

Neural Correlates of Segmental and Tonal Information in Speech Perception

Jack Gandour,¹ Yisheng Xu,¹ Donald Wong,² Mario Dzemidzic,³
Mark Lowe,³ Xiaojian Li,¹ and Yunxia Tong¹

¹Department of Audiology and Speech Sciences, Purdue University, West Lafayette, Indiana

²Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, Indiana

³Department of Radiology, Indiana University School of Medicine, Indianapolis, Indiana

Abstract: The Chinese language provides an optimal window for investigating both segmental and suprasegmental units. The aim of this cross-linguistic fMRI study is to elucidate neural mechanisms involved in extraction of Chinese consonants, rhymes, and tones from syllable pairs that are distinguished by only one phonetic feature (minimal) vs. those that are distinguished by two or more phonetic features (non-minimal). Triplets of Chinese monosyllables were constructed for three tasks comparing consonants, rhymes, and tones. Each triplet consisted of two target syllables with an intervening distracter. Ten Chinese and English subjects were asked to selectively attend to targeted sub-syllabic components and make same-different judgments. Direct between-group comparisons in both minimal and non-minimal pairs reveal increased activation for the Chinese group in predominantly left-sided frontal, parietal, and temporal regions. Within-group comparisons of non-minimal and minimal pairs show that frontal and parietal activity varies for each sub-syllabic component. In the frontal lobe, the Chinese group shows bilateral activation of the anterior middle frontal gyrus (MFG) for rhymes and tones only. Within-group comparisons of consonants, rhymes, and tones show that rhymes induce greater activation in the left posterior MFG for the Chinese group when compared to consonants and tones in non-minimal pairs. These findings collectively support the notion of a widely distributed cortical network underlying different aspects of phonological processing. This neural network is sensitive to the phonological structure of a listener's native language. *Hum. Brain Mapp.* 20:185–200, 2003. © 2003 Wiley-Liss, Inc.

Key words: functional magnetic resonance imaging; fMRI; human auditory processing; language; speech perception; prosody; selective attention; phonology

INTRODUCTION

Recent functional neuroimaging studies of language have been directed to the neural substrates of phonology. The left dorsolateral posterior prefrontal cortex (PFC) has been implicated in the extraction of consonant information [Burton et al., 2000; Demonet et al., 1994; Zatorre et al., 1992, 1996]. It has also been shown that *suprasegmental* as well as segmental units activate a similar subregion of left posterior prefrontal cortex. This suprasegmental information includes pitch information underlying Thai tones [Gandour et al., 1998, 2000, 2002], Chinese tones and intonation [Gandour et al., 2003; Hsieh et al., 2001; Klein et al., 2001], and temporal information underlying Thai vowel length [Gandour et al., 2002a,b]. Collectively, these previous studies point to a functional

Contract grant sponsor: Institutes of Health; Contract grant number: R01DC04584-04; Contract grant sponsor: James S. McDonnell Foundation (to J.G.).

*Correspondence to: Dr. Jack Gandour, Purdue University, Department of Audiology and Speech Sciences, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038.

E-mail: gandour@purdue.edu

Received 19 March 2003; Accepted 27 August 2003

DOI 10.1002/hbm.10137

subregion within inferior PFC as a possible neural substrate for extraction of phonetic information.

Evidence for extraction of phonetic information may be elicited from paired speech stimuli that are distinguished by two or more differences in their segmental composition. Such pairs are called *non-minimal* pairs (e.g., *beep vs. doom*). In non-minimal pairs, subjects must separate the individual phonetic units from the whole syllable in order to make a same-different judgment. In contrast, if there is only one phonetic difference in the discrimination pair (*minimal* pairs), subjects do not necessarily need to tease apart the phonetic units or even be able to classify the paired stimuli into linguistic categories. In minimal pairs (e.g., *bag vs. gag*), a subject's judgment can be based on simple acoustic information. The optimal test for phonetic extraction requires that non-minimal pairs be compared directly to minimal pairs. This direct comparison has been carried out in only one previous study [Burton et al., 2000], which demonstrated that non-minimal pairs activate a posterior and superior subregion of left inferior PFC, whereas minimal pairs do not. They concluded that activation of this subregion reflects either overt segmentation of the phonetic units of the stimulus or, alternatively, verbal rehearsal in working memory for subparts of a stimulus [Burton et al., 2000, p. 687].

Evidence from the lesion literature remains inconclusive on a specific role of the left frontal lobe in phonological segmentation. Patients with impairments in acoustic-phonetic processing, i.e., in their ability to discriminate and identify phonemes, tend to have lesions in the left supra-marginal gyrus and the bordering parietal operculum [Caplan et al., 1995; Gow and Caplan, 1996]. Although the inferior parietal region plays an important role in acoustic-phonetic processing, "it is most likely not the only region involved in this function" [Caplan et al., 1995]. More recently, left-hemisphere-damaged nonfluent aphasics have been reported to exhibit impairments in their ability to discriminate consonant and vowel phonemes, no matter whether they are presented in minimal or non-minimal pairs [Baum, 2002]. Only one of the ten nonfluent aphasics, however, had a lesion circumscribed to the left frontal lobe.

Earlier functional neuroimaging studies of Chinese (Mandarin) tones have consistently revealed activation of the left posterior PFC [Gandour et al., 2003; Hsieh et al., 2001; Klein et al., 2001]. As a tone language, it provides an optimal window for investigating the neural circuitry underlying extraction of phonetic information. In addition to consonants and vowels, Chinese has four lexical tones: e.g., *ma* (1) "mother," *ma* (2) "hemp," *ma* (3) "horse," *ma* (4) "scold". None of these brain imaging studies, however, compared tonal processing in non-minimal relative to minimal pairs. Instead, only minimal pairs were used in the aforementioned studies of Chinese tones. In two studies of Thai tones, minimal pairs were used in one [Gandour et al., 2002], non-minimal pairs in another [Gandour et al., 2000].

Brain imaging evidence has yet to be brought to bear on sub-syllabic components and their relationship to one another. A topic of controversy about syllable structure in

Chinese is whether tones are associated with syllable-internal units, or alternatively, with the entire syllable itself. Phonological evidence has been adduced to support either the rhyme as the unit of reference [Bao, 1990] or the syllable [Wang, 1967]. From a production standpoint, acoustic data appears to favor the rhyme [Howie, 1976] or syllable [Xu, 1998]. Slips-of-the-tongue data, however, are incompatible with the syllable as the unit of reference, revealing a large disparity between tonal and segmental errors [Chen, 1999]. Perceptually, a direct comparison of tonal vs. segmental information in Cantonese syllables reveals that the processing of lexical tones may be slowed relative to the processing of onsets or rhymes [Cutler and Chen, 1997]. Their findings were replicated in Mandarin [Ye and Connine, 1999].

Accordingly, the specific aim of this cross-linguistic, functional magnetic resonance imaging (fMRI) study is to elucidate neural mechanisms involved in the discrimination of Chinese sub-syllabic units: consonants (i.e., consonants in syllable-initial position), rhymes, and tones. By using the same discrimination task with both minimal and non-minimal pairs, we are able to compare brain activation patterns associated with all three sub-syllabic units. By employing native and nonnative speakers of Chinese, we are able to determine the extent to which these phonetic processes are sensitive to language experience.

MATERIALS AND METHODS

Subjects

Ten adult Chinese (5 male; 5 female) and ten adult English (5 male; 5 female) speakers were closely matched in age (Chinese: $M = 28.4$, $SD = 2.3$; English: $M = 27.3$, $SD = 5.0$) and years of formal education (Chinese: $M = 19.1$, $SD = 3.1$; English: $M = 19.9$, $SD = 2.9$). All subjects were strongly right-handed as measured by the Edinburgh Handedness Inventory [Oldfield, 1971] and exhibited normal hearing sensitivity at frequencies of 0.5, 1, 2, and 4 kHz. All subjects gave informed consent in compliance with a protocol approved by the Institutional Review Board of Indiana University Purdue University Indianapolis and Clarian Health.

Stimuli

Triplets of monosyllabic, Chinese morphemes were constructed to allow for tasks of overt comparison of sub-syllabic units: consonant (C), rhyme (R), tone (T). In each triplet (i.e., trial), the two target syllables for comparison, which occurred first and last in the sequence, were either minimal or non-minimal pairs for Chinese consonants, rhymes, and tones (Table I). One stimulus type consisted of minimal pairs; the other consisted of non-minimal pairs. Three sets of triplets were made up of minimal pairs, one for each sub-syllabic unit (C_{minimal} , R_{minimal} , T_{minimal}). Another three sets were made up of non-minimal pairs ($C_{\text{non-minimal}}$, $R_{\text{non-minimal}}$, $T_{\text{non-minimal}}$). The intervening syllable differed from the first and last in all three units (consonant, rhyme, tone) to minimize phonological similarity effects [Baddeley,

TABLE I. Sample minimal and non-minimal pair triplets for segmental and suprasegmental units

Minimal pair	Response	Non-minimal pair	Response
Consonant			
chai ⁴ heng ² chai ⁴	Same	chai ⁴ heng ² cha ¹	Same
mang ¹ you ² fang ¹	Different	mang ¹ you ² ren ⁴	Different
Rhyme			
bo ² liu ⁴ bo ²	Same	bo ² liu ⁴ po ³	Same
rong ² lei ⁴ reng ²	Different	rong ² lei ⁴ deng ³	Different
Tone			
huo ² pian ⁴ huo ²	Same	huo ² pian ⁴ ke ²	Same
shao ² dong ¹ shao ⁴	Different	shao ² dong ¹ fang ⁴	Different

Superscript numbers represent the four lexical tones of Madarin Chinese
 Bold = target, subsyllabic unit.

1966]. An intervening, distracter syllable was included to minimize differences in working memory load for processing minimal and non-minimal pairs. The distracter acts as a suffix to reduce the sensory echo of the first target syllable. Its presence makes it less likely that subjects can employ a simple acoustic strategy to make their discrimination judgments. Instead, it leads subjects to direct their attention to phonetically-encoded aspects of the auditory signal. To eliminate potentially confounding effects of tone sandhi on perceptual judgments [Xu, 1991], no triplet contained any occurrences of two consecutive monosyllables that carried Tone 3.

