

Parametrically Dissociating Speech and Nonspeech Perception in the Brain Using fMRI

Randall R. Benson

University of Connecticut Health Center and Haskins Laboratories

D. H. Whalen

Haskins Laboratories

Matthew Richardson

Haskins Laboratories and Yale University

Brook Swainson

University of Connecticut

Vincent P. Clark and Song Lai

University of Connecticut Health Center

and

Alvin M. Liberman¹

Haskins Laboratories

Candidate brain regions constituting a neural network for preattentive phonetic perception were identified with fMRI and multivariate multiple regression of imaging data. Stimuli contrasted along speech/nonspeech, acoustic, or phonetic complexity (three levels each) and natural/synthetic dimensions. Seven distributed brain regions' activity correlated with speech and speech complexity dimensions, including five left-sided foci [posterior superior temporal gyrus (STG), angular gyrus, ventral occipitotemporal cortex, inferior/posterior supramarginal gyrus, and middle frontal gyrus (MFG)] and two right-sided foci (posterior STG and anterior insula). Only the left MFG discriminated natural and synthetic speech. The data also supported a parallel rather than serial model of auditory speech and nonspeech perception. © 2001 Academic Press

This work was supported by a Faculty Research Grant to Randall R. Benson and NIH Grants HD-01994 and DC-00493 to Haskins Laboratories. This research is dedicated to the memory of Alvin M. Liberman. We thank Jennie Wakefield, Einar Mencl, Annette Jenner, Ken Pugh, Janet Werker, Athena Vouloumanos, Russ Poldrack, and David Caplan for helpful comments. We thank Khalil Iskarous for help with Fig. 2. We thank John Lackey, Jose Maisog, and Leslie Bernstein for technical support.

Address correspondence and reprint requests to Randall R. Benson, Department of Neurology, University of Connecticut Health Center, 263 Farmington Ave., Farmington, CT 06030–2017. Fax: (860) 679–1989. E-mail: benson@nso.uchc.edu.

¹ Deceased.

Key Words: brain; language; speech; phonetic; imaging; fMRI; parametric; Wernicke; auditory; perception.

INTRODUCTION

Fundamental to human language, speech perception, in the narrow sense, is the capacity of a listener to identify and appropriately utilize the phonetic categories from spoken utterances. Recent evidence suggests that perception of phonetic categories is present from birth (Bertoncini, Bijeljic-Babic, Blumstein, & Mehler, 1987) or shortly afterward (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Molfese & Molfese, 1979) but undergoes experience-dependent modification in the first year of life (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984). By age 5 years, without formal training, all neurologically normal children who have been exposed to language can perceive and produce speech at rates of 15–20 phonetic segments per second. These ontogenetic observations strongly suggest that speech, including speech perception, is a genetic predisposition in the brain similar to such perceptual processes as object recognition, sound localization, or stereopsis.

Since the primary source of information about the phonetic segments is acoustic, it has seemed reasonable to posit acoustic primitives for speech perception (Darwin, 1984; Diehl & Kluender, 1989; Kingston, 1992; Kluender & Greenberg, 1989; Stevens & Blumstein, 1978). Auditory theories of speech perception assume an initial representation in terms of acoustic characteristics which are matched to speech templates at a later stage of processing. The specific properties of speech perception are viewed as determined by psychophysical limitations of the auditory system. Consistent with auditory theories of speech perception, impairment of auditory speech comprehension, such as occurs in Wernicke's aphasia, was presumed to be on the basis of disruption of an auditory-based phonological decoding system located in the dominant temporal lobe posterior to Heschl's gyrus (Luria, 1970; Wernicke, 1874/1977). Despite this long held view, however, impairment of phonetic perception is usually not severe in Wernicke's aphasia (Basso, Casati, & Vignolo, 1977; Blumstein, Baker, & Goodglass, 1977; Cappa, Cavallotti, & Vignolo, 1981) and does not correlate well with auditory comprehension disturbance within or across patients (Auerbach, Allard, Naeser, Alexander, & Albert, 1982; Bachman & Albert, 1988; Blumstein, Baker et al., 1977; Blumstein, Cooper, Zurif, & Caramazza, 1977; Cappa et al., 1981). Furthermore, phonetic impairment is not restricted to Wernicke's aphasia but is described in all aphasic syndromes (Basso et al., 1977; Blumstein, Baker et al., 1977; Miceli, Caltagirone, Gainotti, & Payer-Rigo, 1978). On the basis of the poor correlation between auditory comprehension and phonetic perception, it was argued that preserved speech perception capacity in Wernicke's aphasia may be attributable to residual left and/or intact right hemisphere networks (Benson et al., 1973).

The lack of association between phonetic and speech comprehension deficits has lent support for alternative models of speech perception which, unlike auditory theories, consider the articulatory gestures that generate these sounds, rather than the sounds themselves, as primitives. These theories include the Motor Theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985, 1989), Direct Realist theory (Fowler, 1986, 1994; Fowler & Smith, 1986), and Articulatory Phonology (Browman & Goldstein, 1992). In these models, gestural units serve as the primary representation for both speaker and listener alike, and the acoustic signal is merely a carrier for a phonetic module which

extracts only those acoustic cues relevant to the perception of coarticulated gestures. Behavioral support for the gestural models includes the well-studied McGurk effect (McGurk & MacDonald, 1976), in which the perception of heard speech by a listener is found to be strongly influenced by lip reading (visual articulatory information). Additional support comes from experiments in which the same acoustic signal serves as the basis of speech and nonspeech percepts (duplex perception) but with radically different perceptual functions (Whalen & Liberman, 1987; Xu, Liberman, & Whalen, 1997). In one duplex perception experiment, above chance phonetic perception was accomplished despite the nonspeech percept being at chance or below the threshold of hearing, suggesting that a full acoustic analysis is not required for phonetic perception (Xu et al., 1997). According to the gestural models, *early* divergence of speech and nonspeech auditory processing would be reflected by similarly early diverging neural systems performing differing computations.

Currently there is little consensus regarding the neuroanatomy underlying speech perception. The best clinical model may be the acquired syndrome of "pure word deafness," in which a deficit in phonetic perception occurs in the absence of reading and speaking deficits. This syndrome has been found to occur with unilateral left posterior temporal lobe lesions, but more commonly, bilateral lesions of the posterior temporal lobes are responsible (Auerbach et al., 1982), supporting a role for the right temporal lobe in speech perception. Disagreement exists as to the existence of "pre-phonemic temporal" (auditory processing) and "phonemic" behavioral subtypes (Buchman, Garron, Trost-Cardamone, Wichter, & Schwartz, 1986) of this syndrome, but too few cases have been described to form definite conclusions.

Since the mid-1970s, however, functional neuroimaging techniques including positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have afforded the opportunity to study human language with unprecedented safety and detail. Several of these studies have addressed phonetic/phonological processing (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin et al., 1998; Binder, Frost, Hammeke, Rao, & Cox, 1996; Demonet et al., 1992; Demonet, Price, Wise, & Frackowiak, 1994; Fiez, Raichle, Miezin, & Petersen, 1995; Mazoyer et al., 1993; Paulesu, Frith, & Frackowiak, 1993; Petersen, Fox, Posner, Mintun, & Raichle, 1989; Poldrack et al., 1999; Pugh et al., 1997; Rumsey, 1997; Sergent, Zuck, Levesque, & MacDonald, 1992; Zatorre, 1997), implicating brain areas including Broca's area (BAs 44 and 45), premotor cortex (BA 6), prefrontal cortex (BAs 11 and 46), superior temporal gyrus (BAs 42 and 22), middle (BAs 21) and inferior temporal gyri, and temporoparietal cortex (BAs 7, 39 and 40). A metaanalysis by Poeppel (1996) of five of these studies using some type of rhyme judgment task (Demonet et al., 1992; Fiez et al., 1995; Paulesu et al., 1993; Petersen et al., 1989; Sergent et al., 1992) found little convergence of activated regions across these studies. Subsequent article by Poeppel et al. (1996) and Demonet et al. (1996) discussed possible reasons for the differing results across these studies including differing hardware, analysis methods, and subject pool. Furthermore, specifics of the tasks and response type also differed across the studies and were a likely source of variance. Blumstein (1977), among others, has demonstrated that even the seemingly minor difference between discrimination and identification of pairs of sublexical stimuli produces dramatic differences in performance, indicating the different neural processing employed by these tasks. Similarly, attention to the segmental or suprasegmental level of phonological structure changes the specific analytic strategies and neural systems (e.g., working memory) deployed to complete the task (Hickok & Poeppel, 2000).

On the other hand, common features in the experimental designs of the five studies reviewed by Poeppel were the use of the subtraction method and an overt response by subjects. The logic and limitations of subtraction in functional neuroimaging have

been discussed previously (Chertkow & Bub, 1994; Poeppel, 1996; Sidtis, Strother, Anderson, & Rottenberg, 1999) and relate to the nonadditivity of cognitive processes. The use of an overt response in conjunction with subtraction may falsely assume that perceptual and response processing do not interact with each other. This issue becomes particularly important when preattentive perceptual processing, such as occurs while listening to spoken speech, is the process of interest. If perceptual and response processing do interact to a significant degree, then almost any response would preclude an accurate assessment of preattentive perceptual processing. Regarding phonetic perception in particular, where one goal has been to determine whether speech-specific (Liberman & Whalen, 2000) or, alternatively, general auditory neural machinery (Stevens & Blumstein, 1978) decode the speech signal, performance of a task could add an extra layer of (task-dependent) difference between speech and nonspeech, obscuring detection of true differences in the preattentive perception of speech and nonspeech.

The objective of the current investigation was to identify candidate brain regions comprising a neuronal network for speech perception. We employed a passive listening task with a parametric design and multivariate analysis methods to obviate the potential pitfalls of subtraction and overt responses. The general hypothesis was that specific regions of perisylvian cortex will respond preferentially to speech and that these regions, or more likely a subset of these regions, will also respond predictably to parametrically varying stimulus phonological content distinctly from variations in acoustic content, reflecting a role in phonological decoding. Still other regions, we predicted, would respond selectively to differences in acoustic content distinct from phonological content. Similar types of parametric paradigms have been used recently in fMRI mapping (see Buchel, Holmes, Rees, & Friston, 1998). In principle, this should allow for functional and anatomical separation of phonological from nonphonological auditory processing regions. In practice, we presented semantically meaningless auditory stimuli, matched for duration and amplitude, to 12 normal subjects during fMRI scanning. Specifically, a three-factor design was employed consisting of speech/nonspeech, complexity, and natural/synthetic dimensions. The speech/nonspeech dimension contrasted nonword monosyllables ("speech") with stimuli composed of complex periodic waveforms not perceived as speech ("nonspeech"). The complexity dimension consisted of three levels for both the speech and nonspeech stimuli. The natural/synthetic dimension consisted of natural and synthetic tokens of both speech and non-speech stimuli. Our intention in the latter comparison was to compare the brain's response to acoustically differing but phonologically equivalent (i.e., same targets) speech and whether a similar response is elicited by the equivalent difference in the nonspeech stimuli.

