* 8:2511-2516.

Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394:475-495.

- Henschen SE (1918-1919) On the hearing sphere. *Acta Otolaryngol* 1: 423-486.
- Herbster AN, Mintun MA, Nebes RD, Becker JT (1997) Regional cerebral blood flow during word and nonword reading. *Human Brain Map* 5: 84-92.
- Hirano S, Naito Y, Okazawa H, Kojima H, Honjo I, Ishizu K, Yenokura Y, Nagahama Y, Fukuyama H, Konishi J (1997) Cortical activation by monaural speech sound stimulation demonstrated by positron emission tomography. *Exp Brain Res* 113:75-80.
- Howard D, Patterson K, Wise R, Brown WD, Friston K, Weiller C, Frackowiak R (1992) The cortical localization of the lexicons. *Brain* 115:1769-1782.
- Howard MA, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA (1996) A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* 724:260-264.
- James W (1890) *Principles of psychology*, vol. 1. New York: Dover Publications.
- Jäncke L, Shah NJ, Posse S, Grosse-Ryken M, Müller-Gärtner H-W (1998) Intensity coding of auditory stimuli: an fMRI study. *Neuropsychologia* 36:875-883.
- Kaas JH, Hackett TA (1998) Subdivisions of auditory cortex and levels of processing in primates. *Audiol Neurootol* 3:73-85.
- Kanshepolsky J, Kelley JJ, Waggener JD (1973) A cortical auditory disorder: clinical, audiologic and pathologic aspects. *Neurology* 23: 699-705.
- Kapur S, Rose R, Liddle PF, Zipursky RB, Brown GM, Stuss D, Houle S, Tulving E (1994) The role of the left prefrontal cortex in verbal processing: semantic processing or willed action? *NeuroReport* 5: 2193-2196.
- Kingston J, Diehl R (1994) Phonetic knowledge. *Language* 70:419-454.
- Klatt DH (1989) Review of selected models of speech perception. In: *Lexical representation and process* (Marslen-Wilson WD, ed.), pp. 169-226. Cambridge, MA: MIT Press.
- Kuhl PK (1991) Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Percept Psychophys* 50:93-107.
- Leonard CM, Voeller KKS, Lombardino IJ, Morris MK, Hynd GW, Alexander AW, Anderson HG, Garofalakis M, Honeyman JC, Mao J, Agee OF, Staab EV (1993) Anomalous cerebral structure in dyslexia revealed with magnetic resonance imaging. *Arch Neurol* 50:461-469.
- Lhermitte F, Chain F, Escourolle R, et al. (1972) Etude des troubles perceptifs auditifs dans les lésions temporales bilatérales. *Rev Neurol* 24:327-351.
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M (1967) Perception of the speech code. *Psychol Rev* 74:431-461.
- Lieberman P, Blumstein SE (1988) *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.
- Liegeois-Chauvel C, Musolino A, Chauvel P (1991) Localization of the primary auditory area in man. *Brain* 114:139-153.
- Lisker L, Abramson AS (1964) A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20:384-422.
- Macleod CM (1991) Half a century of research on the Stroop effect: an integrative review. *Psychol Bull* 109:163-203.
- Marcel AJ (1983) Conscious and unconscious perception: experiments on visual masking and word recognition. *Cogn Psychol* 15:197-237.
- Marslen-Wilson WD, Welsh A (1978) Processing interactions and lexical access during word recognition in continuous speech. *Cogn Psychol* 10:29-63.
- Massaro DW, Oden GC (1980) Evaluation and integration of acoustic features in speech perception. *J Acoust Soc Am* 67:996-1013.
- Mazoyer BM, Tzourio N, Frak V, Syrota A, Murayama N, Levrier O, Salamon G, Dehaene S, Cohen L, Mehler J (1993) The cortical representation of speech. *J Cogn Neurosci* 5:467-479.