Because bisyllabic words predominate in the Chinese lexicon [Chao, 1968], triplets were designed so that no two adjacent morphemes matched an existing lexical item. The use of monosyllables also eliminated word length effects. Chinese consonants, vowels, and tones occurred with about equal frequency in the target syllables in the first and last position as well as in the intervening syllables. In the target syllables, there were 432 occurrences each of consonants, rhymes, and tones; in the intervening syllables, 216 occurrences of each. In the target syllables, 21 of 21 Chinese consonant phonemes were represented; 32 of 37 rhymes; and all 4 lexical tones.

Thirty-six triplets were constructed for each speech task. Half of the triplets were comprised of minimal pairs for a designated, sub-syllabic unit (i.e., consonant, rhyme, tone), the other half non-minimal pairs. Speech stimuli for the consonant, rhyme, and tone tasks were identical phonotactically for both minimal and non-minimal pairs. Thirty-six corresponding triplets of non-speech stimuli, hums (H), were created (*Praat*) for the passive listening baseline tasks by resynthesizing the speech stimuli with six formants (bandwidths): 600 (50), 1,400 (100), 2,400 (200) 3,400 (300), 4,500 (400), and 5,500 (500) Hz. Half of the non-speech triplets that were used as a passive listening control for the consonant task was derived from minimal pairs from C_{minimal} triplets; the other half was derived from non-minimal pairs from $C_{\text{non-minimal}}$ triplets. Non-speech triplets corresponding to the rhyme and tone triplets were similarly derived. The non-speech stimuli effectively held spectral information con-

stant, while preserving duration and fundamental frequency properties of the speech stimuli. Moreover, regardless of the derivational source, non-speech hums were not significantly different from one another in terms of either mean duration [$F(2,231) = 0.00, P = 0.9993$] or voice fundamental frequency [$F(2,231) = 0.25, P = 0.7756$].

Tasks

In all six speech tasks—C, R, and T in minimal and non-minimal pairs—subjects were required to direct their attention to the *first* and *last* target syllables of each triplet, and make discrimination judgments of consonants, rhymes, or tones in those syllables only, ignoring the intervening, distracter syllable. Subjects were not informed of the two different stimulus types (minimal pairs vs. non-minimal pairs). Thus, they were unable to consciously differentiate their response strategies according to stimulus type. They were asked to press the left or right mouse button to indicate their same/different judgment during a 2-s response interval following each triplet. Two-thirds of the responses were “different,” one-third “same.”

In minimal pairs, target syllables for different responses differed by one phonetic feature, whereas same responses differed by zero features. In non-minimal pairs, target syllables for different responses differed by three phonetic features, same responses by two features. Pooling across minimal and non-minimal pairs, same responses were elicited from target syllables that differed by one less phonetic feature than from those on which different responses were based. By this design, the phonetic distance between same and different responses was held constant across minimal and non-minimal pairs. Neither minimal nor non-minimal pairs were judged more than once in any task, thus eliminating confounding effects of prior exposure.

The consonant, rhyme, and tone tasks were designed to elicit discrimination judgments about Chinese phonological units in a non-syntactic context. As such, they should permit us to identify brain areas involved in perceptual processing of segmental (consonant, rhyme) and suprasegmental (tone) information. The passive listening control task was designed to address cognitive processes inherent to automatic, non-directed perceptual analysis of non-speech, auditory signals whose duration, intensity, and F_0 properties closely match those of the Chinese speech stimuli. In these control tasks (H), subjects were asked to listen passively without directing their attention to anything in particular, and alternately press the left and right mouse button after each triplet.

Scanning runs consisted of two tasks presented in blocked format (32 s) alternating with 16-s rest periods. Each of six scanning runs contained eight blocks, four blocks per each of two tasks. One task required active, discrimination judgments of Chinese consonants, rhymes, or tones; the other task required only passive listening to non-speech stimuli. Each block was made up of nine trials. The order of scanning runs and trials within blocks were randomized for each subject. Instructions were delivered to subjects in their native language via headphones during rest periods immedi-

ately preceding each task: “listen” for passive listening to non-speech stimuli (H); “consonant” for same-different judgments on Chinese consonants (C_{minimal} , $C_{\text{non-minimal}}$); “rhyme” for same-different judgments on Chinese rhymes (R_{minimal} , $R_{\text{non-minimal}}$); and “tone” for same-different judgments on Chinese tones (T_{minimal} , $T_{\text{non-minimal}}$). Average trial duration was about 3.6 s, including two interstimulus intervals of 200 ms each and a response interval of 1,650 ms. Average syllable duration was 520 ms. Total duration of auditory stimulation (excluding the interstimulus and response intervals) during a 32-s block was about 14 s.

Accuracy, reaction time, and subjective ratings of task difficulty were collected to measure task performance. Before scanning, Chinese subjects were trained to a high level of accuracy on all tasks accuracy using stimuli different from those used during the scanning runs: consonant ($M = 95.9$, $SD = 1.2$); rhyme ($M = 98.2$, $SD = 0.9$); tone ($M = 97.7$, $SD = 1.2$). English subjects were similarly trained to reasonably high levels of accuracy on the consonant ($M = 83.3$, $SD = 1.9$), rhyme ($M = 91.4$, $SD = 1.7$), and tone ($M = 85.9$, $SD = 3.4$) tasks.

Imaging Protocol

Scanning was done on a 1.5T Signa GE LX Horizon scanner (Waukesha, WI) equipped with birdcage transmit-receive radiofrequency head coils. Each of three 200-volume echo-planar imaging (EPI) series was begun with a rest interval consisting of 8 baseline volumes (16 s), followed by 184 volumes during which the two comparison tasks (32 s) alternated with intervening 16-s rest intervals, and ended with a rest interval of 8 baseline volumes (16 s). Gradient-echo EPI images were acquired with the following parameters: repetition time/echo time (TR/TE) 2 s/50 ms; matrix 64×64 ; flip angle (FA) 90° ; 24×24 cm field of view (FOV); receiver bandwidth 125 kHz. Fifteen 7.5-mm thick axial slices without an interslice gap were prescribed to image the entire cerebrum.

Subjects were scanned with eyes closed and room lights dimmed. The effects of head motion were minimized by using a head-neck pad and dental bite bar. Each scan was analyzed for head motion [Woods et al., 1992]. Data with maximal peak-to-peak displacement greater than 0.15 mm were rejected because image registration algorithms do not completely remove motion effects [Jiang et al., 1995]. Data for two out of 22 subjects were excluded due to excessive head motion. Pooling across scanning runs, average peak-to-peak displacement was 0.01 mm for four subjects, 0.02 mm for ten subjects, 0.03 mm for four subjects, and 0.04, 0.05, and 0.07 mm for each of three subjects, respectively. All fMRI data were Hamming-filtered spatially, which increased the BOLD (blood oxygen level dependent) contrast-to-noise ratio with only a small loss of spatial resolution [Lowe and Sorenson, 1997].

Prior to functional imaging scans, high-resolution, anatomic images were acquired in 124 contiguous axial slices using a 3D Spoiled-Grass (3D SPGR) sequence (slice thickness 1.1–1.2 mm; TR/TE 35/8 ms; 1 excitation; 30° FA;

matrix 256×128 ; FOV 24×24 cm; receiver bandwidth 32 kHz) for purposes of anatomic localization and co-registration.

Imaging Analysis

A comparison of the tasks of interest was accomplished by constructing a reference function described below. Each fMRI scan consisted of blocks of three states: rest, state 1, and state 2. For example, one fMRI scan had a rest condition in 16s blocks, $C_{\text{non-minimal}}$ in 32-s blocks, and H in 32-s blocks. The reference function of interest (state 1-state 2) was defined as:

$$\begin{aligned} \text{reference} &= 1 \text{ i} = [12,27], [60,75], [108,123], [156,171] \\ &= -1 \text{ i} = [36,51], [84,99], [132,147], [180,195] \end{aligned}$$

where i is the index of the i th-acquired volume (first volume: $i = 1$). The first 11 images were discarded to account for pre-saturation effects (images 1–8) and subsequent hemodynamic delay effects (images 9–11; [Bandettini et al., 1993]). A least-squares method was used to calculate a Student’s t -statistic value for each task comparison by comparing the derived reference function to the acquired data [Lowe and Russell, 1999].

Individual whole brain statistical maps were interpolated to $256 \times 256 \times 256$ cubic voxels (0.9375 mm/side). Individual anatomic images and single-subject interpolated activation maps were projected into a standardized stereotaxic coordinate system [Talairach and Tournoux, 1988]. The Student’s t maps were then summed pixel-by-pixel to produce within-group activation maps. Significance levels for the resulting maps were calculated using the fact that, for the null hypothesis, the summed t -values are distributed approximately as $N(0, \sqrt{n})$, where $\langle n \rangle$ is the number of subjects’ maps summed. The thresholded group maps were then displayed on anatomic images from a representative subject. Stereotaxic location of activation peaks and extent of activation were identified by drawing regions of interest (ROIs) around activation foci at a Student’s t -statistic threshold (1-tailed, uncorrected) of $t(\infty) = 6.0$, $P < 9.6 \times 10^{-10}$ for within-group comparisons involving a passive listening control (e.g., T_{minimal} vs. H; $T_{\text{non-minimal}}$ vs. H).

Data were then analyzed using the double subtraction technique [Poldrack et al., 1999]. Within-group comparisons that we performed included: non-minimal vs. minimal pairs for each phonological unit (e.g., $T_{\text{non-minimal}}$ vs. T_{minimal}), and consonant vs. rhyme vs. tone for either non-minimal or minimal pairs (e.g., $T_{\text{non-minimal}}$ vs. $R_{\text{non-minimal}}$) ($t(\alpha) = 4.25$, $P < 1.1 \times 10^{-5}$, 1-tailed, uncorrected). Between-group comparisons were also carried out on each phonological unit for either non-minimal or minimal pairs (e.g., $(T_{\text{non-minimal}} \text{ vs. } H)_{\text{Chinese}} - (T_{\text{non-minimal}} \text{ vs. } H)_{\text{English}}$) ($t(\alpha) = 4.25$, $P < 1.1 \times 10^{-5}$, 1-tailed, uncorrected).