MATERIALS AND METHODS

Subjects

Twelve right-handed normal subjects (6 male and 6 female) ages 21–45 years (mean = 31), with no known language impairment participated in this study. All were native English speakers with no first-degree sinistrality. Subjects gave written informed consent, and the study was approved by the Institutional Review Board at the University of Connecticut Health Center.

Procedure

The task involved passive listening to speech and nonspeech stimuli arranged in blocks. A $2 \times 2 \times 3$ factorial design was used where speech (speech vs. nonspeech), naturalness (natural vs. synthetic), and complexity (three levels) were independent variables. Presentation of stimulus conditions was pseu-

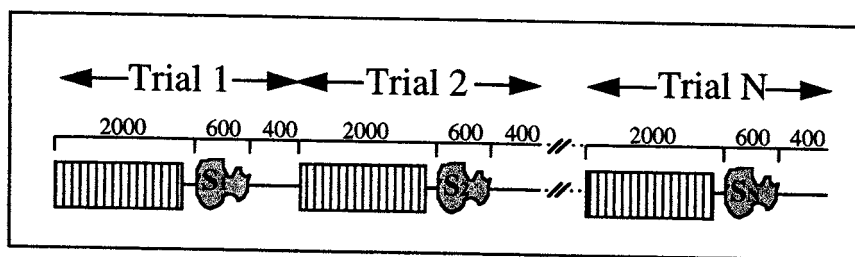


FIG. 1. Schematic of experimental trials. Each 3000-ms trial consisted of an initial 2000-ms offset during which 20 gradient pulses occurred, followed by a 600-ms auditory stimulus, followed by a second offset of 400 ms, which preceded the next train of scanner pulses. Stimulus presentation was triggered by TTL pulse from scanner electrically coupled to the PC.

derandomized and counterbalanced within and between subjects. Subjects were scanned over two sessions on different days. Stimuli were played through a Tucker Davis Technologies (TDT; Gainesville, FL) array processor using SigPlay software (TDT). Stimulus delivery to the subject utilized an auditory amplifier, transducer, and an air conduction system with high-attenuating headphones (Resonance Technology, Van Nuys, CA). Subjects were instructed to "listen to each sound presented through the headphones" with eyes closed.

Stimuli

All speech and nonspeech stimuli were 600 ms in duration, digitally sampled at 44.1 kHz, and matched for root-mean-squared power. Image acquisition and stimulus presentation did not overlap so as to minimize acoustic masking of the stimuli. Each stimulus trial lasted 3000 ms, consisting of an initial offset of 2000 ms, during which the 20 image planes were collected, followed by stimulus presentation of 600 ms, and separated from the next trial by a 400-ms final offset (see Fig. 1). Each of the 12 runs consisted of 12 blocks, each 30 s in duration (10 trials), which alternated between speech and nonspeech, natural and synthetic, and the three levels of complexity. Each run also contained three blocks (blocks 1, 7, and 15) during which no stimulus was presented.

Speech Stimuli. The speech stimuli consisted of nonsense syllables, which varied in phonetic complexity from steady-state vowels (V) to consonant-vowel (CV) syllables to consonant-vowel-consonant syllables (CVC). The natural speech stimuli were all produced by a male speaker of a midwestern dialect of American English, and the synthetic stimuli were generated with a cascade formant resonance synthesizer with values based on the natural productions. There were five instances of each category: Vs—/a, e, i, o, u/*; CVs—/ta, ze/, gi/, vo*, du*/; and CVCs—/tag, ze'k, gi'p, vo*s, du't/. Only vowels which occur in open syllables were used, since these occur naturally without a following consonant, they would not seem inappropriate in isolation as would, say, the vowel of "hit." The averaged spectral profile of all three speech types was virtually identical (see Fig. 2A).

The synthetic syllables were based on the natural versions, using SenSyn PPC v1.0b3 (Sensimetrics Corp., Somerville, MA). Formant values were copied from the natural tokens and modified until they reliably conveyed the intended phonetic categories to naive listeners (see Appendix for details). Spectral content of synthetic speech was virtually identical to the natural speech up to 3500 Hz above which naturals were more intense (see Fig. 2B).

Non speech stimuli. These stimuli were, by design, not perceivable as speech. They were tonal, increasing in complexity from notes to chords to chord progressions. The spectral profiles of the speech and nonspeech signals were also extremely similar (see Fig. 2C); the broad spacing of the harmonics for the nonspeech in the higher frequency regions led to a sparser representation, seen in the fluctuations in the nonspeech spectrum. The spectral profile of the three levels of complexity in the nonspeech were also virtually identical (see Fig. 2D). A natural/synthetic dimension existed for the nonspeech stimuli, similar to the speech stimuli. The "natural" stimuli were piano notes and chords produced by a piano synthesizer. Although these were technically synthetic, they were based on sampling of natural piano production. The ability to control the duration of these stimuli made them more natural than the edited versions of naturally struck piano notes would have been. The synthetic stimuli consisted of sinusoids of the fundamental and first four harmonics of each note.

The natural piano notes and chords were recorded on a Yamaha Clavinova Digital Piano CLP-611 ("Piano 1" sound) Roland RD-250s. The piano notes in the experiment were C4, D4, E4, F4, and G4. Each of the five piano chords used one of these piano notes as its lowest note. The piano chords were as follows: the major triad (1-4-7) C4E4G4, the minor triad (1-3-7) D4F4A4, the diminished triad (1-3-6) E4G4B^b4, the major triad 2nd inversion (1-5-9) F4B^b4D5, and the minor triad 2nd inversion

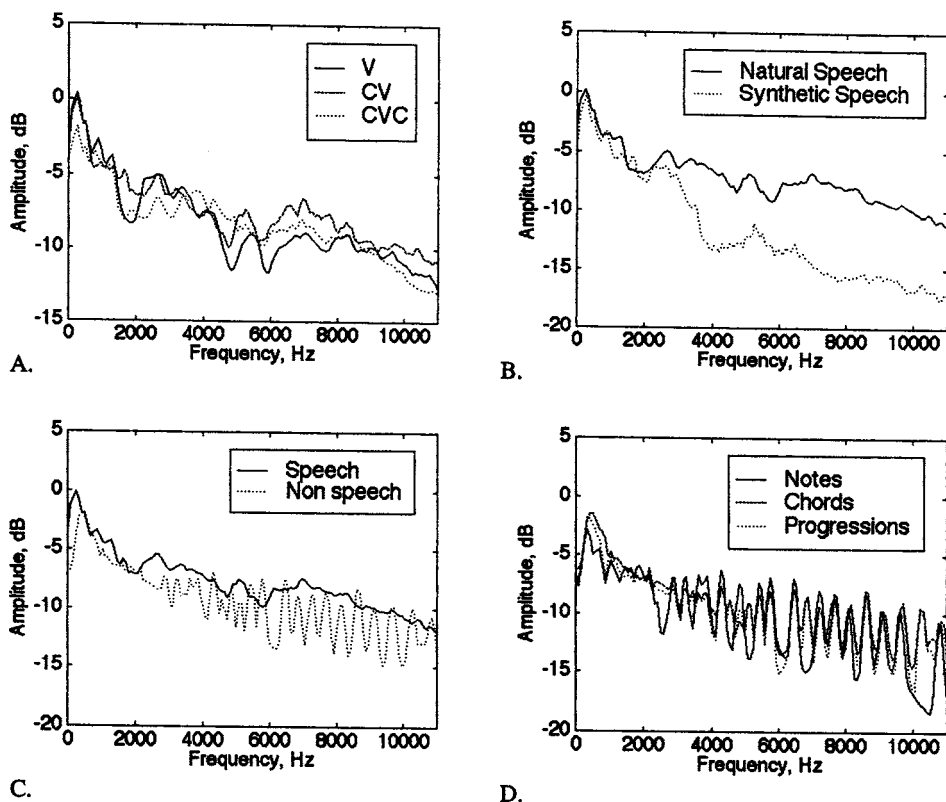


FIG. 2. Averaged power spectra for acoustic stimuli. (A) Speech stimuli. (B) Natural and synthetic speech stimuli. (C) Speech and nonspeech stimuli. (D) Nonspeech stimuli. V = Vowel; C = Consonant.

(1-5-8) G4C5E5 (see Appendix). The chord progressions contained three chords, the first of which was one of the five chords just described. The next chord had a higher first note, and the third chord had a first note intermediate in pitch between that of the first and second chords.

The synthetic stimuli were created using SWS (written by P. E. Rubin at Haskins Laboratories). The 600-ms stimulus duration included 3-ms intensity onset and offset ramps. Harmonic complexes were made by combining the fundamental with the next four harmonics of pure tones. The five pure tones' fundamentals were equal to those of the piano notes C4, D4, E4, F4, and G4 (261.63, 293.66, 329.63, 349.23, and 392.00 Hz, respectively). Each harmonic decreased in input amplitude by 0.2, starting from a nominal "full" amplitude of 1.0. The five synthetic chords also used the frequencies of the natural stimuli produced with the piano. They employed the pure tones/sinusoids C4 through G4 as the lowest frequency note of the chord. The frequencies for each chord are given in the Appendix. Finally, the harmonic chord progressions were constructed from the harmonic chords in the same way as for the piano chord progressions.

Imaging

Imaging was performed on a Siemens Magnetom Vision scanner (Siemens Medical Systems, Erlangen, Germany) operating at 1.5T and equipped for single shot echoplanar imaging. Subjects were scanned on 2 separate days, consisting of anatomical scanning and six experimental runs on each day. On the first scanning day experimental runs included an additional run of noise vs silence which preceded the others and allowed the subjects to accommodate to the scanner environment. Subjects were instructed to close their eyes during experimental runs. Twenty contiguous, interleaved oblique-axial slices 7 mm thick were collected using a gradient-echo echoplanar sequence (flip angle = 90° , TE = 40 ms, TR = 3 s, FOV = 19.2 cm, 64×64 matrix in about 1.8 s and every 3 s. Each scanning run collected 132 image volumes over 396 s. Data writing required about 3 min between successive runs. An electrical pulse emitted by the scanner with each train of images triggered the stimulus presentation from a Micron PC running SigPlay software (Tucker-Davis Technologies, Gainesville, FL). Interleaving the scanner noise with the stimulus presentation ensured that subjects could easily hear each stimulus without interfer-

ence from the 85-dB scanner pulse noise at the external ear. Interleaving also allowed for consistent, if slight, separation of the fMRI responses evoked by scanner pulses and stimuli. Avoidance of acoustic masking is particularly relevant for the synthetic speech where the phonetic cues are less robust than for the natural analogs. All studies included anatomical scout images and a high resolution T1-weighted 3D inversion recovery sequence (1.75 mm thick, TI = 300 ms, TR = 11.4 ms, TE = 4.4 ms, 152 slices, flip angle = 15°).