- McClelland JL, Elman JL (1986) Interactive processes in speech perception: the TRACE model. In: *Parallel distributed processing*. Vol. 2. Psychological and biological models (McClelland JL, Rumelhart DE, eds), pp. 58-121. Cambridge, MA: MIT Press.
- McClelland JL, Rumelhart DE (1981) An interactive activation model of context effects in letter perception. Part 1. An account of basic findings. *Psychol Rev* 88:375-407.
- Mendez MF, Geehan GR (1988) Cortical auditory disorders: clinical and psychoacoustic features. *J Neurol Neurosurg Psychiatr* 51:1-9.
- Mesulam MM (1990) Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann Neurol* 28:597-613.
- Miller G, Heise G, Lichten W (1951) The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol* 41: 329-335.
- Morel A, Garraghty PE, Kaas JH (1993) Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol* 335:437-459.
- O'Leary DS, Andreasen NC, Hurtig RR, Hichwa RD, Watkins GL, Boles-Ponto LL, Rogers M, Kirchner PT (1996) A positron emission tomography study of binaurally and dichotically presented stimuli: effects of level of language and directed attention. *Brain Lang* 53: 20-39.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97-113.
- Pandya DN, Sanides F (1973) Architectonic parcellation of the temporal operculum in rhesus monkey and its projection pattern. *Z Anat Entwicklungs* 139:123-161.
- Penhune VR, Zatorre RJ, MacDonald JD, Evans AC (1996) Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb Cortex* 6:661-672.
- Perani D, Dehaene S, Grassi F, Cohen L, Cappa SF, Dupoux E, Fazio F, Mehler J (1996) Brain processing of native and foreign languages. *NeuroReport* 7:2439-2444.
- Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1988) Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331:585-589.
- Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME (1989) Positron emission tomographic studies of the processing of single-words. *J Cogn Neurosci* 1:153-170.
- Pisoni DB (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept Psychophys* 13: 253-260.
- Pope KS, Singer JL (1976) Regulation of the stream of consciousness: toward a theory of ongoing thought. In: *Consciousness and self-regulation* (Schwartz GE, Shapiro D, eds), pp. 101-135. New York: Plenum Press.
- Price CJ, Wise RJS, Watson JDG, Patterson K, Howard D, Frackowiak RSJ (1994) Brain activity during reading. The effects of exposure duration and task. *Brain* 117:1255-1269.
- Price CJ, Wise RSJ, Frackowiak RSJ (1996a) Demonstrating the implicit processing of visually presented words and pseudowords. *Cereb Cortex* 6:62-70.
- Price CJ, Wise RJS, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RSJ, Friston KJ (1996b) Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain* 119:919-931.
- Price CJ, Moore CJ, Friston KJ (1997a) Subtractions, conjunctions, and interactions in experimental design of activation studies. *Hum Brain Map* 5:264-272.
- Price CJ, Moore CJ, Humphreys GW, Wise RJS (1997b) Segregating semantic from phonological processes during reading. *J Cogn Neurosci* 9:727-733.
- Pugh KR, Shaywitz BA, Shaywitz SE, Fulbright RK, Byrd D, Skudlarski P, Shankweiler DP, Katz L, Constable RT, Fletcher J, Lacadie C, Marchione K, Gore JC (1996) Auditory selective attention: an fMRI investigation. *Neuroimage* 4:159-173.
- Rademacher J, Caviness VS, Steinmetz H, Galaburda AM (1993) Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping and neurobiology. *Cereb Cortex* 3: 313-329.
- Rauschecker JP (1998) Cortical processing of complex sounds. *Curr Opin Neurobiol* 8:516-521.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111-114.
- Rauschecker JP, Tian B, Pons T, Mishkin M (1997) Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol* 382:89-103.
- Repp BH (1979) Relative amplitude of aspiration noise as a cue for syllable-initial stop consonants. *Lang Speech* 22:947-950.