TABLE II. Behavioral data*

Language group	Task	Stimulus type	Accuracy (%)	Reaction time (ms)	Self-rating ^a
Chinese	Consonant	Minimal	93.9 (2.6)	316.4 (93.0)	2.4 (1.3)
		Non-minimal	88.9 (7.4)	429.2 (59.0)	
	Rhyme	Minimal	97.2 (2.6)	384.9 (88.9)	2.6 (1.0)
		Non-minimal	93.0 (4.8)	496.3 (105.6)	
	Tone	Minimal	99.4 (1.2)	350.4 (81.5)	1.5 (0.8)
		Non-minimal	96.7 (4.3)	487.7 (136.2)	
English	Consonant	Minimal	84.6 (9.6)	283.7 (85.7)	2.5 (1.1)
		Non-minimal	86.9 (3.9)	484.4 (168.4)	
	Rhyme	Minimal	93.1 (4.6)	399.5 (94.8)	2.8 (0.9)
		Non-minimal	86.7 (5.5)	475.8 (76.9)	
	Tone	Minimal	94.4 (5.7)	390.6 (107.5)	3.3 (1.3)
		Non-minimal	78.6 (12.7)	494.7 (137.0)	

*Values are expressed as mean and standard deviation (in parentheses).

^a Scalar units are from 1 to 5 (1 = easy; 3 = medium; 5 = hard) for consonant, rhyme, and tone tasks across stimulus types.

RESULTS

Behavioral Performance

Behavioral measures of task performance by Chinese and English groups are given in Table II. Repeated measures ANOVAs were performed on response accuracy, reaction time, and subjects' ratings of task difficulty. Response accuracy revealed a significant 3-way interaction between task, stimulus type, and group [$F(2,36) = 7.79, P = .0015$]. Pooling across stimulus types, simple main effects ($\alpha = .01$) of group for each task (C, R, T) showed that Chinese listeners were more accurate than English listeners. Chinese listeners were also more accurate than English in all task/stimulus type combinations except for the consonant/non-minimal. For both groups, accuracy was lower in the consonant task than in the rhyme or tone [$F(2,36) = 10.67, P = .0016$], as well as being higher in minimal pairs as compared to non-minimal [$F(1,18) = 38.45, P < .0001$].

Reaction times yielded significant main effects of task [$F(2,36) = 7.77, P = .0002$] and stimulus type [$F(1,18) = 58.62, P < .0001$]. Post hoc Tukey-adjusted multiple comparisons ($\alpha = .01$) revealed that reaction times for both groups were shorter in the consonant task than in the rhyme or tone, and longer for non-minimal pairs as compared to minimal.

Each task (consonant, rhyme, tone) was self-rated by listeners on a 5-point difficulty scale (1 = easy, 3 = medium, 5 = hard). Subjective ratings of task difficulty showed an interaction between task and group [$F(2,36) = 5.73, P = .0069$]. Tests of the simple main effects of group for task indicated that the tone task was easier for the Chinese group than for the English.

Between-Group Comparisons Tasks (Consonant, Rhyme, Tone) by Minimal Pairs

A comparison of C_{minimal} , R_{minimal} , and T_{minimal} to passive listening (H) revealed significant increases of activation in frontal, temporal, and parietal regions predominantly in the

left hemisphere (LH) in the Chinese group relative to the English group (Figs. 1–3; Table III). In the frontal lobe, a common peak of activation was observed across tasks in the left posterior middle frontal gyrus (MFG) (Figs. 1,2; top, $x = -41, C_{\text{minimal}}$, R_{minimal} ; focus for T_{minimal} listed in Table III only). There was increased activity in the left anterolateral MFG for R_{minimal} and in the left superior frontal gyrus (SFG) for C_{minimal} and R_{minimal} . Temporal lobe activity was restricted to the left posterior middle temporal gyrus (MTG) across tasks (Figs. 1–3; top, $x = -51, -47$). Parietal lobe foci were restricted to the LH, centered in both the superior and inferior parietal lobules in C_{minimal} and R_{minimal} in the inferior parietal lobule in T_{minimal} (Figs. 1,2, top, $x = -47, -41$; Fig. 3, top, $x = -47, -43$).

Tasks (Consonant, Rhyme, Tone) by Nonminimal Pairs

A comparison of $C_{\text{non-minimal}}$, $R_{\text{non-minimal}}$, and $T_{\text{non-minimal}}$ relative to passive listening (H) also revealed significant increases of activation in frontal, temporal, and parietal regions predominantly in the LH in the Chinese group relative to the English group (Figs. 1–3; Table III). In the frontal lobe, the left anterolateral MFG was activated in common across tasks (Figs. 1,2, bottom panel, $x = -41$; Fig. 3, bottom, $x = -43$). Increased activity in the right anterior MFG was observed in $R_{\text{non-minimal}}$ only (Fig. 2, bottom panel, $x = +33$). Posterior prefrontal foci were predominantly left-sided across tasks, though activation was observed in the right posterior inferior frontal gyrus (IFG) in $R_{\text{non-minimal}}$ (Table III). In $C_{\text{non-minimal}}$ and $R_{\text{non-minimal}}$, the left posterior prefrontal foci were centered in the posterior MFG, extending ventrally across the inferior frontal sulcus into the IFG (Figs. 1,2, bottom, $x = -47, -41$). In $T_{\text{non-minimal}}$, the posterior prefrontal focus was centered in the inferior segment of the precentral sulcus, extending rostrally into posterior and dorsal aspects of the IFG (Fig. 3, bottom, $x = -51, -47$). Activation foci in the left posterior SFG or superior frontal sulcus were observed across tasks except for T_{minimal} (Table

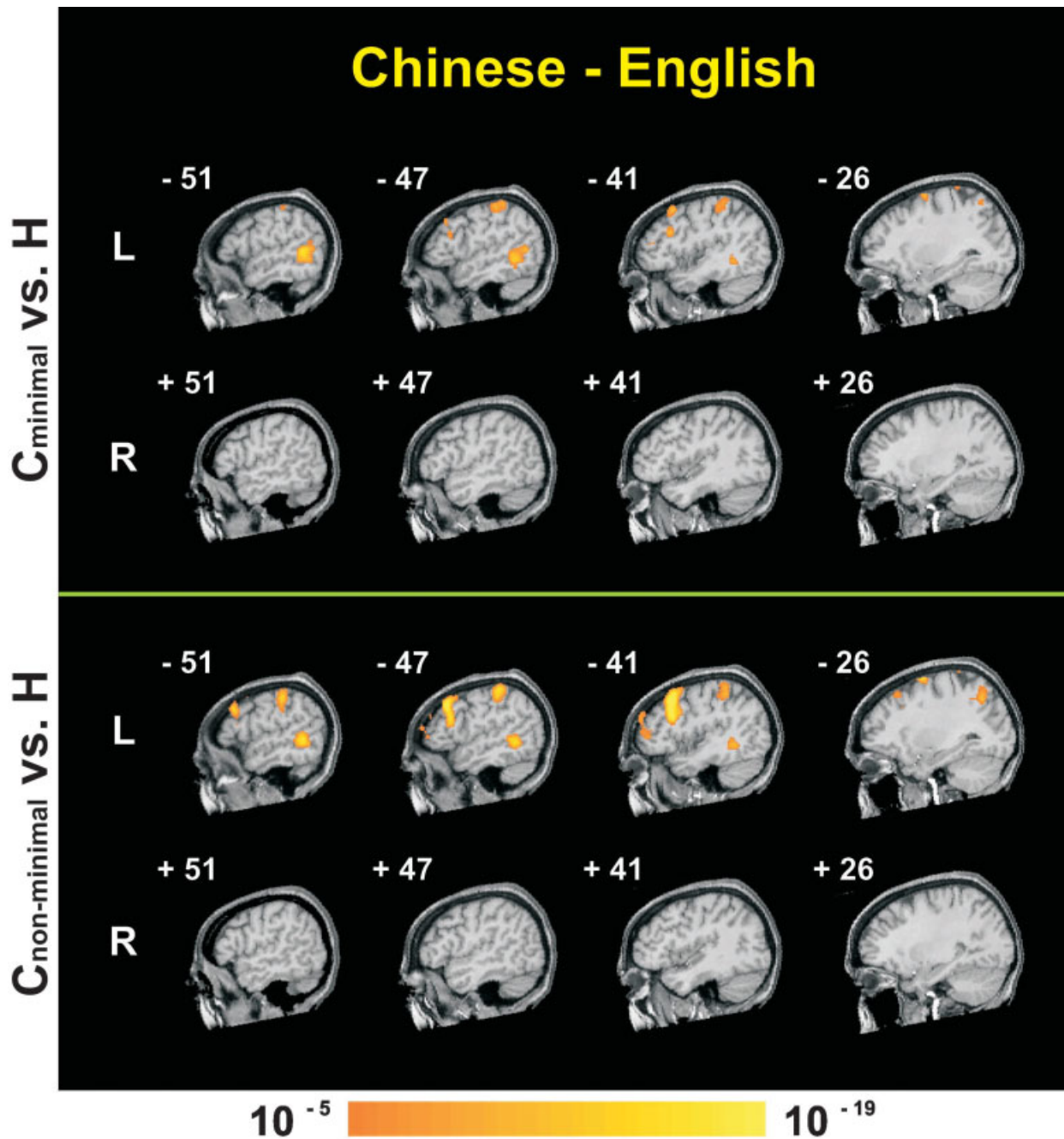


Figure 1.