Data Analysis

The functional images of each study were realigned to the first volume using SPM96 (Friston et al., 1995) with adjustment for first-order movement effects and spin history (Friston, Williams, Howard, Frackowiak, & Turner, 1996). Images were then spatially smoothed in-plane using a Gaussian kernel of FWHM = 6 mm using Functional Imaging Data Analysis Platform (FIDAP). Smoothed images were then spatially normalized using SPM96 using both linear and nonlinear matching. As a first step, one subject's functional (T2-weighted EPI) images were normalized to the SPM T2-weighted template, but the remaining 11 subjects' data were normalized to the first subject's normalized images, since this produced more reliable normalizations than the SPM template. The data from the 12 pseudorandomized runs for each subject over the two scanning sessions (1560 volumes) were reordered to be the same and then averaged over the 12 subjects. A multiple-regression module (Haxby, Maisog, & Courtney, 2001) within FIDAP was used to partial out linear drift within runs and baseline shifts between runs on the averaged data prior to calculation of statistical maps. Multiple-regression and ANOVA routines within FIDAP were used to assess the statistical significance of stimulus-evoked changes in MRI signal intensity on each voxel after correction for global signal change in the images. The regressor functions for each independent variable were delayed and smoothed using the empirically derived 4.8-s delay and 1.8-s dispersion in order account for hemodynamic effects (Maisog, Clark, Courtney, & Haxby, 1995). The first two time points of each run were discarded since these were obtained prior to T1 stabilization. Multiple regression was used to compute main effects in the data with ANOVA used to examine significant interactions (i.e., between speech and nonspeech). The multiple-regression *F* ratio map was transformed to a *Z* map after adjustment for temporal autocorrelation using the experimentally derived correction factor (0.6 for TR = 3s) (Maisog et al., 1995). Regions of interest (ROIs) were located using a test of spatial contiguity in FIDAP by identifying clusters of at least 50 contiguous voxels (volume = 3.2 cc) significant at $Z > 3.09$ ($p < .001$), resulting in a region-wise probability of $p < .05$ after accounting for spatial smoothing. For each ROI attaining significance, we reported the anatomical location, peak *Z* score, and the Brodmann's area (BA) of the local maximum of the cluster. Since the SPM96 template is larger than the atlas of Talairach and Tournoux (1988), a linear transformation (using SPM96) of the SPM96 template to the SPM95 PET template (which is closer to the atlas of Talairach and Tournoux, (1988) was performed to better approximate canonical Talairach space. Individual ROIs were then assessed for speech or nonspeech selectivity by analyzing the behavior (*Z* score) of the region across a set of independent contrasts (conjunction analysis): (1) speech vs rest, (2) speech vs nonspeech, (3) speech complexity, (4) interaction of speech status (speech or nonspeech) and complexity, (5) nonspeech vs rest, and (6) nonspeech complexity.

RESULTS

Analysis by Contrast

Auditory vs Rest. For the auditory stimulus (collapsed across all stimulus types) vs rest comparison, several regions were activated. These regions included bilateral prefrontal, left inferior frontal gyrus, i.e., Broca's area (BA 44/45); bilateral temporal (superior, middle, and inferomedial); right greater than left ventral extrastriate cortex (BAs 37, 18 and 19); and cerebellar locations. A graph of the grand averaged, normalized signal for the ROI in left inferior frontal gyrus (and four other ROIs) revealed that the chord progressions and consonantal syllables elicited the highest signal (see Fig. 3). Comparison of nonspeech vs baseline (see Table 1) and speech vs baseline maps (see Table 2) revealed that speech vs baseline evoked greater activation in posterior superior temporal gyrus (pSTG) bilaterally (see Tables 1 and 2) while nonspeech vs baseline produced more activation than speech in pre- and postcentral gyri on the left.

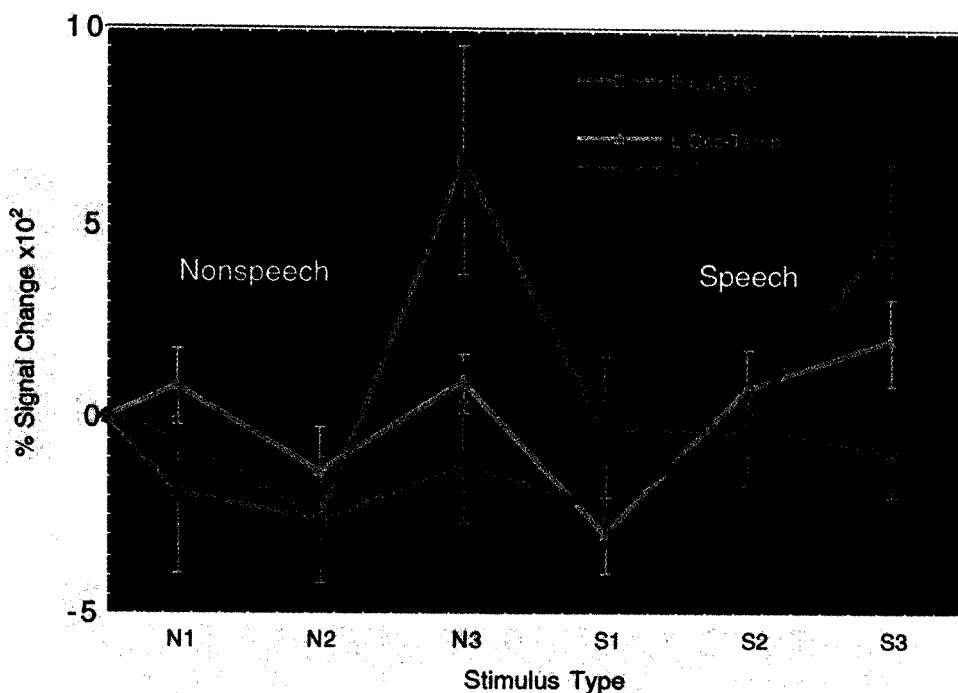


FIG. 3. Grand averaged signal data averaged by condition (collapsing over the natural/synthetic difference) after reordering blocks which were pseudorandomized across the 12 subjects. Graph of data from four ROIs demonstrates differing selectivities. Bilateral aSTG is most sensitive to nonspeech complexity and the left temporoparietal ROI is most sensitive to the speech complexity dimension; the other two ROIs were intermediate. Error bars indicate standard error of the mean. Percentage of signal change was given relative to resting baseline. N1 = notes and harmonic complexes; N2 = chords; N3 = chord progressions; and S1 = Vs, S2 = CVs, and S3 = CVCs; B/L aSTG = bilateral anterior Superior Temporal Gyrus; L Inf. Frontal = left inferior frontal; L Occ-Temp = left occipitotemporal; L Temp-Par = left temporoparietal.

Speech vs nonspeech (see Table 3a) (Fig. 4A). A regression Z map of the main effect of speech (speech vs nonspeech, collapsing over natural/synthetic and complexity contrasts) revealed that there were no subcortical foci which demonstrated a preference for speech compared to nonspeech or vice versa. On the other hand, speech was more active than nonspeech for bilateral foci in posterior temporal cortex (BAs 42 and 22). This large activated region (Fig. 4A1) extended anteriorly to include the most posterior portion of Heschl's gyrus, inferiorly into middle temporal gyrus (MTG) (BA 21), and posteriorly to the anterior angular gyrus (BA 39). This region had a greater dorsal extent ($z = +28$), including the supramarginal gyrus (BA 40) (Fig. 4A2), on the left compared with the right hemisphere ($z = +17$). Within the inferior frontal gyrus, only a small focus in Broca's area (BA 45; $-45, 25, 2$; mean $Z = 4.1$) was more active for speech than nonspeech (Fig. 4A3) but this was only a trend ($p = .31$). An additional focus in left posterior middle frontal gyrus (BA 9) bordering premotor cortex (BA 6) was more active for speech than nonspeech (Fig. 4A4). Nonspeech produced greater activation than speech for foci centered in the right supramarginal gyrus (BA 40) (Fig. 4A5) and bilateral anterior STG (aSTG) just anterior to Heschl's gyrus (BA 22) (Fig. 4A6), which, on the left, extended to include the anterior portion of Heschl's gyrus. The average signal plot of this bilateral anterior STG region revealed that the chord progressions, by far, elicited the greatest signal increase (see Fig. 3), which was responsible for the greater activation for nonspeech.

TABLE 1
Nonspeech vs Baseline

Z max	x	y	z	Hemisphere	Lobule	Brodmann's area
Increases						
7.07	38	-73	-1	Right	Inferior occipital gyrus	18
6.54	20	-81	-15	Right	Fusiform gyrus	19
6.5	20	-93	23	Right	Cuneus	19
6.34	27	-69	-18	Right	Fusiform gyrus	19
6.34	31	-93	2	Right	Middle occipital gyrus	18
5.88	-22	-69	-7	Left	Lingual gyrus	18
5.69	10	51	13	Right	Medial frontal gyrus	10
5.67	-22	28	29	Left	Middle frontal gyrus	46
5.67	-15	55	17	Left	Superior frontal gyrus	10
5.4	31	-50	-13	Right	Fusiform gyrus	37
5.17	52	-19	23	Right	Postcentral gyrus	2
5.16	45	-23	27	Right	Inferior parietal lobule	40
5.13	-54	-19	27	Left	Postcentral gyrus	3, 1, 2
5.07	-36	-54	-10	Left	Fusiform gyrus	37
5.03	-25	-96	5	Left	Middle occipital gyrus	18
4.99	-50	-30	44	Left	Inferior parietal lobule	40
4.96	20	32	5	Right	Middle frontal gyrus	10 (deep)
4.91	13	32	26	Right	Cingulate gyrus	32
4.85	-36	28	29	Left	Middle frontal gyrus	46
4.71	-47	24	12	Left	Inferior frontal gyrus	45
4.55	-15	47	34	Left	Superior frontal gyrus	9
4.43	-36	35	16	Left	Middle frontal gyrus	46
4.43	-15	-58	14	Left	Posterior cingulum	31
4.4	28	-36	42	Right	Inferior parietal lobule	40
4.28	-36	-15	24	Left	Postcentral gyrus	3, 1, 2 (deep)
4.27	-36	-38	47	Left	Inferior parietal lobule	40
4.07	-29	4	28	Left	Precentral gyrus	6
Decreases						
5.84	-1	-50	50	Left	Precuneus	7
5.58	3	-85	48	Right	Precuneus	7
5.52	-4	-62	53	Left	Precuneus	7
5.43	13	-85	48	Right	Precuneus	7
5.42	-1	-27	27	Left	Precuneus	7
5.31	3	-7	13	Left	Caudate	—
5.19	-11	-7	17	Left	Cingulate gyrus	23
4.75	45	15	35	Right	Middle frontal gyrus	9
4.52	44	-50	51	Right	Inferior parietal lobule	40
4.39	2	12	59	Right	Superior frontal gyrus	6

Note. Z max, Z score of maximal peak; stereotaxic coordinates are derived from Talairach and Tournoux (1988) and refer to medial-lateral position x (mm) relative to midline (positive is right), anterior-posterior position y (mm) relative to the anterior commissure (positive is anterior), and superior-inferior position z (mm) relative to the commissural line (positive is superior).

Complexity

Nonspeech (see Table 3b) (Fig. 4b). Only Heschl's gyrus (BA 41) and immediately surrounding cortex (i.e., BAs 22 and 42) bilaterally increased activity with increasing nonspeech acoustic complexity (Fig. 4B1). No region of the temporal lobe posterior to secondary auditory cortex (BA 42) showed correlated activity to the nonspeech complexity dimension. Deactivation was seen with increasing non-speech complexity in right and left prefrontal cortex (Fig. 4B2) and left medial occipitotemporal cortex (Fig. 4B3).