- Repp BH (1982) Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychol Bull* 92:81-110.
- Samuel AG (1982) Phonetic prototypes. *Percept Psychophys* 31:307-314.
- Steinmetz H, Volkman J, Jäncke L, Freund HJ (1991) Anatomical

- left-right asymmetry of language-related temporal cortex is different in left- and right-handers. *Ann Neurol* 29:315-319.
- Summerfield Q, Haggard M (1977) On the dissociation of spatial and temporal cues to the voicing distinction in initial stop consonants. *J Acoust Soc Am* 62:435-448.
- Symmes D (1971) On the use of natural stimuli in neurophysiological studies of audition. *Hearing Res* 4:203-214.
- Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical Publishers.
- Tanaka Y, Yamadori A, Mori E (1987) Pure word deafness following bilateral lesions: a psychophysical analysis. *Brain* 110:381-403.
- Teasdale JD, Proctor L, Lloyd CA, Baddeley AD (1993) Working memory and stimulus-independent thought: effects of memory load and presentation rate. *Eur J Cogn Psychol* 5:417-433.
- Tian B, Rauschecker JP (1994) Processing of frequency-modulated sounds in the cat's anterior auditory field. *J Neurophysiol* 71:1959-1975.
- Tian B, Rauschecker JP (1998) Processing of frequency-modulated sounds in the cat's posterior auditory field. *J Neurophysiol* 79:2629-2642.
- Toga AW, Jones AS, Rothfield JM, Woods RP, Payne BA, Huang C, Mazziotta JC, Cai R (1993) Anatomic variability as measured with a 3D reconstructed Talairach atlas. In: Quantification of brain function (Uemura K, Lassen NA, Jones T, Kanno I, eds), pp. 449-456. New York: Elsevier Science.
- Vandenberghe R, Price C, Wise R, Josephs O, Frackowiak RSJ (1996) Functional anatomy of a common semantic system for words and pictures. *Nature* 383:254-256.
- Van Orden GC (1987) A ROWS is a ROSE: spelling, sound, and reading. *Mem Cogn* 15:181-198.
- Van Orden GC, Johnston JC, Hale BL (1988) Word identification in reading proceeds from spelling to sound to meaning. *J Exp Psychol (Learn Mem Cogn)* 14:371-385.
- Warburton E, Wise RJS, Price CJ, Weiller C, Hadar U, Ramsay S, Frackowiak RSJ (1996) Noun and verb retrieval by normal subjects. Studies with PET. *Brain* 119:159-179.
- Warren RM, Obusek CJ (1971) Speech perception and phonemic restorations. *Percept Psychophys* 9:358-362.
- Weisskoff RM, Davis TL (1992) Correcting gross distortion on echo-planar images. In: Book of abstracts, 11th Annual Meeting, Society for Magnetic Resonance in Medicine, p. 4515. Berkeley, CA: SMRM.
- Wernicke C (1874) *Der aphasische Symptomenkomplex*. Breslau: Cohn Weigert.
- Wise R, Chollet F, Hadar U, Friston K, Hoffner E, Frackowiak R (1991) Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain* 114:1803-1817.
- Witelson SF, Kigar DL (1992) Sylvian fissure morphology and asymmetry in men and women: bilateral differences in relation to handedness in men. *J Comp Neurol* 323:326-340.
- Wohlfart G, Lindgren A, Jernelius B (1952) Clinical picture and morbid anatomy in a case of 'pure word deafness'. *J Nerv Ment Dis* 116: 818-827.
- Wong EC, Bandettini PA, Hyde JS (1992) Echo-planar imaging of the human head using a three axis local gradient coil. In: Book of abstracts, 11th Annual Meeting, Society for Magnetic Resonance in Medicine, p. 108. Berkeley: SMRM.
- Wood NL, Cowan N (1995) The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry. *J Exp Psychol (Gen)* 124:243-262.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846-849.