Averaged fMRI activation maps obtained from comparison of consonant_{minimal} (**top**) and consonant_{non-minimal} (**bottom**) discrimination judgments relative to a passive listening baseline between the two language groups (Chinese minus English). Each panel shows left/right sagittal sections through stereotaxic space super-

imposed onto a representative brain anatomy. Stereotaxic coordinates (mm) are derived from the human brain atlas of Talairach and Tournoux [1988]. C = consonant; H = hums. Both panels show activation foci in left-sided frontal, temporal, and parietal regions.

III). In the temporal lobe, the left posterior MTG was activated across tasks (Figs. 1–3, bottom, $x = -51, -47$). In the parietal lobe, peak foci were found in the inferior parietal lobule in C_{non-minimal} and R_{non-minimal} (Figs. 1,2, bottom, $x = -47, -41$).

Peak activation was located in the banks of the intraparietal sulcus in T_{non-minimal} (Fig. 3, bottom, $x = -47, -43$), projecting extensively into both the inferior and superior parietal lobules (not shown in Fig. 3).

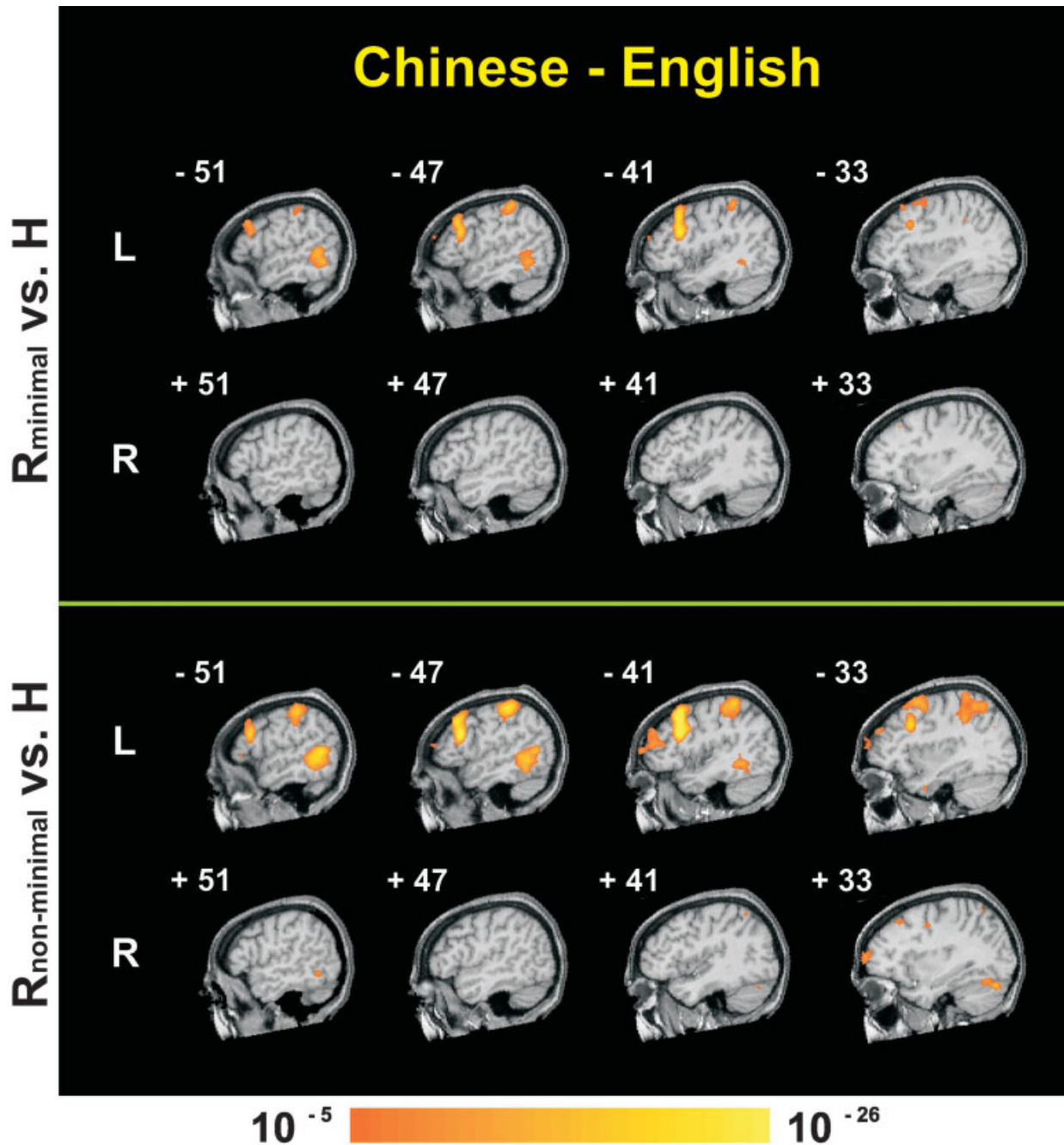


Figure 2.

Averaged fMRI activation maps obtained from comparison of rhyme_{minimal} (**top**) and rhyme_{non-minimal} (**bottom**) discrimination judgments relative to a passive listening baseline between the two

language groups (Chinese minus English). R = rhyme; H = hums. Both panels show activation foci in left-sided frontal, temporal, and parietal regions. See also caption to Figure 1.

Within-Group Comparisons

Non-minimal vs. minimal pairs by task (consonant, rhyme, tone)

A comparison of C_{non-minimal} relative to C_{minimal} showed significant increases of activation in the left inferior parietal

region in the Chinese group only (Table IV). In R_{non-minimal} minus R_{minimal}, an activation focus in the right IFG was centered in the pars triangularis, extending medially into the anterior insula. Activation in anterior regions of the MFG occurred bilaterally. Also noted was activity in the medial frontal gyrus. No significant differences in activation were

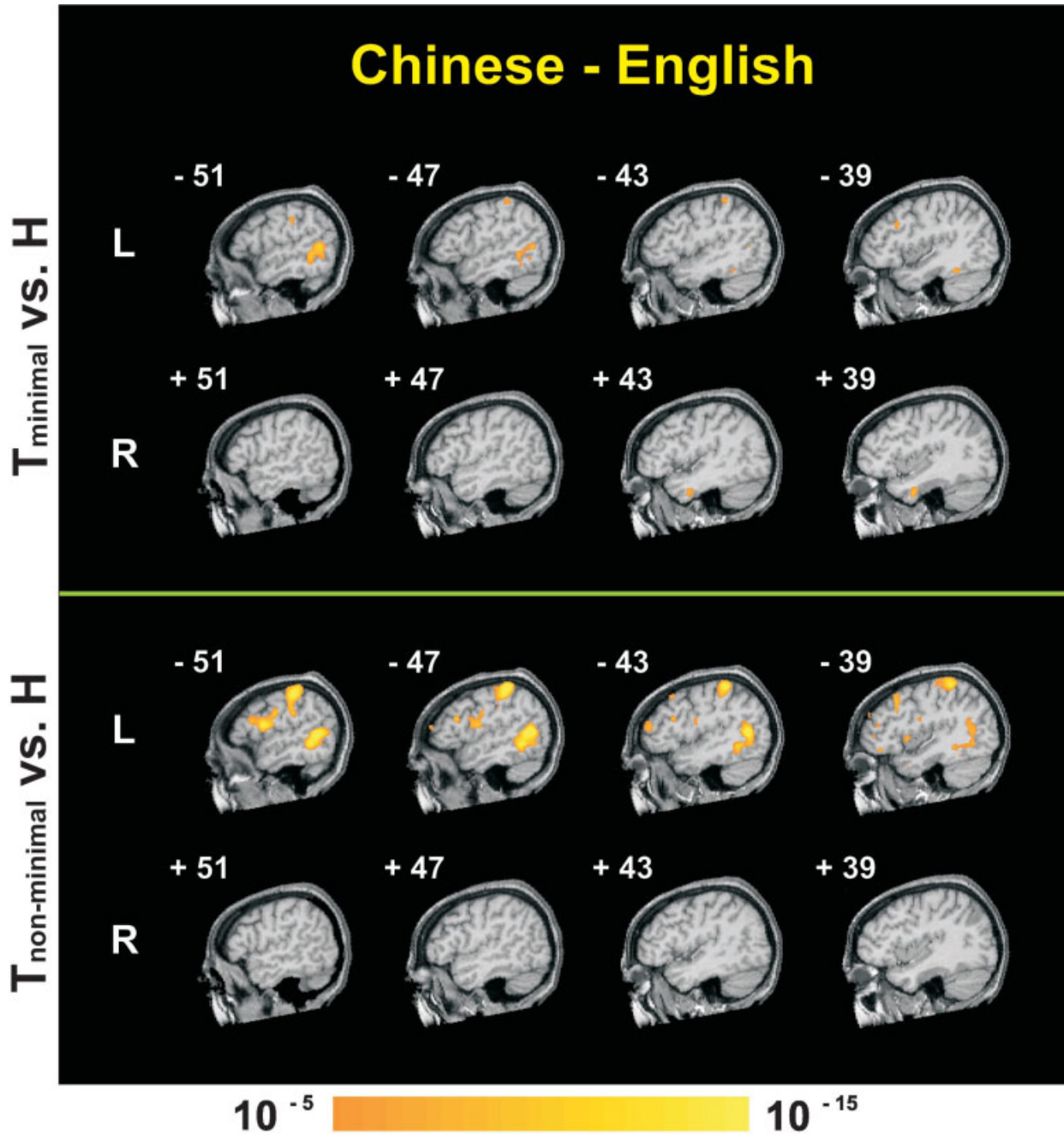


Figure 3.

Averaged fMRI activation maps obtained from comparison of tone-minimal (**top**) and tone_{non-minimal} (**bottom**) discrimination judgments relative to a passive listening baseline between the two language groups (Chinese minus English). T = tone; H = hums.

The top and bottom panels, respectively, show activation foci in left-sided and bilateral frontal, temporal, and parietal regions. See also caption to Figure 1.

observed for the English group when comparing either $C_{\text{non-minimal}}$ minus C_{minimal} or $R_{\text{non-minimal}}$ minus R_{minimal} .