TABLE 2
Speech vs Baseline

Z max	x	y	z	Hemisphere	Location	Brodmann's area
Increases						
8.91	-57	-42	12	Left	Superior temporal gyrus	22
8.72	-61	-15	6	Left	Superior temporal gyrus	22/42
8.59	56	-11	-4	Right	Middle temporal gyrus	21
8.46	-54	-46	15	Left	Superior temporal gyrus	22
8.06	56	-15	6	Right	Superior temporal gyrus	22/42
7.98	-61	-23	20	Left	Postcentral gyrus	40
7.57	52	-23	2	Right	Superior temporal gyrus	22
6.28	41	-73	-4	Right	Inferior occipital gyrus	19
6.13	-18	55	13	Left	Superior frontal gyrus	18
6.13	17	-81	-15	Right	Lingual gyrus	10
6.04	10	51	13	Right	Medial frontal gyrus	10
5.19	-47	24	12	Left	Inferior frontal gyrus	45
5.13	10	43	27	Right	Medial frontal gyrus	9
5.12	20	-93	23	Right	Cuneus	19
5.05	-18	28	26	Left	Anterior cingulate gyrus	32
5.04	31	-69	-18	Right	Fusiform gyrus	19
4.98	-43	16	32	Left	Middle frontal gyrus	9
4.80	17	28	22	Right	Anterior cingulate	32
4.58	-40	28	1	Left	Inferior frontal gyrus	47
4.53	-19	-68	-5	Left	Lingual gyrus	19
4.50	27	-96	2	Right	Cuneus	18
4.38	41	-93	9	Right	Medial occipital gyrus	19
4.25	-32	43	2	Left	Subgyral	10
4.10	20	16	15	Right	Caudate	—
Decreases						
6.07	41	-50	50	Right	Inferior parietal lobule	40
6.05	-50	-69	-43	Left	Posterior lobe cerebellum	—
5.95	3	-11	66	Right	Medial frontal gyrus	6
5.94	-1	-50	50	Left	Precuneus (PCu)	7
5.94	13	-81	48	Right	Precuneus (PCu)	7
5.65	20	8	60	Right	Middle frontal gyrus	6
5.49	-25	-42	64	Left	Postcentral gyrus	5
5.47	24	4	56	Right	Subgyral	6
5.33	-8	-65	56	Left	Superior parietal lobule	7
5.17	-1	12	60	Left	Superior frontal gyrus	6
5.12	-1	12	39	Left	Cingulate gyrus	23
4.96	3	-11	13	Right	Thalamus	Medial dorsal N.
4.92	-15	-81	45	Left	Precuneus (PCu)	7
4.56	-15	-11	17	Left	Thalamus	Ventral lateral N.
4.52	-1	-27	27	Left	Cingulate gyrus	23
4.27	20	-38	65	Right	Postcentral gyrus	3
4.21	27	-50	61	Right	Superior Parietal lobule	7

See note for Table 1.

Speech (see Table 3c) (Fig. 4c). Increasing speech complexity was associated with increasing activation of superior temporal sulcus (STS) bilaterally (Fig. 4C1) and more posterior left STG (BA 22) bordering angular gyrus posteriorly (Fig. 4C2). Additional foci were located in right supramarginal gyrus (Fig. 4C3), left inferior occipitotemporal cortex (Fig. 4B4), and the right anterior insula (Fig. 4C5). Primary and secondary auditory cortex *did not* demonstrate correlated activity to the speech complexity dimension. Signal decrease associated with increased speech complexity

TABLE 3a
Speech vs Nonspeech

Z max	x	y	z	Hemisphere	Location	Brodmann's area
Increases						
14.45	48	-30	2	Right	Superior temporal sulcus	22
13.24	56	-15	-1	Right	Superior temporal gyrus	22
11.09	-57	-42	12	Left	Superior temporal gyrus	22
10.96	-57	-23	6	Left	Superior temporal gyrus	42
9.73	-64	-23	16	Left	Supramarginal gyrus	40
6.91	53	-27	2	Right	Superior temporal sulcus	22
6.87	-57	-32	10	Left	Superior temporal gyrus	42/22
4.13	-47	8	35	Left	Middle frontal gyrus	9
4.09	63	-50	11	Right	Posterior superior t. gyrus	22
3.93	11	-75	-30	Right	Cerebellum	—
3.86	-49	8	37	Left	Middle frontal gyrus	9
3.73	-54	1	46	Left	Middle frontal gyrus	6
3.72	-10	49	-4	Left	Medial frontal gyrus	10
3.50	-8	47	-8	Left	Medial frontal gyrus	10
3.46	-11	51	-5	Left	Medial frontal gyrus	10
Decreases						
5.10	44	1	-3	Right	Anterior superior t. gyrus	22
4.98	-49	-6	1	Left	Anterior superior t. gyrus	22
4.20	30	-49	50	Right	Supramarginal gyrus	40
4.02	-52	-36	45	Left	Supramarginal gyrus	40
3.94	52	-31	28	Right	Supramarginal gyrus	40
3.78	13	-76	49	Right	Precuneus	7
3.78	33	-7	56	Right	Superior frontal gyrus	6
3.75	-28	38	32	Left	Superior frontal gyrus	9
3.73	5	3	64	Right	Medial frontal gyrus	6
3.67	-28	1	54	Left	Superior frontal gyrus	6
3.63	32	-86	16	Right	Occipital gyrus	19

See note for Table 1.

occurred in cerebellar vermis (Fig 4C6), bilateral superior and medial frontal gyri (BAs 9 and 10) (Fig. 4C7), and left dorsomedial thalamus (Fig. 4C8). Of these deactivations, only the prefrontal deactivation replicated with the nonspeech condition. Figure 5 shows a surface rendering of the positively correlating activation foci from Figs. 4B and 4C, with speech complexity in red and nonspeech complexity in green.

Interaction of speech status with complexity (see Table 3d) (Figure 4D). No brain region demonstrated a congruent response to both speech and nonspeech complexity dimensions. Brain regions demonstrating a dissociation to the speech and non-speech complexity dimensions were detected using a two-way analysis of variance (speech status x complexity). Regions demonstrating sufficiently different response functions to the two complexity dimensions were indicated by significance of the interaction term. The sign of the resulting *F*-map was given by the sign of the voxelwise subtraction of the *Z*-maps of speech and nonspeech complexity (i.e., Fig. 4C minus Fig. 4B), since this difference approximates the difference between the slopes of the respective regression lines for the three levels of complexity. Of the foci which tracked either the speech or nonspeech complexity dimensions (Tables 3a and 3b), eight showed significant interactions between speech and nonspeech. Four foci showed sensitivity to speech complexity greater than nonspeech complexity: (1) left pSTG (BA 22; -60, -52, 12) (Fig. 4D1), (2) left angular gyrus (BA 39; -52, -68, 16) (Fig. 4D2) extending posteriorly to lateral extrastriate cortex BA 19, (3) left inferior occipito-

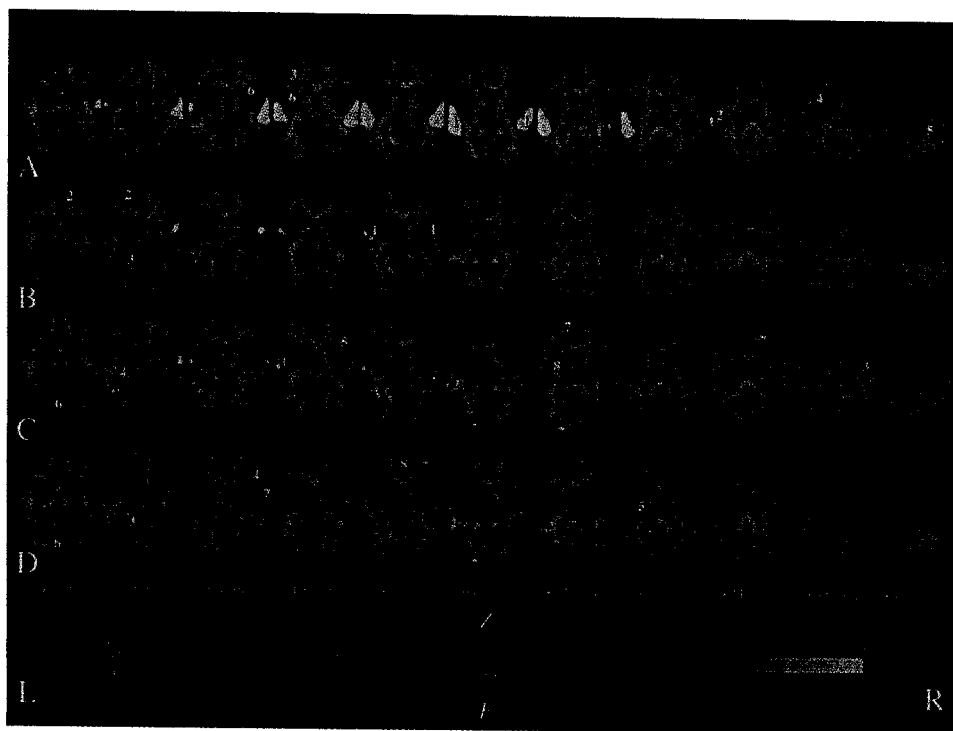


FIG. 4. Grand-averaged, spatially normalized, thresholded statistical maps colorized and overlaid onto a single subject's anatomical images for all 12 subjects on (A) speech vs nonspeech, (B) nonspeech complexity, (C) speech complexity, (D) Interaction of speech status and complexity. (A–C) Statistical maps are Z scores (top color bar scale) while D map is F ratio (bottom color bar scale). Red/yellow scale assigned to positive values; blue/green scale assigned to negative values. From left to right across a row of slices is bottom to top of the brain with axial slice location in Talairach z dimension (inferior/superior) given in millimeters just below row D. Left side of the brain is left of the page. White numbers on images are explained in the text.

TABLE 3b
Nonspeech Complexity

Z max	x	y	z	Hemisphere	Location	Brodmann's area
Increases						
5.8	48	-3	0	Right	Superior temporal gyrus	22
4.65	-54	-7	6	Left	Superior temporal gyrus	22
4.49	-43	-15	6	Left	Extranuclear	13, 41
3.44	38	-19	9	Right	Extranuclear	13, 41
3.31	-64	-15	13	Left	Transverse temporal gyrus	42
3.3	-61	-19	16	Left	Postcentral gyrus	40
Decreases						
3.89	-38	-52	-5	Left	Subgyral	19/37
3.85	-12	24	35	Left	Cingulate gyrus	32
3.66	16	34	23	Right	Anterior cingulum	32
3.60	21	19	1	Right	Extranuclear	L. caudate/ant. insula

See note for Table 1.