A comparison of $T_{\text{non-minimal}}$ relative to T_{minimal} showed significant increases of activation in frontal and parietal regions in both groups (Fig. 4; Table IV). Anterior frontal

activations were observed bilaterally in the Chinese group (Fig. 4, top, $x = \pm 41, \pm 37$), but were lateralized to the right hemisphere (RH) in the English group (Fig. 4, bottom, $x = +47$). Posterior prefrontal activations were observed in dorsal aspects of the IFG bilaterally for the Chinese group

TABLE III. Significant activation foci for between-group comparisons of each task and stimulus type*

Region	Chinese - English				Peak <i>t</i> value	Extent (ml)
	BA	<i>x</i>	<i>y</i>	<i>z</i>		
Consonant [minimal pairs] vs. passive listening to hums						
Frontal						
L posterior MFG	9	-37	15	31	6.68	1.14
L posterior SFG	6	-17	-3	64	8.35	1.45
Temporal						
L posterior MTG	21/37	-54	-54	10	8.48	3.48
Parietal						
L inferior parietal lobule	40	-46	-33	52	6.48	2.10
L superior parietal lobule	7	-17	-62	61	6.77	1.56
Consonant [non-minimal pairs] vs. passive listening to hums						
Frontal						
L anterior MFG	9/46	-37	45	29	7.37	1.89
L posterior MFG	8/6	-43	16	42	11.64	7.42
L posterior SFG	6	-19	-3	64	9.44	2.55
Temporal						
L posterior MTG	21	-50	-54	3	8.35	3.21
Parietal						
L inferior parietal lobule	40	-46	-37	51	8.31	4.56
L superior parietal lobule	7	-28	-65	44	6.43	1.16
Rhyme [minimal pairs] minus passive listening to hums						
Frontal						
L anterior MFG	46/9	-45	41	22	5.27	0.31
L posterior MFG	9/44	-40	13	29	10.46	5.78
L posterior SFS	6	-28	14	58	6.59	2.46
Temporal						
L posterior MTG	21/37	-54	-53	7	8.01	3.24
Parietal						
L inferior parietal lobule	40	-46	-39	54	7.43	2.55
L superior parietal lobule	7	-26	-65	44	5.74	0.97
Rhyme [non-minimal pairs] minus passive listening to hums						
Frontal						
L anterior MFG	46	-38	45	29	7.38	3.68
R anterior MFG	10	34	59	10	6.48	0.47
L posterior MFG	6/8	-41	14	42	14.29	8.68
R posterior IFG	44	47	15	17	5.46	0.26
L posterior SFS	6	-21	15	49	9.89	9.47
M medial frontal gyrus	6/32	2	9	44	7.80	1.76
Temporal						
L posterior MTG	21/37	-52	-58	9	9.97	6.62
Parietal						
L intraparietal sulcus	7/40	-46	-33	51	10.22	16.45
Tone [minimal pairs] vs. passive listening to hums						
Frontal						
L posterior MFG	9	-39	15	31	5.38	0.16
Temporal						
L posterior MTG	21/37	-54	-55	8	7.05	2.35
Parietal						
L inferior parietal lobule	40	-43	-40	55	5.49	0.48
Tone [non-minimal pairs] vs. passive listening to hums						
Frontal						
L anterior MFG	46	-42	39	17	5.72	0.69
L precentral sulcus	6/44	-52	0	17	8.73	2.94
L posterior SFG	6	-17	-2	64	7.19	3.52
Temporal						
L posterior MTG	37/21	-47	-65	10	11.52	7.12
Parietal						
L intraparietal sulcus	7/40	-45	-40	55	9.48	15.95

*L = left, R = right, M = medial; BA = Brodmann's areas; IFG = inferior frontal gyrus, MFG = middle frontal gyrus, MTG = middle temporal gyrus, SFG = superior frontal gyrus, SFS = superior frontal sulcus. Stereotaxic coordinates (mm) are derived from a human brain atlas [Talairach and Tournoux, 1988], and refer to the peak *t* value for each region. *x* = distance (mm) to right (+) or left (-) of the midsagittal plane; *y* = distance anterior (+) or posterior (-) to vertical plane through the anterior commissure; *z* = distance above (+) or below (-) the intercommissural (AC-PC) line. Extent (> 0.15 ml) refers to the size of activation (number of voxels > threshold).

TABLE IV. Significant activation foci for within-group comparisons of non-minimal vs. minimal stimulus types per task*

Region	Chinese						English					
	BA	x	y	z	Peak <i>t</i> value	Extent (ml)	BA	x	y	z	Peak <i>t</i> value	Extent (ml)
Consonant_{non-minimal} minus Consonant_{minimal} ($C_{non-minimal} - C_{minimal}$)												
Parietal												
L inferior parietal lobule	40	-54	-30	43	5.59	0.82						
Rhyme_{non-minimal} minus Rhyme_{minimal} ($R_{non-minimal} - R_{minimal}$)												
Frontal												
L anterior MFG	46	-38	32	14	5.30	0.18						
R anterior MFG	10	34	55	10	5.38	0.33						
R IFG	45	48	24	-2	5.78	1.11						
M medial frontal gyrus	6	-2	7	61	6.99	3.15						
Tone_{non-minimal} minus Tone_{minimal} ($T_{non-minimal} - T_{minimal}$)												
Frontal												
L anterior MFG	46	-41	41	19	5.76	1.40						
R anterior MFG	10	40	53	10	9.04	4.28	46	47	36	26	6.02	0.68
L posterior MFG							8/9	-39	9	28	6.26	1.61
R posterior MFG							9	50	19	38	6.46	3.54
L posterior IFG	44	-43	11	23	6.15	1.60						
R posterior IFG	44	55	16	29	6.72	1.88						
L anterior insula		-30	14	10	5.69	0.79						
R frontal operculum	45	36	26	10	6.65	1.73	45	34	25	7	5.44	0.65
M medial frontal gyrus	6/32	7	18	46	6.30	3.28	8	8	28	37	4.93	0.24
Parietal												
L interior parietal lobule	40	-51	-33	44	5.45	1.38	40	-47	-40	36	6.27	1.25
R intraparietal sulcus	40/7	39	-42	44	5.00	0.27	40	50	-41	44	6.26	1.91

*L = left, R = right, M = medial; BA = Brodmann's areas; IFG = inferior frontal gyrus, MFG = middle frontal gyrus. See also note to Table III.

(Fig. 4, top, $x = \pm 55, \pm 51, \pm 47, \pm 41$), extending ventrally to the Sylvian fissure on the right side (Fig. 4, top, $x = +55$). For the English group, activation foci in the posterior MFG also were observed bilaterally, more extensively in the RH (Fig. 4, bottom, $+55, +51, \pm 47, \pm 41$). Increased activity in the left anterior insula was observed in the Chinese group only (Fig. 4, top, $x = -31$). Activity was also observed in the medial frontal gyrus in both groups, extending ventrally into the cingulate gyrus in the Chinese group. Frontal foci were observed in the right frontal operculum in both groups (Fig. 4, top/bottom, $x = +37$), extending laterally into the rostroventral IFG in the Chinese group (Fig. 4, top, $x = +41, +47, +51$). In the parietal lobe, activity was predominantly left-sided in the Chinese group, bilateral in the English group (Fig. 4, top/bottom, $x = \pm 51, \pm 47$).

Consonant vs. Rhyme vs. Tone Tasks by Stimulus Type (minimal, non-minimal)

In the Chinese group, a comparison of $R_{non-minimal}$ relative to either $C_{non-minimal}$ or $T_{non-minimal}$ revealed activation foci in the left posterior MFG, extending ventrally into the inferior frontal sulcus (Fig. 5; Table V). When comparing $R_{minimal}$ to $T_{minimal}$, the focus was located more ventrally in the posterodorsal IFG, though clearly projecting into the inferior frontal sulcus. No other direct contrasts of tasks by

stimulus type revealed any areas of significant activation for the Chinese group.

In the English group, a comparison of $T_{non-minimal}$ relative to $C_{non-minimal}$ revealed activation bilaterally in the posterior MFG and inferior parietal lobule (Fig. 5; Table V). The posterior MFG activation was much more extensive in the RH. The right-sided focus extended rostrally throughout the MFG as far front as frontopolar cortex; ventrolaterally into the IFG, especially contiguous to the anterior ascending ramus involving both pars opercularis and pars triangularis; and ventromedially into the frontal operculum. The activation pattern for $T_{non-minimal}$ minus $R_{non-minimal}$ was approximately identical to that for $T_{non-minimal}$ minus $C_{non-minimal}$. The left MFG focus, however, was about 26 mm anterior to that in $T_{non-minimal}$ minus $C_{non-minimal}$. No other direct contrasts of tasks by stimulus type revealed any areas of significant activation for the English group.

DISCUSSION

Neural Circuitry Underlying Phonological Processing

Our major findings reveal that extraction of phonological units in Chinese, segmental and suprasegmental alike, elicits

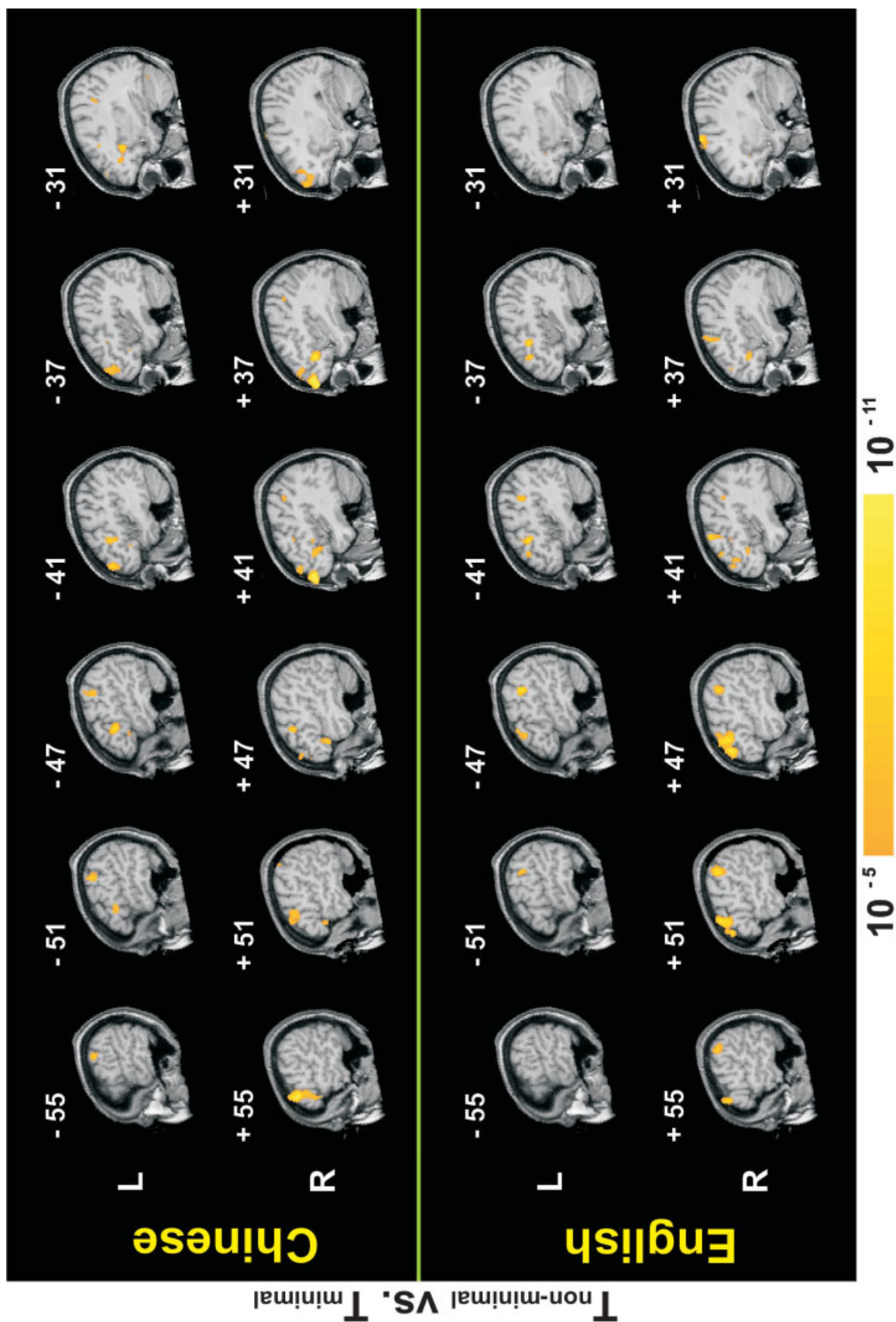


Figure 4.

Averaged fMRI activation maps obtained from comparison of tone_{non-minimal} vs. tone_{minimal} discrimination judgments within the Chinese (**top**) and English (**bottom**) language groups. T = tone. Both panels show activation foci in anterior and posterior regions of the frontal lobe. However, the posterior foci are centered in the inferior frontal gyrus in the Chinese group, the middle frontal gyrus in the English group. The top and bottom panels, respectively, show predominantly left-sided and bilateral activation in the parietal lobe. See also caption to Figure 1.

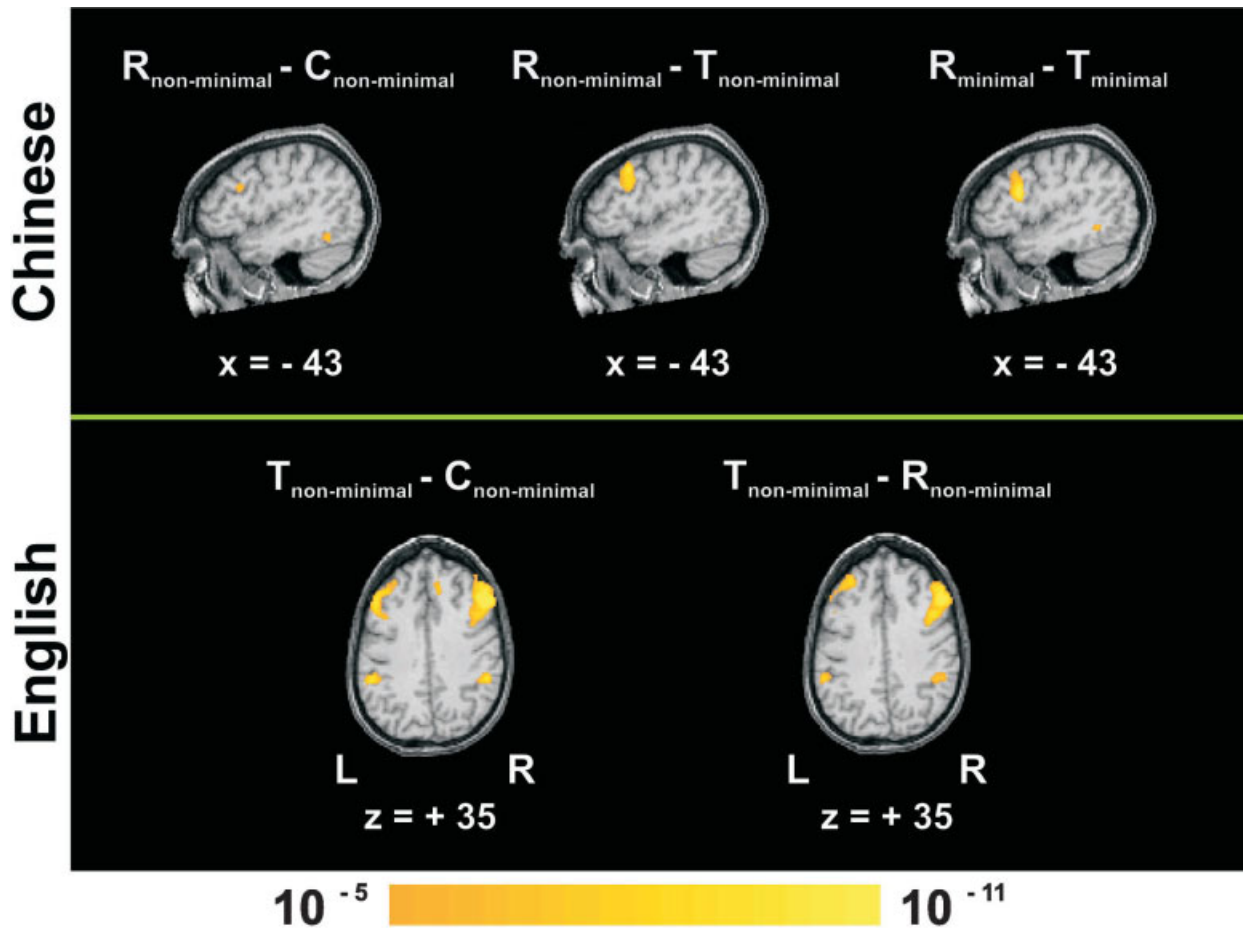


Figure 5.

Averaged fMRI activation maps obtained from comparison of consonant vs. rhyme vs. tone tasks by condition within the Chinese (**top**) and English (**bottom**) language groups. C = consonant; R = rhyme; T = tone; L = left hemisphere; R = right hemisphere. The Chinese group shows a common focus of activation in the left posterior middle frontal gyrus in comparisons of $R_{\text{non-minimal}}$ minus

$C_{\text{non-minimal}}$, $R_{\text{non-minimal}}$ minus $T_{\text{non-minimal}}$, and R_{minimal} minus T_{minimal} . The English group, on the other hand, shows activation of the posterior middle frontal gyrus and inferior parietal lobule bilaterally in comparisons of $T_{\text{non-minimal}}$ minus $C_{\text{non-minimal}}$ and $T_{\text{non-minimal}}$ minus $R_{\text{non-minimal}}$.

activation of a widely-distributed cortical network in the LH (see Chinese vs. English). Direct comparisons of non-minimal to minimal pairs show that the process of phonetic extraction varies as a function of the sub-syllabic unit itself - C, R, or T (see Non-minimal vs. minimal pairs). In either non-minimal or minimal pairs, direct comparisons of sub-syllabic units (C, R, T) indicate that the rhyme elicits greater activation in the left posterior MFG when compared to consonants and tones (see Consonant vs. rhyme vs. tone).

Chinese vs. English

Direct between-group comparisons in both minimal and non-minimal pairs (Figs. 1-3; Table III) reveal increased activation for the Chinese group in predominantly left-sided frontal, parietal, and temporal regions, all of which have been associated with attention-modulated phonological pro-

cessing [Shaywitz et al., 2001]. In the frontal lobe, there are two discrete prefrontal cortical areas of activation. One area of activation, irrespective of stimulus type, is in the posterior inferior MFG extending into the IFS and bordering aspects of the IFG (BA 6/9/44). It has been proposed that dorsolateral posterior aspects of PFC mediate subvocal rehearsal in verbal working memory [Paulesu et al., 1993, 1996]. The other area of activation is in the dorsolateral anterior MFG extending ventrally into bordering aspects of the IFG (BA 46). This region has been implicated with selective attention [Gehring and Knight, 2002] and with active manipulation of information held in short-term storage [Smith and Jonides, 1999]. Moreover, activity in the anterior MFG varies depending on both stimulus type and sub-syllabic component. This finding might reflect differential use of selective attention mechanisms as applied to consonants, rhymes, and tones.