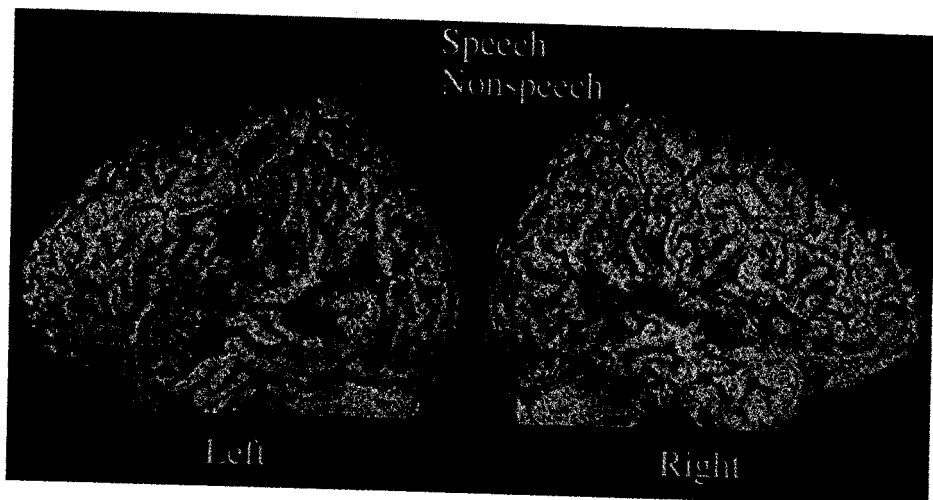


FIG. 5. Surface rendering of Fig. 4b and 4c with nonspeech complexity in green, speech complexity in red, and overlap in yellow. Only positive tails depicted. The figure depicts an anterior-to-posterior gradient for nonspeech and speech complexity, respectively, as well as the more distributed representation for speech compared with nonspeech bilaterally in the brain. Note also the minimal overlap between speech and nonspeech complexity areas indicating largely dissociated processing of speech and nonspeech acoustic complexity in the brain.

nonspeech could be on the basis of acoustic differences, since the speech stimuli (CV and CVC) contain rapid modulations in spectra not present in the nonspeech. On the other hand, the greater response to speech, including the steady-state vowel syllables, suggests that both phonetic and acoustic processing may be occurring within these regions. Finer inspection of these regions, in fact, revealed that this large

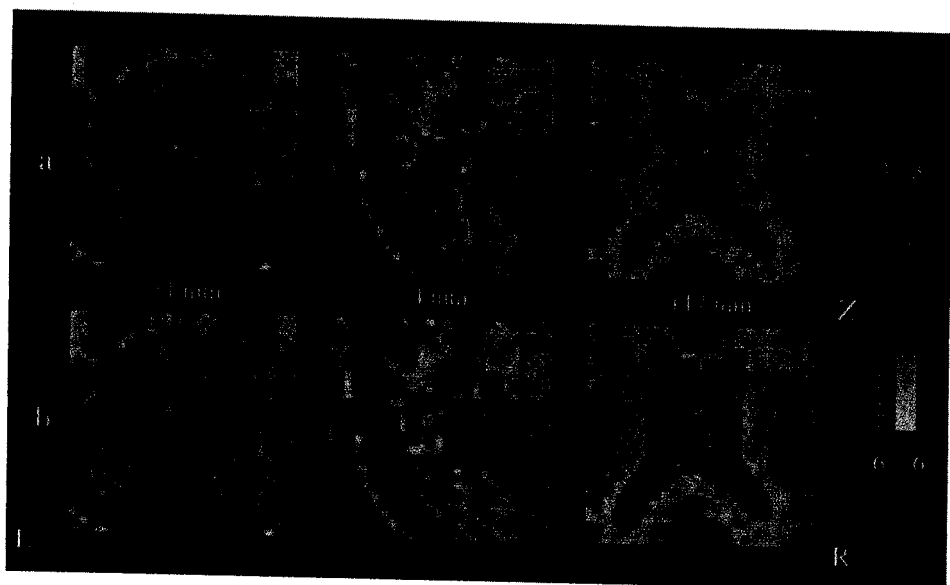


FIG. 6. Three ROIs respond in opposite fashion to nonspeech (a) and speech (b) complexity, respectively. Activity in the cerebellar vermis ($z = -11$ mm) and left dorsal thalamus ($z = +17$ mm) was positively correlated with nonspeech complexity and negatively correlated with speech complexity, while the left occipitotemporal region was positively correlated with speech complexity and negatively correlated with nonspeech complexity. Left side of brain is left of the page.

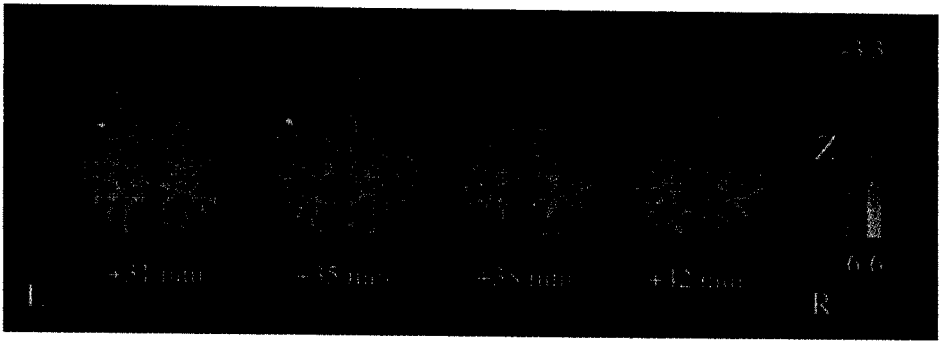


FIG. 7. Main effect of natural vs synthetic speech showing unilateral left frontal focus centered on the precentral sulcus between middle frontal gyrus BA 9 and premotor cortex BA 6. From left to right of the page is bottom to top of the brain with axial slice location in Talairach z-dimension (inferior/superior). Voxels with significantly greater signal for natural speech ($z > 3$) are colored red to yellow. Greater signal for synthetic speech colored blue to green.

bilateral region contained multiple functional domains (four on the left, and three on the right), indicated by multiple local maxima (see Table 2) which responded differently to the other contrasts discussed below.

Several studies in PET and fMRI have attempted to image phonological processing regions (Belin et al., 1998, 2000; Binder et al., 1994; 1996; Celsis et al., 1999; Demonet et al., 1992; Fiez et al., 1995; Mazoyer et al., 1993; Paulesu et al., 1993; Sergent et al., 1992; Tzourio et al., 1997; Wise et al., 1991; Zatorre, 1997; Zatorre, Evans, Meyer, & Gjedde, 1992; Zatorre, Meyer, Gjedde, & Evans, 1996). Only three of these however, (i.e., Fiez et al., 1995; Binder et al., 1996; Celsis et al., 1999), compared passive listening to speech with nonspeech and reported group averaged results in standard space, allowing for comparison of results with the current study.

Fiez et al. (1995), using PET (Fiez et al., 1995), found no difference in the regions activated between speech (vowels, CVs, CVC words) and nonspeech (tone triplets) stimuli in either passive listening or target detection modes. In their study, three bilateral regions along the superior temporal gyrus, but none posterior to secondary auditory cortex, (BA 42) activated significantly to any of their stimulus conditions. Since our study *did* find temporoparietal activation despite using monosyllabic stimuli similar to Fiez et al., which were also repeated many times across runs, we believe the critical difference between the two studies was a difference in total stimulus duration (10-fold difference) and the number of trials per condition (4-fold difference). This underscores the practical importance of ensuring adequate sampling of experimental conditions before interpreting negative results, which may be a strength of fMRI compared with PET, since PET requires a washout period of about 10 min between runs. FMRI, on the other hand, requires neither contrast agent nor washout period, allowing for maximal data collection.

On the other hand, Binder et al. (1996), imaging only the left hemisphere, did find the posterior temporal cortex to differentially treat monosyllabic words and sinewave tones on a passive listening task, finding that a focus centered on the left STS ($-56, -12, -2$) activated more for the words. Similarly, for the speech vs nonspeech contrast, our data revealed large bilateral foci in pSTG which *included* the STS bilaterally. Despite the convergent results, it remains unknown whether these bilateral activations are a response to differential acoustic or phonetic features, since in each study the speech and nonspeech stimuli differed both acoustically and phonetically. Binder also found a focus within left SMG ($-47, -36, 24$) which activated preferentially to tones compared with words. Similarly, our data revealed non-speech-preferring

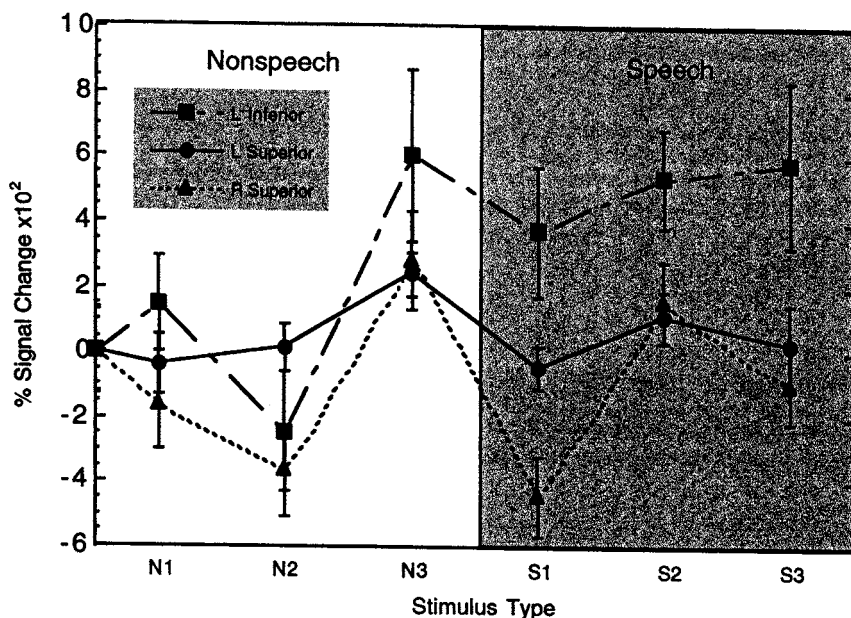


FIG. 9. Grand-averaged signal data averaged by condition (collapsing over the natural/synthetic difference) after reordering blocks which were pseudorandomized across the 12 subjects. Graph of data from three distinct ROIs in the supramarginal gyrus (SMG) which obtained from cluster analysis of the statistical maps. L = left; R = right. Error bars indicate standard error of the mean. Percentage of signal change was given relative to resting baseline. N1 = notes and harmonic complexes; N2 = chords, N3 = chord progressions; and S1 = V, S2 = CV, and S3 = CVC. Note the response of left inferior SMG to both the chord progressions and speech and the similarity of response of the left and right superior SMG ROIs.

speech or nonspeech. Increasing nonspeech complexity was associated, however, with significant increases in bilateral Heschl's gyrus and anterior STG activity but not posterior temporal regions. Conversely, increasing syllable complexity was associated with increased activation of left posterior STG (BA 22) but no increase in Heschl's gyrus, anterior STG. It is conceivable that the lack of an increment in signal in Heschl's gyrus [which contains primary auditory cortex (AI)] with increasing syllable complexity was due to similar spectral content across the three syllable types or to saturation of auditory cortex activity by the vowel syllables. The data, however, do not support this, since the chord progressions activated AI more than any of the syllable types (see Fig. 2), proving that AI was *not* saturated by the syllables (and scanner noise). Furthermore, sensitivity to frequency modulation differences across the syllable types (V-CV-CVC) by AI might be expected to increase activity in Heschl's gyrus, as has been observed in cats (Mendelson & Cynader, 1985). These results suggest that the increased phonetic processing demands of the more complex syllable types do not entail additional "low level" auditory cortical processing (in and immediately surrounding AI). In addition, three other regions, i.e., cerebellar vermis, dorsomedial thalamus, and left ventral occipitotemporal cortex demonstrated *opposite* responses to speech and nonspeech complexity. That each of these regions should respond to both phonetic and acoustic complexity, albeit in opposite directions, on the surface, suggests that these regions are not engaged in exclusively speech-specific processing. One possibility is that the activity of these regions is subject to ongoing top-down influence from "higher" speech and nonspeech cortical regions, explaining the dual response character. Such modulation of nonprimary cortical areas has been observed in a cross-modal attentional study using a visualauditory target detection task (Woodruff et al., 1996).