TABLE V. Significant activation foci for within-group comparisons of consonant, rhyme, and tone tasks per stimulus type*

Region	BA	x	y	z	peak <i>t</i> value	extent (ml)
Chinese group						
Rhyme _{non-minimal} minus Consonant _{non-minimal} ($R_{non-minimal} - C_{non-minimal}$)						
Frontal						
L posterior MFG	9	-43	13	28	4.99	0.21
L fusiform gyrus	37	-42	-59	-13	5.52	0.28
Rhyme _{non-minimal} minus Tone _{non-minimal} ($R_{non-minimal} - T_{non-minimal}$)						
Frontal						
L posterior MFG	9/8	-43	12	43	6.56	2.14
Rhyme _{minimal} minus Tone _{minimal} ($R_{minimal} - T_{minimal}$)						
Frontal						
L posterior IFG	9/44	-43	10	24	7.06	2.55
R frontal operculum	45	34	25	5	4.94	0.23
Parietal						
L superior parietal lobule	7	-24	-65	47	6.00	0.91
English group						
Tone _{non-minimal} minus Consonant _{non-minimal} ($T_{non-minimal} - C_{non-minimal}$)						
Frontal						
L posterior MFG	6/8	-44	9	42	6.63	8.14
R posterior MFG	6	40	7	51	8.92	21.22
Parietal						
L inferior parietal lobule	40	-49	-43	33	6.58	1.17
R inferior parietal lobule	40	50	-40	40	7.75	2.63
Tone _{non-minimal} minus Rhyme _{non-minimal} ($T_{non-minimal} - R_{non-minimal}$)						
Frontal						
L anterior MFG	9	-31	35	37	6.07	3.26
R posterior MFG	6	30	5	63	8.42	19.32
Parietal						
L inferior parietal lobule	40	-52	-43	33	5.92	0.63
R inferior parietal lobule	40	49	-40	45	7.34	5.47

*See also note to Table III.

Activation is seen in both the parietal and temporal lobes, irrespective of stimulus type. Parietal activation may be related to attentional orientation [Corbetta et al., 2000; Posner et al., 1984]. In our task, subjects had to engage/disengage attention to the first and last target syllable, and switch their attention to targeted sub-syllabic components. Activation was greater for Chinese listeners because of their implicit knowledge of the structural elements of a Mandarin syllable. Parietal regions are also considered to be part of a network that mediate the short-term storage and retrieval of phonologically coded verbal material [Jonides et al., 1998]. Temporal lobe activity may reflect involvement in phonetic recognition from the auditory modality [Shaywitz et al., 2001]. Activation is expected to be greater for Chinese listeners because of their implicit knowledge of Chinese phonology.

Non-minimal vs. minimal pairs

Within-group comparisons of non-minimal and minimal pairs show that frontal and parietal activity varies for each sub-syllabic component (Fig. 4; Table IV). The Chinese group shows bilateral activation of the anterior MFG for

$R_{non-minimal}$ minus $R_{minimal}$ and $T_{non-minimal}$ minus $T_{minimal}$ only. In temporal order, syllable-initial consonants occur before rhymes or tones, whereas rhymes and tones occur concurrently. Perhaps, more attentional resources are recruited in non-minimal pairs to extract phonetic information from units (R, T) that occur simultaneously than for units that are presented sequentially (C).

Bilateral posterior IFG activity is seen in $T_{non-minimal}$ minus $T_{minimal}$. No activation of the left/right posterior MFG or IFG was observed for either consonants ($C_{non-minimal}$ minus $C_{minimal}$) or rhymes ($R_{non-minimal}$ minus $R_{minimal}$). This unique effect for tonal processing is likely due to an increase in subvocal rehearsal due to late arrival of tonal information [Cutler and Chen, 1997]. Our speeded-response task may have increased the computational demands of extracting tonal information from non-minimal pairs. Activation in the right posterior IFG may reflect concurrent recruitment of pitch mechanisms [Zatorre et al., 2002].

The absence of significant brain activity in ventral aspects of the IFG for consonants and rhymes is noteworthy. In working memory studies, the former has been implicated in

articulatory rehearsal. Segmental (consonant or rhyme) matching on written Chinese characters activates ventral aspects of the IFG [Siok et al., 2003; Tan et al., 2001, 2003]. However, this activation may reflect the conversion of visual text to speech. In this study, such conversion is unnecessary because the input modality is auditory, even though articulatory recoding from audition to speech may be necessary.

Indeed, in the auditory modality, it has been demonstrated that phonetic extraction of Chinese tones reveals a left-sided dorsal frontoparietal network that recruits mediational processes related to selective attention and internal guidance [Li et al., in press]. These processes apply to other domains besides language. In the language domain, these mediational processes make it possible for a listener to extract linguistically significant parameters in his/her native language. It is suggested that the left posterior MFG, in particular, may be responsible for encoding the perceived signal, storing it for immediate matching, and transferring the matched code to the neighboring motor area for a response.

These findings complement and extend the seminal work of Burton et al. [2000]. Surprisingly, we find no significant differences in a left dorsolateral posterior subregion of prefrontal cortex for either consonants ($C_{\text{non-minimal}}$ minus C_{minimal}) or rhymes ($R_{\text{non-minimal}}$ minus R_{minimal}). Our study, however, is not a replication of their paradigm applied to a tone language. The major methodological difference lies with the trial design. No intervening distracters were employed in their study. The presence of a distracter makes the working memory demands of minimal pairs more similar to those of non-minimal pairs. Its presence militates against a simple acoustic strategy in judging minimal pairs, which, in turn, maximizes the likelihood of Chinese listeners using a segmentation strategy for both stimulus types. Moreover, subjects were given the same instructions for both minimal and non-minimal pairs. These instructions make it even more likely that subjects did not consciously differentiate their segmentation strategies according to the stimulus type. Another possibly relevant factor is the interstimulus interval (ISI). In Burton et al. [2000], the ISI was 50 ms; in this study, 200 ms. The longer ISI also militates against a simple acoustic strategy.

In contrast, the English group shows bilateral posterior MFG activation as well as activation foci in the right anterior MFG and frontal operculum for non-minimal vs. minimal pairs of tones only. This predominantly right-sided prefrontal activation may reflect involvement of working memory in non-linguistic pitch processing [Zatorre et al., 1992, 1994].

Consonant vs. rhyme vs. tone

Rhymes induce greater activation in the left posterior MFG for the Chinese group when compared to C and T in non-minimal pairs (Fig. 5; Table V). Temporally speaking, rhyme information arrives later than consonant information. Rhymes are also longer in duration than consonants [Howie, 1976]. These factors may account for why rhymes require more time for subvocal rehearsal than consonants.

But why do rhymes require more time for subvocal rehearsal than tones? Rhymes appear to be pivotal for integrating tonal information with the Mandarin syllable [Ye and Connine, 1999], resulting in increased activation for subvocal rehearsal regardless of stimulus type. There are 37 rhymes but only four tones in Mandarin. Only 170 syllables carry all four Mandarin tones [Dictionary, 1971]. The comparatively small number of tonal categories and their co-occurrence restrictions with rhymes make it easier for Chinese listeners to extract tonal information.

In contrast, tones induce greater activation in the right posterior MFG for the English group when compared to C and R in non-minimal pairs (Fig. 5; Table V). The increased activation for subvocal rehearsal presumably is due to English listeners' lack of prior experience with Chinese tones. This right-sided bias for extracting pitch information points to a critical role for the RH in mediating pitch perception [Zatorre et al., 2002].

In a recent study of phoneme vs. syllable processing in written Chinese [Siok et al., 2003], it is found that the left middle frontal cortex mediates syllables, whereas the left inferior prefrontal gyrus mediates the processing of phonemes (initial consonants) and phonological segmentation. Our findings, on the other hand, show that patterns of activation in left prefrontal cortex vary depending on sub-syllabic segmental (C, R) and suprasegmental (T) units. Isolating the functions of these frontal regions in both spoken and written Chinese clearly warrants further empirical investigation.

Phonological vs. Semantic Processing in the Frontal Lobe

Our findings are consistent with emerging evidence in favor of a functional distinction between anterior/ventral and posterior/dorsal regions within left inferior prefrontal cortex that are differentially activated by attention to semantics and to phonology, respectively. Specifically, activations attributed to semantic processing are generally located anteriorly within the ventral inferior IFG (BA 47/10), whereas activations attributed to phonological processing have been more frequently located posteriorly within the MFG/IFG (BA 6/44) [Burton, 2001; McDermott et al., 2003; Poldrack et al., 1999]. Direct between-group comparisons (Figs. 1–3; Table III) show phonologically-related areas of peak activation posteriorly in the left PFC (BA 6/9/44) across sub-syllabic constituents (C, R, T) irrespective of stimulus type (minimal, non-minimal). No activation was observed in semantically-related anterior/ventral regions (BA 47) of the left frontal lobe. Though the current design cannot exclude the involvement of semantic information in tone processing, it appears likely that activation of the left posterior/dorsal functional subregion in this study is driven by attention to phonology instead of semantics.

Our findings on consonants, rhymes, and tones cannot be attributed to lexico-semantic effects. In the $T_{\text{non-minimal}}$ minus T_{minimal} comparison for the Chinese group (Fig. 4; Table IV), we observe bilateral activation in the posterior dorsal

aspects of the IFG. In the $R_{\text{non-minimal}}$ minus $C_{\text{non-minimal}}$ or $T_{\text{non-minimal}}$ comparisons (Fig. 5; Table V), the common area of activation is located in the left posterior MFG. None of these comparisons show activation in areas that have been associated with semantically-related activity [McDermott et al., 2003]. Therefore, these posterior prefrontal activations are likely to reflect phonologically-related activity [see Wong, 2002].

Competing Models of Chinese Syllables

Our findings favor a hierarchical, instead of flat, model of Mandarin syllable structure. A flat model cannot account for the increased activation in the left posterior MFG/IFG associated with R as compared to C and T for the Chinese group (Fig. 5; Table V). One hierarchical model that has been proposed provides for a separate level of tonal representation, but allows for highly interconnected links between tones and rhymes [Ye and Connine, 1999]. Their findings demonstrate that tonal processing may be a graded effect dependent on task demands rather than an all-or-none phenomenon. In our study, task demands emphasized the pre-eminent role of the rhyme as the carrier of lexical tone in Mandarin. This asymmetry between rhymes and tones is likely due to higher-level processing demands instead of perceptual constraints.