Supported by the above findings and the poor correlation between speech and non-speech psychophysical data in behavior studies (Whalen & Liberman, 1987; Xu et al., 1997), we would argue that the increase in phonetic processing was achieved largely independently of lower level auditory processing. These findings do not support auditory theories of speech perception (Darwin, 1984; Diehl & Kluender, 1989; Kingston, 1992; Kluender & Greenberg, 1989; Stevens & Blumstein, 1978) in which speech signals receive a full acoustic analysis before the resulting auditory percept is passed on to the speech system. The data are more compatible with a model in which speech signals are treated differently from the earliest stage in cortical processing. Proponents of the Motor Theory of speech perception (Liberman & Mattingly, 1985) and other gestural theories, such as Fowler's Direct Perception theory and Browman and Goldstein's Articulatory Phonology (1992), hold that the acoustic signal is interpreted in the context of its possible origin as a gesture of the vocal tract and is processed as such by a speech-specific module. While evidence for the gestural component of the theory is outside the scope of the present study, early separation of speech and nonspeech auditory processing is strongly supported.

Natural vs Synthetic Speech

Speech perception research has long relied on synthetic speech to investigate the effects of manipulating a single acoustic variable (Cooper, Liberman, & Borst, 1951). The goal of this approach is to identify the cues used in perception of speech. If natural and synthetic speech each receive a full acoustic analysis before being processed as speech, we might expect to see differences in auditory processing reflecting the acoustically richer natural speech (e.g., greater number of formants and greater energy at high frequencies). Conversely, if the signals are treated as speech "gestures" from the earliest stage, we might expect to see very little difference in auditory processing except as a consequence of paraphonetic, i.e., timber and voice, attributes.

In the current study, a significant dissociation between natural and synthetic speech was observed only in the left dorsolateral precentral sulcus, being more active for natural than synthetic speech. The left precentral sulcus focus may be related to differences in the degree to which mnemonic processes are engaged by the intrinsically more familiar natural speech (Wagner et al., 1998). No such dissociation occurred in the temporal lobe suggesting similar acoustic treatment; however, there was a trend in the right secondary auditory cortex (AII), which was more active for the natural than synthetic stimuli. The left precentral and right AII regions could be involved in paraphonetic attributes of the natural speech not contained in the synthetic speech. The right AII activation may be related to the vocal quality of the natural speech, as Belin et al. found a right STS focus which preferred vocal sounds over matched control sounds (Belin et al., 2000). Ross (1993) has found evidence of the right hemisphere's contribution to paralinguistic aspects of speech, e.g., prosody, through studies of patients with nondominant hemisphere lesions.

More generally, the differential activation between natural and synthetic speech is consistent with behavioral results of Gow et al. (1996), who found poor correlations by subject and item for natural and synthetic speech stimuli on identification and discrimination tasks. These data, together with our imaging results, reveal some of the differences between natural and synthetic speech which need to be accounted for in future investigations of speech recognition mechanisms.

Speech and Nonspeech Preference by Region

We evaluated each ROI's speech and nonspeech selectivity by observing the response across a set of six relevant contrasts. We assumed that an ROI which (1)

activated to speech but not nonspeech, (2) followed the speech complexity dimension but not the nonspeech complexity dimension, and (3) showed a significant interaction between speech and nonspeech complexity was highly speech selective. The converse criteria operationally defined nonspeech selectivity.

Left posterior superior temporal gyrus. A portion of the left pSTG demonstrated the highest degree of speech selectivity of the 12 ROIs evaluated (see Fig. 9). Given the range of empirical evidence supporting the left pSTG's involvement in linguistic processing (Marie & Foix, 1917; Nielsen, 1946; Penfield & Roberts, 1959; Wernicke, 1874), and phonological processing more specifically (Belin et al., 1998; Binder et al., 1996; Blumstein, Baker et al., 1977; Demonet et al., 1992; Mazoyer et al., 1993; Miceli, Gainotti, Caltagirone, & Masullo, 1980; Ojemann, 1990; Petersen, Fox, Posner, Mintun, & Raichle, 1988; Wise et al., 1991; Zatorre et al., 1996), this region's association with phonetic perception is not surprising. Whether the sensitivity of this region to phonetic structure is based on speech-specific cues or to complex auditory cues uniquely embodied by the syllable progression, as mentioned, cannot readily be determined from this first experiment, since the speech and nonspeech progressions differed acoustically, with the speech stimuli alone containing rapidly modulated spectra. Therefore, one possibility is that the apparent preference for speech-specific information of the left pSTG could be a preference for acoustic features, i.e., brief and continuously changing spectra. Some psychophysical (Kitterle, 1991) and unilateral lesion data (Blumstein, 1984; Robin, Tranel, & Damasio, 1990) suggest that the left hemisphere may be superior to the right at resolving temporal events, e.g., rapidly modulated spectra, while the right is superior at resolving spectral events (Zatorre, 1997). If this is true, then one would expect that the left pSTG would activate more for the speech which were more frequency modulated over time.

Right posterior superior temporal gyrus A homologous but smaller region of the right pSTG tracked the speech complexity dimension and was comparable in speech specificity to the left pSTG. In fact, only the left pSTG was more speech specific than the right pSTG. While this was a reliable finding in this group of strongly right-handed subjects, the specific contributions of the right pSTG, or the right hemisphere more generally, to speech perception is less certain.

One way that the linguistic capacities of the right hemisphere have been investigated is through commissurotomy (split-brain) patients. While results have varied across these patients, some right-handed patients have demonstrated that the isolated right hemisphere is capable of comprehending syntactically simple speech (Gazzaniga, 1995), and even when comprehension is poor, profound phonetic deficits are not the rule (Zaidel, 1985). These observations seem to conflict with the frequent occurrence of phonetic deficits with left hemisphere lesions, since the right hemisphere retains its ability. A recent case report by Boatman et al. (1998) highlights this apparent paradox: Intracarotid sodium amobarbital testing on a patient with refractory epilepsy revealed strongly left lateralized language function but no impairment of syllable discrimination with left carotid injection, suggesting speech perception ability in the right hemisphere. Electrical stimulation mapping of the left pSTG, however, did impair syllable discrimination. Boatman et al. reasoned that the left pSTG may be the preferred path of speech perception even when lesioned. Only when the entire hemisphere is nonfunctional, as with amobarbital anesthesia, is the right hemisphere "disinhibited." These findings are consistent with the concept of "metacontrol," first put forth by Levy and Trevarthen in 1976, which suggested that ability differences between the two hemispheres are not the sole determinants of dominance.

Prior functional imaging of auditory speech has also demonstrated bilateral activity in pSTG. Hickok et al. recently reviewed ERP, MEG, PET, and fMRI studies of auditory speech perception and concluded that bilateral pSTG activation was the most

consistent finding across studies (Hickok & Poeppel, 2000). Another recent study (Vouloumanos, Kiehl, Werker, & Liddle, 2000) compared the activation pattern to monosyllabic nonsense-word speech with that for three-formant sinewave analogs of the nonsense-word tokens using event-related fMRI in an oddball task and found bilateral temporal activation greater for the natural speech condition. Additionally, attempts to use fMRI to determine hemisphere language dominance have generally found auditory language tasks to be less lateralizing than analogous reading tasks because of bilateral temporal lobe activations (Constable, 2000; FitzGerald et al., 1997), suggesting that the processes underlying speech perception are more bilaterally represented than those of reading. Whether the right pSTG is performing different or redundant processing relative to the left pSTG, or is being driven passively through transcallosal connections with the left, is not clear but should be empirically testable by varying stimulus and task parameters.

Left ventral occipitotemporal cortex. Left ventral occipitotemporal cortex demonstrated a strong coupling with phonetic structure (preferring consonantal syllables) but a seemingly paradoxical response to nonspeech: activation of nonspeech over baseline but a significant negative correlation with complexity. The pattern of response to the six contrasts was very similar to that of the right anterior insula and opposite those of left dorsal thalamus and midline cerebellum. These findings suggest a complex role in acoustic and phonetic processing.

The left ventral occipitotemporal cortex (BA 37=basal posterior temporal) has increasingly been shown to have linguistic function (Nielsen, 1946; Rumsey, 1997). Lüders demonstrated prominent naming deficits at low-intensity electrical stimulation of this region and global aphasia with higher intensities (Lüders et al., 1991). Lesions in this area have been demonstrated to produce aphasia acutely but with no lasting deficit (Lüders et al., 1991; Penfield & Roberts, 1959). These results from Lüders suggest a complex but noncritical role for this region in language processing. Recently, functional neuroimaging studies suggested a role in phonological processing for this region (Rumsey, 1997; Shaywitz et al., submitted). Pugh and colleagues, based on work in dyslexics and normal readers (Shaywitz et al., submitted) believe this ventral temporal region plays a vital role in the fast transcoding between orthography and phonology and may function inefficiently in dyslexics. Other imaging studies have implicated this region in semantic processing, e.g., object naming (Moore & Price, 1999), concrete (Kiehl et al., 1999), and animate (Perani et al., 1999) word processing. Others have concluded that this region is amodal (Buchel, Price, & Friston, 1998; Damasio & Damasio, 1994), possibly concerned with linking converging inputs from multiple sensory modalities (Buchel, Price et al., 1998). More work exploring the effects of within- and across-modal manipulations and functional connectivity with other areas should help to elucidate the role of this region in language.

Left middle frontal/premotor cortex. This region activated to natural speech tokens but not synthetic speech or nonspeech. Fuster discusses the fundamental biological importance of prefrontal cortex in cognitive experience and the concept of an "action-perception cycle," which is required for learning and appropriate behaving (Fuster, 1989). Natural speech may be more efficiently perceived as a product of a human vocal tract, and on this basis may more effectively activate an action-perception "loop" in a listener. Consistent with this notion, neurons in this region, which preferred natural speech, may be related to Rizzolatti and Arbib's "mirror neurons," found in primate premotor cortex to activate to *both* the perception and execution of a specific action (Rizzolatti & Arbib, 1999). Recently, McGuire, in a PET study of intrinsic speech, found that imagined listening to a speaker activated left premotor cortex, while imagining one's own speech activated left inferior frontal cortex, suggesting a role for the premotor cortex in the perception of extrinsic speech (McGuire et al., 1996).