Hemispheric Specialization for Segmental and Suprasegmental Units

A clearer picture now begins to emerge regarding the significance of the segmental/suprasegmental dichotomy as it applies to differential hemispheric roles in speech perception. In this study, consonants, rhymes, and tones all resulted in a predominantly left-sided fronto-temporo-parietal pattern of activation regardless of whether the unit is segmental or suprasegmental.

Our findings are consistent with previous brain imaging data that show LH lateralization of tonal perception [Gandour et al., 2000, 2002; Hsieh et al., 2001]. However, all suprasegmentals are not mediated by the LH exclusively. The RH appears to play an important role in the processing of sentence-level prosodic units [Gandour et al., 2003]. Imaging data from speech production similarly argue for recruitment of the RH in processing speech prosody at the sentence level, regardless of linguistic or emotional functions [Dogil et al., 2002]. Cross-linguistic group comparisons reveal that processing strategies for segmentals and suprasegmentals depend on language experience.

Effects of Task Performance on Brain Activation Patterns

Task performance of Chinese and English subjects cannot account for their differential patterns of brain activation. Although Chinese listeners achieved higher levels of accuracy than English listeners across tasks, the latter group reached 85% or higher on all tasks except tone (79%). Nevertheless, between-group comparisons revealed common ac-

tivation in fronto-temporo-parietal areas irrespective of task. For both groups, reaction times were longer in the R or T task than in the C, and longer in non-minimal pairs than in minimal. Yet within-group comparisons by task and stimulus type revealed that brain activity of the Chinese group differed substantially from that of the English group. Therefore, it seems reasonable to attribute these differences in brain activity to language-specific effects on the processing of sub-syllabic components.

Conclusions

We may conclude that extraction of sub-syllabic phonological units, segmental and suprasegmental alike, elicits activation of a widely-distributed cortical network in the LH. The activation of this network in processing non-minimal pairs shows that phonetic segmentation varies as a function of the sub-syllabic unit (C, R, T) in relation to both task demands (e.g., attention, memory) and syllable structure. Our findings are the first to demonstrate how anterior MFG regions may play a crucial role in extraction of phonological information from non-minimal pairs. The fact that the rhyme elicits greater activation in the left posterior MFG than either C or T further indicates that not all phonological units are treated alike with respect to phonetic segmentation. Future research is to be directed to teasing apart these attention and memory components as they relate to phonological processing in the brain.

ACKNOWLEDGMENTS

Funding was provided in part by an NIH postdoctoral traineeship (XL). This article was based in part on a report submitted in partial fulfillment of requirements for a Masters degree at Purdue University (Y.X.). We are grateful to J. Lowe, T. Osborn, and J. Zimmerman for their technical assistance in the MRI laboratory. Mark Lowe is currently affiliated with the Cleveland Clinic Foundation, Mario Dziedzic with MDZ Consulting Inc.

REFERENCES

- Baddeley A (1966): Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Q J Exp Psychol* 18:362–365.
- Bandettini PA, Jesmanowicz A, Wong EC, Hyde JS (1993): Processing strategies for time-course data sets in functional MRI of the human brain. *Magn Reson Med* 30:161–173.
- Bao Z (1990): Fanqie languages and reduplication. *Linguist Inquiry* 21:317–350.
- Baum S (2002): Consonant and vowel discrimination by brain-damaged individuals: effects of phonological segmentation. *J Neurolinguist* 15:447–461.
- Burton M (2001): The role of the inferior frontal cortex in phonological processing. *Cogn Sci* 25:695–709.
- Burton M, Small S, Blumstein S (2000): The role of segmentation in phonological processing: an fMRI investigation. *J Cogn Neurosci* 12:679–690.
- Caplan D, Gow D, Makris N (1995): Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology* 45:293–298.

- Chao YR (1968): A grammar of spoken Chinese. Berkeley, CA: University of California Press.
- Chen J-Y (1999): The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Appl Psycholinguist* 20:289–301.
- Corbetta M, Kincade JM, Ollinger JM, McAvooy MP, Shulman GL (2000): Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat Neurosci* 3:292–297.
- Cutler A, Chen HC (1997): Lexical tone in Cantonese spoken-word processing. *Percept Psychophys* 59:165–179.
- Demonet JF, Price C, Wise R, Frackowiak RS (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.
- Dictionary XH (1971): *Xin Hua Dictionary*. Beijing: Shangwu Press.
- Dogil G, Ackermann H, Grodd W, Haider H, Kamp H, Mayer J, Riecker A, Wildgruber D (2002): The speaking brain: a tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *J Neurolinguist* 15:59–90.
- Gandour J, Wong D, Hutchins G (1998): Pitch processing in the human brain is influenced by language experience. *Neuroreport* 9:2115–2119.
- Gandour J, Wong D, Hsieh L, Weinzapfel B, Van Lancker D, Hutchins GD (2000): A crosslinguistic PET study of tone perception. *J Cogn Neurosci* 12:207–222.
- Gandour J, Wong D, Lowe M, Dzemidzic M, Sathamnuwong N, Tong Y, Li X (2002a): A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *J Cogn Neurosci* 14:1076–1087.
- Gandour J, Wong D, Lowe M, Dzemidzic M, Sathamnuwong N, Tong Y, Lurito J (2002b): Neural circuitry underlying perception of duration depends on language experience. *Brain Lang* 83:268–290.
- Gandour J, Dzemidzic M, Wong D, Lowe M, Tong Y, Hsieh L, Sathamnuwong N, Lurito J (2003): Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain Lang* 84:318–336.
- Gehring WJ, Knight RT (2002): Lateral prefrontal damage affects processing selection but not attention switching. *Brain Res Cogn Brain Res* 13:267–279.
- Gow DW, Jr., Caplan D (1996): An examination of impaired acoustic-phonetic processing in aphasia. *Brain Lang* 52:386–407.
- Howie JM (1976): *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press. xxiii, 279.
- Hsieh L, Gandour J, Wong D, Hutchins GD (2001): Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain Lang* 76:227–252.
- Jiang A, Kennedy D, Baker J, Weiskoff R, Tootel R, Woods R, Benson R, Kwong K, Thomas J, Brady B, Rosen B, Belliveau J (1995): Motion detection and correction in functional MRI imaging. *Human Brain Mapping* 3:224–235.
- Jonides J, Schumacher EH, Smith EE, Koeppel RA, Awh E, Reuter-Lorenz PA, Marshuetz C, Willis CR (1998): The role of parietal cortex in verbal working memory. *J Neurosci* 18:5026–5034.
- Klein D, Zatorre R, Milner B, Zhao V (2001): A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage* 13:646–653.
- Li X, Gandour J, Talavage T, Wong D, Dzemidzic M, Lowe M, Tong Y. (in press): Selective attention to lexical tones recruits left dorsal frontoparietal network. *NeuroReport*.
- Lowe MJ, Russell DP (1999): Treatment of baseline drifts in fMRI time series analysis. *J Comput Assist Tomogr* 23:463–473.
- Lowe MJ, Sorenson JA (1997): Spatially filtering functional magnetic resonance imaging data. *Magn Reson Med* 37:723–729.
- McDermott KB, Petersen SE, Watson JM, Ojemann JG (2003): A procedure for identifying regions preferentially activated by attention to semantic and phonological relations using functional magnetic resonance imaging. *Neuropsychologia* 41:293–303.
- Oldfield RC (1971): The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Paulesu E, Frith CD, Frackowiak RS (1993): The neural correlates of the verbal component of working memory. *Nature* 362:342–345.
- Paulesu E, Frith U, Snowling M, Gallagher A, Morton J, Frackowiak RS, Frith CD (1996): Is developmental dyslexia a disconnection syndrome? Evidence from PET scanning. *Brain* 119:143–157.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD (1999): Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10:15–35.
- Posner M, Walker J, Friedrich F, Rafal R (1984): Effects of parietal injury on covert orienting of attention. *J Neurosci* 4:1863–1874.
- Shaywitz BA, Shaywitz SE, Pugh KR, Fulbright RK, Skudlarski P, Mencl WE, Constable RT, Marchione KE, Fletcher JM, Klorman R, Lacadie C, Gore JC (2001): The functional neural architecture of components of attention in language-processing tasks. *Neuroimage* 13:601–612.
- Siok WT, Jin Z, Fletcher P, Tan LH (2003): Distinct brain regions associated with syllable and phoneme. *Hum Brain Mapp* 18:201–207.
- Smith EE, Jonides J (1999): Storage and executive processes in the frontal lobes. *Science* 283:1657–1661.
- Talairach J, Tournoux P (1988): *Co-planar stereotaxic atlas of the human brain: three-dimensional proportional system. An approach to cerebral imaging*. New York: Thieme Medical Publishers.
- Tan LH, Liu HL, Perfetti CA, Spinks JA, Fox PT, Gao JH (2001): The neural system underlying Chinese logograph reading. *Neuroimage* 13:836–846.
- Tan LH, Spinks JA, Feng CM, Siok WT, Perfetti CA, Xiong J, Fox PT, Gao JH (2003): Neural systems of second language reading are shaped by native language. *Hum Brain Mapp* 18:158–166.
- Wang WS-Y (1967): Phonological features of tone. *Int J Am Linguist* 33:93–105.
- Wong P (2002): Hemispheric specialization of linguistic pitch patterns. *Brain Res Bull* 59:83–95.
- Woods R, Cherry S, Mazziotta J (1992): Rapid automated algorithm for aligning and reslicing PET images. *J Comput Assist Tomogr* 16:620–633.
- Xu Y (1991): Depth of phonological recoding in short-term memory. *Mem Cognit* 19:263–273.
- Xu Y (1998): Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55:179–203.
- Ye Y, Connine C (1999): Processing spoken Chinese: the role of tone information. *Lang Cogn Process* 14:609–630.
- Zatorre R, Evans A, Meyer E, Gjedde A (1992): Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849.
- Zatorre R, Evans A, Meyer E (1994): Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.
- Zatorre R, Meyer E, Gjedde A, Evans A (1996): PET studies of phonetic processing of speech: review, replication, and reanalysis. *Cereb Cortex* 6:21–30.
- Zatorre RJ, Belin P, Penhune VB (2002): Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* 6:37–46.