The difference between natural and synthetic speech may also be related to the adequacy of the representation that each engenders in working memory. Mnemonic function has been previously ascribed to this region, including monitoring and verification of information within working memory (Petrides, 1994) and encoding during memory retrieval (Fletcher et al., 1995). Similar to the action-perception model, one can speculate that the more familiar, natural stimulus is a more effective substrate of verbal working memory. Clearly, further investigation aimed at sorting out this area's linguistic, mnemonic, and perceptuomotor properties is needed to more fully understand its role.

Left inferior frontal cortex (Broca's area) The left inferior frontal cortex had a marginal preference for speech, indicated by trend toward greater activation for speech compared with nonspeech and slight tracking of the speech complexity dimension which did not attain significance. On the other hand, no difference was seen between the most complex nonspeech condition and the consonant-containing syllables. These results are similar to Fiez et al. (1995), who found Broca's activation for CV syllables and tone triplets and much less for vowels in isolation. They concluded that Broca's activation was in response to the rapid acoustic temporal modulations intrinsic to specific stimuli. More recent literature, however, supports there being no modal specificity to Broca's area (Klingberg, Kawashima, & Roland, 1996). Furthermore, that Broca's area should activate for a covert, nonlanguage condition is no longer an unexpected finding. Several studies have now reported Broca's area activation for nonlinguistic conditions (Fiez et al., 1995; Fox, Petersen, Posner, & Raichle, 1988; Grafton, Arbib, Fadiga, & Rizzolatti, 1996; Ingvar, 1983; Ojemann, 1983; Seitz & Roland, 1992; Tallal & Newcombe, 1978), including imaged grasping (Rizzolatti & Arbib, 1999), direction discrimination (Cornette et al., 1998), orientation discrimination (Taira, Kawashima, Inoue, & Fukuda, 1998), and multimodal short-term memory (Klingberg et al., 1996). What these studies have in common is not clear but may be related to processing of hierarchically organized information (Greenfield, 1991; Grossman, 1980; Wilkins & Wakefield, 1995) regardless of the sensory modality. Alternatively, one cannot rule out functional subdivisions within Broca's area, including areas specialized for aspects of syntactic (Stromswold, Caplan, Alpert, & Rauch, 1996), semantic, and phonological processing (Poldrack et al., 1999).

Supramarginal gyrus. The supramarginal gyrus (BA 40) appears to be comprised of two functionally different regions. The bilateral anterior/superior SMG exhibited an overall preference for nonspeech over speech, being driven by the more acoustically complex chord progressions. The right superior SMG, despite an overall preference for nonspeech, did show sensitivity to phonetic complexity, suggesting sensitivity to acoustic features embodied in the consonantal syllables. In contrast, left inferior SMG exhibited a preference for speech, despite not tracking either complexity dimension. Differential functioning of anterior/superior and posterior/inferior subregions of the SMG is supported by multiple lines of evidence from lesion and functional imaging data. Varney & Damasio (Varney and Damasio, 1986) in a CT study of aphasics found that strokes involving the left SMG were just as likely as not to be associated with impairment in sound recognition. Finer spatial analysis within the SMG was not reported. These results imply separability of auditory speech and nonspeech perception within supramarginal gyrus. Caplan, Gow, and Makris (1995) used regression and correlation analyses to relate acoustic/phonetic processing in aphasics to lesion location and extent on MRI and found the left posterior SMG to be the best predictor of phonetic processing deficits. Using MEG, (Nishitani, Nagamine, and Shibasaki (1998) found modality-specific subregions within the SMG, with the auditory M400 13.6 mm anterior and superior to the visual M400 ($Z = 28$ mm vs 20 mm; $Y = -32$ mm vs -40 mm) on a target detection oddball task. For the left SMG foci

in our study, the auditory nonspeech-preferring region of SMG was 11.3 mm more superior and anterior to the speech-preferring region ($Z = 27$ vs 16 mm; $Y = -23$ mm vs -31 mm). Rauschecker has found evidence in primates of ventral and dorsal processing streams analogous with the “what” and “where” pathways in the visual system. The ventral stream, involving posterior superior temporal cortex, is involved in auditory pattern analysis and is specialized for species-specific communication, while the dorsal stream, involving inferior parietal cortex, subserves auditory spatial processing (Rauschecker, 1998). An alternative model proposed by Hickok and Poeppel states that a ventral stream composed of posterior/superior temporal lobe is “the primary substrate for constructing sound-based representations of speech,” whereas a dorsal stream composed of inferior parietal lobe, situated between postsuperior temporal lobe and inferior frontal lobe, “reflects the operation of an auditory–motor integration network for speech” (Hickok & Poeppel, 2000). While our study did not specifically address these proposed dichotomies, our results and the other referenced studies appear to converge on dissociation of communicative and noncommunicative auditory processing across the temporal/parietal cortices.

Anterior superior temporal gyrus The bilateral aSTG showed a definite preference for nonspeech auditory stimuli, specifically the temporally modulated, acoustically more complex chord progressions. This region was contiguous with the anterior portion of Heschl’s gyrus and tracked the nonspeech complexity dimension, which was driven by a large response to chord progressions. This region would appear to be a relatively “pure” auditory processing region and may be homologous with the monkey rostral field of the core area or a field within the belt region surrounding the core (Kaas, Hackett, & Tramo, 1999). To our knowledge this is the first report describing such a region anterior to Heschl’s gyrus in the human, although Wessinger et al. did find expansion of activation circumferentially around the core as the stimulus changed from pure tones to band-passed noise in an fMRI experiment (Wessinger, Bharucha, Justus, & VanMeter, 2000). Further work exploring the response properties to modulations of acoustic spectra is needed to further characterize this bilateral region.

Subcortical regions. Left dorsomedial thalamus and midline cerebellum showed a similar pattern of activation across the contrasts, while the right anterior insula showed nearly the mirror opposite pattern. Left dorsomedial thalamus and midline cerebellum both showed deactivation to auditory stimuli compared with rest, and both regions correlated with the complexity dimensions, positively for nonspeech and negatively for speech. This pattern of results suggests that these regions are tonically active at rest and are actively inhibited during auditory stimulation. A likely candidate for inhibition of the dorsomedial thalamus is prefrontal cortex with which it has rich reciprocal connections (Akert & Hartmann-von Monakow, 1980) as well as indirect connections through other subcortical nuclei (Montaron, Deniau, Menetrey, Glowinski, & Thierry, 1996). Harder to explain, however, is the positive correlation with nonspeech complexity and the negative correlation with speech complexity. These regions are not generally considered auditory or speech processing regions, although a speech comprehension deficit may result from a left thalamic lesion (Graff-Radford, Eslinger, Damasio, & Yamada, 1984). One possibility is that these regions are part of an auditory processing module and are inhibited by higher centers during speech input. Another possibility is that these regions are not domain specific but have domain neutral functions, e.g., mnemonic or temporal processing, which are differentially deployed by higher centers.

Mirroring the dorsomedial thalamus and midline cerebellum, the right anterior insula activated to both speech and nonspeech above baseline, correlated with speech complexity and was anticorrelated with nonspeech complexity. The function of the

insular cortex is not well understood but has been related to visceral functions such as cardiovascular, gustatory and nociceptive perception, and mood. Recently, two PET studies have found the right anterior insula to activate to swallowing (Hamdy et al., 1999; Zald & Pardo, 1999), while another found the anterior insula to activate with articulation (Wise, Greene, Buchel, & Scott, 1999). Given the rightward laterality of the anterior insular activation observed in our study in association with increasing phonetic complexity, one could attribute activation of the anterior insula to recruitment of common articulatory/motor neurons serving both speech perception and production, consistent with the Motor Theory of speech perception. This preliminary finding requires further confirmation.

SUMMARY AND CONCLUSIONS

The main objective of this study was to identify candidate nodes in the neural network subserving phonetic processing in normal subjects. Using a passive listening task, we successfully identified seven spatially distributed brain regions which were strongly associated with speech processing, indicated by a strong preference for speech stimuli or correlation with phonetic complexity. These regions were located in the left posterior STG (extending to left angular gyrus), right posterior STG, left ventral occipitotemporal cortex (BAs 21 and 37), left inferior SMG (BA 40), left middle frontal gyrus (BA 6/9), right superior SMG, and right anterior insula. Of these, pSTG, or classical Wernicke's area, was most selective for speech, activating selectively to speech and phonetic complexity. No other region was as speech selective, but ventral occipitotemporal and right insula were strongly associated with the speech complexity dimension, indicating a speech or phonetic processing function for these regions. Only the left middle frontal gyrus differentiated natural and synthetic speech, suggesting a mimetic or mnemonic role.

Regarding brain regions implicated previously in phonetic/phonological perception, the data strongly suggests that the dominant pSTG, including angular gyrus, is engaged in the decoding of phonetic/phonological information. In contrast, inferior frontal gyrus, or Broca's area, was equally active for both speech and complex non-speech, suggesting a nonlinguistic role of this region. The SMG was found to be composed of a left posterior/inferior speech-preferring subregion and a bilateral anterior/superior auditory nonspeech preferring subregion. On the other hand, the *right* superior SMG responded to both acoustic and phonetic features, lending support to previous observations of hemispheric differences in the perception of specific phonetic features, e.g., place of articulation vs voicing. Further acoustic/phonetic analysis of the seven distributed nodes comprising the core phonetic network defined in this study, in addition to describing the individual and collective behavior of these nodes (e.g., functional connectivity assessment), are the necessary steps to construct an accurate neurolinguistic model of speech perception. Defining the functional network subserving phonetic perception as well as interindividual variability should be useful as a benchmark against which many developmental and acquired language disorders can be analyzed and neuroplasticity evaluated.

One important question addressed by this study, which has been a subject of longstanding debate, is whether speech receives a full acoustic analysis before it is processed as speech, i.e., a serial model, or whether speech is processed differently at a very early stage, i.e., a parallel model. The results presented in this article, while not conclusive, suggest early separation of speech and nonspeech auditory processing. In order to more directly address this question, we are currently using acoustically matched speech and nonspeech stimuli in addition to sparse temporal sampling to deconvolute the hemodynamic responses to scanner and stimuli.

APPENDIX

Natural Speech Stimuli

Natural Vs	Natural CVs	Natural CVCs
[a]	[ta]	[tag]
[e ^j]	[ze ^j]	[ze ^k k]
[i]	[gi ^j]	[gi ^j p]
[o ^w]	[vo ^w]	[vo ^w s]
[u ^w]	[ðu ^w]	[ðu ^w t]

Synthetic Speech Stimuli

The synthetic speech stimuli were based on the vowel formant frequencies in the natural speech. The values for the formants and the durations of each stimulus are given below.

Synthetic Vs

Duration (same across stimuli) = 600 ms

F0 (normalized for all Vs) = 120 Hz (until 500 ms for [a] but 400 for others), 134 Hz for [o^w, u] with drop to 85 Hz (at 600 ms for [a]), 90 Hz for [e^j, o^w], 100 Hz for [i], 95 Hz for [u]

Input amplitude of voicing (normalized across stimuli during vowel) = 56 dB (for [a]), 52 dB (for others)

(starts to drop to 0 dB starting at 576 ms for [a], 572 for [e^j], 567 for [o^w], and 564 for [i])

(ramps up from 28 dB at 29 ms for [i], from 35 dB at 37 ms for [u])

	F1	F2	F3
[a]	845	1340	2775
[e ^j]	509–496(336 ms)–375(600 ms)	2239(0–320 ms)–2440(600 ms)	2614
[i]	214	2828	2989
[o ^w]	496 (362 ms–362(600 ms)	979–952(379 ms)–724(600 ms)	2614
[u]	241	697(0–297 ms)–630(569 ms)	2239

Synthetic CVs

Duration (Fixed) = 600 ms

F0 (should be normalized in part) = 120 Hz with drop at 400 ms to 90.4 Hz [ta]
 93.8 Hz (0 ms)–125 (94–400 ms)–90 (600 ms) [ze^j]
 107.2 (0 ms)–134(49–400 ms)–100 (600 ms) [gi^j]
 125 (0–400 ms)–85.4 (600 ms) [vo^w]
 134 (0–400 ms)–93.8 (600 ms) [ðu^w]

	F1	F2	F3
[ta]	791 (100 ms)–804 (367 ms)	1300 (100)–1287(367)	2641(100)–2654(367 ms)
[ze ^j]	228–470 (176–470)–335	1997–1984(184)–2440(534)	3150–2708(184)–2694(538)

- Basso, A., Casati, G., & Vignolo, L. A. (1977). Phonemic identification defect in aphasia. *Cortex*, **13**, 84–95.
- Belin, P., Zatorre, R., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, **403**(6767), 309–312.
- Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience*, **10**(4), 536–540.
- Benson, D. F., Sheremata, W. A., Bouchard, R., Segarra, J. M., Price, D., & Geschwind, N. (1973). Conduction aphasia: a clinicopathological study. *Archives of Neurology*, **28**, 339–346.
- Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CVs. *Journal of the Acoustical Society of America*, **82**(1), 31–37.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain*, **119**, 1239–1247.
- Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., Wong, E. C., Estkowski, L. D., Goldstein, M. D., Haughton, V. M., & Hyde, J. S. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology*, **35**, 662–672.
- Blumstein, S. E. (1984). Acoustic cues for the perception of place of articulation in aphasia. *Brain and Language*, **22**, 128–149.
- Blumstein, S. E., Baker, E., & Goodglass, H. (1977). Phonological factors in auditory comprehension in aphasia. *Neuropsychologia*, **15**, 19–30.
- Blumstein, S. E., Cooper, W. E., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset time in aphasia. *Neuropsychologia*, **15**, 371–383.
- Boatman, D., Hart, J. J., Lesser, R. P., Honeycutt, N., Anderson, N. B., Miglioretti, D., & Gordon, B. (1998). Right hemisphere speech perception revealed by amobarbital injection and electrical interference. *Neurology*, **51**(2), 458–464.
- Brown, C. P., & Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica*, **49**(3/4), 155–180.
- Buchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus–response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage*, **8**(2), 140–148.
- Buchel, C., Price, C., & Friston, K. (1998). A multimodal language region in the ventral visual pathway. *Nature*, **394**(6690), 274–277.
- Buchman, A. S., Garron, D. C., Trost-Cardamone, J. E., Wichter, M. D., & Schwartz, M. (1986). Word deafness: one hundred years later. *Journal of Neurology, Neurosurgery, and Psychiatry*, **49**, 489–499.
- Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic–phonetic processing deficits. *Neurology*, **45**, 293–298.
- Cappa, S., Cavallotti, G., & Vignolo, L. A. (1981). Phonemic and lexical errors in fluent aphasia: correlation with lesion site. *Neuropsychologia*, **19**, 171–177.
- Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., & Chollet, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *NeuroImage*, **9**, 135–144.
- Chertkow, H., & Bub, D. (Eds.). (1994). *Functional activation and cognition: The O¹⁵ subtraction method*. San Diego: Academic Press.
- Constable, R. T. (2000). Auditory tasks are less lateralized than visual language tasks.
- Cooper, F. S., Liberman, A. M., & Borst, J. M. (1951). The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proceedings of the National Academy of Science*, **37**, 318–325.
- Cornette, L., Dupont, P., Rosier, A., Sunaert, S., Van Hecke, P., Michiels, J., Mortelmans, L., & Orban, G. (1998). Human brain regions involved in direction discrimination. *Journal of Neurophysiology*, **79**(5), 2749–2765.
- Damasio, A. R., & Damasio, H. (Eds.). (1994). *Large scale neuronal theories of the brain*. Cambridge, MA: MIT Press.
- Darwin, C. J. (Ed.). (1984). *Auditory processing and speech perception*.
- Demonet, J., Chollet, R., Ramsay, S., Cardebat, D., Nespoulous, J., Wise, R., Rascol, A., & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain*, **115**, 1753–1768.

- Demonet, J. F., Fiez, J. A., Paulesu, E., Petersen, S. E., & Zatorre, R. J. (1996). PET studies of phonological processing: A critical reply to Poeppel. *Brain and Language*, *55*, 352–379.
- Demonet, J. F., Price, C., Wise, R., & Frackowiak, R. S. (1994). A PET study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain*, *117*, 671–682.
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*, 121–144.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*(968), 303–306.
- Fiez, J. A., Raichle, M. E., Miezin, F. M., & Petersen, S. E. (1995). PET studies of auditory and phonological processing: Effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience*, *7*(3), 357–375.
- FitzGerald, D. B., Cosgrove, G. R., Ronner, S., Jiang, H., Buchbinder, B. R., Belliveau, J. W., Rosen, B. R., & Benson, R. R. (1997). Language localization in cortex: A comparison between fMRI and electrocortical stimulation. *AJNR: American Journal of Neurology*, *18*, 1529–1539.
- Fletcher, P. C., Frith, C. D., Grasby, P. M., Shallice, T., Frackowiak, R. S., & Dolan, R. J. (1995). Brain systems for encoding and retrieval of auditory-verbal memory. An *in vivo* study in humans. *Brain*, *118*, 401–416.
- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.
- Fowler, C. A. (1994). Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation. *Perception and Psychophysics*, *55*, 597–610.
- Fowler, C. A., & Smith, M. (Eds.). (1986). *Speech perception as "vector analysis": An approach to the problems of segmentation and invariance*. Hillsdale, NJ: Erlbaum.
- Fox, P. T., Petersen, S., Posner, M., & Raichle, M. E. (1988). Is Broca's area language-specific? *Neurology*, *38*, 172.
- Friston, K. J., Ashburner, J., Poline, J. B., Frith, C. D., Heather, J. D., & Frackowiak, R. S. J. (1995). Spatial registration and normalisation of images. *Human Brain Mapping*, *2*, 165–189.
- Friston, K. J., Williams, S. R., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*, 346–355.
- Fuster, J. M. (1989). *The prefrontal cortex: anatomy, physiology, and neuropsychology of the frontal lobe* (2nd ed.). New York: Raven Press.
- Gazzaniga, M. S. (1995). Principles of human brain organization derived from split-brain patients. *Neuron*, *14*, 217–228.
- Gow, D. W., & Caplan, D. (1996). An examination of impaired acoustic-phonetic processing in aphasia. *Brain and Language*, *52*, 386–407.
- Graff-Radford, N. R., Eslinger, P. J., Damasio, A. R., & Yamada, T. (1984). Nonhemorrhagic infarction of the thalamus: Behavioral, anatomical and physiological correlates. *Neurology*, *34*, 14–23.
- Grafton, S. T., Arbib, M. A., Fadiga, L., & Rizzolatti, G. (1996). Localization of grasp representations in humans by positron emission tomography. *Experimental Brain Research*, *112*(1), 103–111.
- Greenfield, P. M. (1991). Language, tools, and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *The Behavioral and Brain Sciences*, *14*, 531–595.
- Grossman, M. (1980). A central processor for hierarchically structured material: Evidence from Broca's aphasia. *Neuropsychologia*, *18*, 299–308.
- Hamdy, S., Rothwell, J. C., Brooks, D. J., Bailey, D., Aziz, Q., & Thompson, D. G. (1999). Identification of the cerebral loci processing human swallowing with H2(15)O PET activation. *Journal of Neurophysiology*, *81*(4), 1917–1926.
- Haxby, J. V., Maisog, J. M., & Courtney, S. C. (2001). Multiple regression of effects of interest in fMRI time series. In P. Fox, J. Lancaster, & K. Friston (Eds.), *Mapping and modeling the human brain*. New York: Wiley.
- Hickok, G., & Poeppel, G. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Neurosciences*, *4*(4), 131–138.
- Ingvar, D. H. (1983). Serial aspects of language and speech related to prefrontal cortical activity: A selective review. *Human Neurobiology*, *2*(3), 177–189.
- Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology*, *9*, 164–170.
- Kiehl, K. A., Liddle, P. F., Smith, A. M., Mendrek, A., Forster, B. B., & Hare, R. D. (1999). Neural

- Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2000). *Neurological bases of speech and non-speech processing: An event-related functional magnetic resonance imaging study*. Paper presented at the Cognitive Neuroscience Society Annual Meeting, San Francisco.
- Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, **281**, 1188–1191.
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, **75**(6), 1866–1878.
- Wernicke, C. (1874). *Der aphasische symptomcomplex*. Breslau: Max Cohn and Weigert.
- Wernicke, C. (1874/1977). Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis. In G. H. Eggert (Ed.), *Wernicke's works on aphasia: A sourcebook and review* (pp. 91–145). New York: Mouton.
- Wessinger, C. M., Bharucha, J., Justus, T., & VanMeter, J. W. (2000). *Differential activation of auditory and visual cortices with auditory objects*. Paper presented at the Society for Neuroscience, 30th Annual Meeting, New Orleans.
- Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, **237**, 169–171.
- Wilkins, W. K., & Wakefield, J. (1995). Brain evolution and neurolinguistic preconditions. *Behavioral and Brain Sciences*, **18**(1), 161–226.
- 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.
- Wise, R. J., Greene, J., Buchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *Lancet*, **353**(9158), 1056–1061.
- Woodruff, P. W., Benson, R. R., Bandettini, P. A., Kwong, K. K., Howard, R. J., Talavage, T., Belliveau, J. W., & Rosen, B. R. (1996). Modulation of auditory and visual cortex by selective attention is modality-dependent. *NeuroReport*, **7**(12), 1909–1913.
- Xu, Y., Liberman, A. M., & Whalen, D. H. (1997). On the immediacy of phonetic perception. *Psychological Science*, **8**(5), 358–362.
- Zaidel, E. (1985). *Language in the right hemisphere*. New York: Guilford Press.
- Zald, D. H., & Pardo, J. V. (1999). The functional neuroanatomy of voluntary swallowing. *Annals of Neurology*, **46**(3), 281–286.
- Zatorre, R. J. (1997). Cerebral correlates of human auditory processing: Perception of speech and musical sounds. In J. Syka (Ed.), *Acoustical signal processing in the central auditory system* (pp. 453–468). New York: Plenum Press.
- Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, **256**, 846–849.
- Zatorre, R. J., Meyer, E., Gjedde, A., & Evans, A. C. (1996). PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cerebral Cortex*, **6**(1), 21–30.