A Cross-Linguistic fMRI Study of Perception of Intonation and Emotion in Chinese

Jack Gandour,[1]* **Donald Wong,**[2] **Mario Dzemidzic,**[3] **Mark Lowe,**[3] **Yunxia Tong,**[4] **and Xiaojian Li**[1]

[1]*Department of Audiology and Speech Sciences, Purdue University, West Lafayette, Indiana*
[2]*Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, Indiana*
[3]*Department of Radiology, Indiana University School of Medicine, Indianapolis, Indiana*
[4]*School of Science, Purdue University, West Lafayette, Indiana*

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆
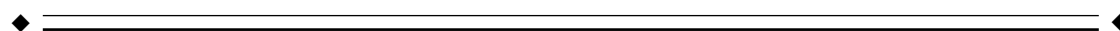
Abstract: Conflicting data from neurobehavioral studies of the perception of intonation (linguistic) and emotion (affective) in spoken language highlight the need to further examine how functional attributes of prosodic stimuli are related to hemispheric differences in processing capacity. Because of similarities in their acoustic profiles, intonation and emotion permit us to assess to what extent hemispheric lateralization of speech prosody depends on functional instead of acoustical properties. To examine how the brain processes linguistic and affective prosody, an fMRI study was conducted using Chinese, a tone language in which both intonation and emotion may be signaled prosodically, in addition to lexical tones. Ten Chinese and 10 English subjects were asked to perform discrimination judgments of intonation (I: statement, question) and emotion (E: happy, angry, sad) presented in semantically neutral Chinese sentences. A baseline task required passive listening to the same speech stimuli (S). In direct between-group comparisons, the Chinese group showed left-sided frontoparietal activation for both intonation (I vs. S) and emotion (E vs. S) relative to baseline. When comparing intonation relative to emotion (I vs. E), the Chinese group demonstrated prefrontal activation bilaterally; parietal activation in the left hemisphere only. The reverse comparison (E vs. I), on the other hand, revealed that activation occurred in anterior and posterior prefrontal regions of the right hemisphere only. These findings show that some aspects of perceptual processing of emotion are dissociable from intonation, and, moreover, that they are mediated by the right hemisphere. *Hum. Brain Mapping 18:149–157, 2003.*   © **2003 Wiley-Liss, Inc.**

Key words: functional magnetic resonance imaging; human auditory processing; language; speech perception; suprasegmental; selective attention; pitch
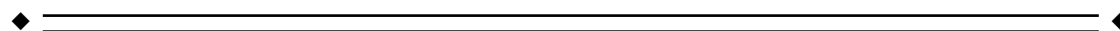
◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

INTRODUCTION

The neurobiological substrate underlying emotional prosody has yet to be clearly identified. Data from brain-damaged patients are conflicting and often implicate both the left (LH) and right (RH) hemispheres [Baum and Pell, 1999]. Working memory of verbal and emotional content of aurally presented information shares a distributed neural network bilaterally in fronto-parieto-occipital regions [Rama et al., 2001]. Yet certain operations associated with emotional percep-

tion appear to be mediated by the RH [Pell, 1998]. One major hypothesis claims that the RH is dominant for the production and perception of emotion [Ross, 2000]. Neuroimaging studies (EEG, fMRI, PET) show increased activation of right frontal regions during emotional prosody recognition [George et al., 1996; Wildgruber et al., 2002]. In tone languages (e.g., Chinese), patients with unilateral RH lesions exhibit selective deficits in comprehension and production tasks related to the affective component of prosody [Edmondson et al., 1987; Gandour et al., 1995; Hughes et al., 1983].

Chinese (Mandarin) has four lexical tones that can be described phonetically as high level, high rising, falling rising, and high falling. Less well known is its use of linguistic prosody to signal intonation. Of special interest is an unmarked question intonation that exhibits an overall higher pitch contour compared to a statement [Shen, 1990]. Voice fundamental frequency (F_0) level and range may be used to signal affective prosody [Ho, 1977].

Chinese tone and intonation span over syllable- and sentence-level prosodic frames, respectively. Earlier functional neuroimaging studies on linguistic prosody perception in Chinese have demonstrated that pitch contours associated with syllable-level lexical tones are processed in the LH, whereas those associated with sentence-level intonation engage neural mechanisms in both hemispheres but predominantly in the RH [Gandour et al., in press; Hsieh et al., 2001]. These findings suggest that linguistic prosody is subserved by both hemispheres, and that differential lateralization depends on the size of the prosodic frame.

We investigated neural regions involved in processing emotional and linguistic prosody at the sentence level. We predicted that lateralization of prosodic cues would be sensitive to affective properties of auditory stimuli, and that some aspects of emotion could be dissociated from linguistic prosody. By holding the prosodic frame constant, we avoided potential confounds arising from the use of differently sized hierarchical units in emotional and linguistic stimuli [see Buchanan et al., 2000]. By presenting Chinese sentences in linguistic and affective contexts to Chinese and English listeners, we could tease apart those effects on brain processing that are sensitive to language-specific factors.

SUBJECTS AND METHODS

Subjects

Ten adult Chinese (5 men, 5 women) and ten adult English (5 men, 5 women) speakers were closely matched in age (Chinese: $M = 26.1$; English: $M = 28.0$) and years of formal education (Chinese: $M = 16.9$; English: $M = 21.2$). All subjects were strongly right handed [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

Stimuli consisted of 36 pairs of Chinese sentences. The sentences were designed with two intonation patterns (*declarative, interrogative*) in combination with three emotional states (*happy, angry, sad*). Verbal content was neutral so that sentences were not biased to any particular emotion. Stimuli were balanced to eliminate valence as a potential confound [Sackheim et al., 1982]. Stimulus pairs were of three types: same sentence intonation, different emotion (16 pairs); different intonation, same emotion (16 pairs); same intonation, same emotion (4 pairs).

Tasks

The fMRI paradigm consisted of two *active judgment* conditions (I = intonation; E = emotion) and a *passive listening* condition (S = Chinese speech). In active conditions, subjects were instructed to direct their attention to particular features of auditory stimuli, make discrimination judgments, and respond by pressing a mouse button (Table I). In the intonation condition, subjects directed their attention to intonation (statement/question), and ignored any differences in emotion (happy/sad/angry). In the emotion condition, subjects directed their attention to emotion, and ignored any differences in intonation. In the passive listening condition, subjects listened passively to Chinese speech and responded by alternately pressing the left and right mouse buttons.

The speech tasks were designed to compare intonation (I) and emotion (E) to a passive listening baseline (S) as well as directly to one another. The baseline was designed to capture cognitive processes inherent to automatic, perceptual analysis, executive functions, and motor response formation. In I vs. S and E vs. S, we were able to compare focused attention to Chinese intonation and emotion, respectively, relative to automatic, unfocused attention to the same stimuli. In I vs. E, we were able

TABLE I. fMRI paradigm for Chinese speech discrimination*

Conditions	Sentence pairs within trials		English gloss	Response
Intonation	ta^1 zai^4 jia^1 (A/S)	ta^1 zai^4 jia^1 (A/Q)	He's at home	Different
	ta^1 lai^2 wan^2 (A/S)	ta^1 lai^2 wan^2 (H/S)	He's visiting	Same
	gua^1 feng^1 le (S/S)	gua^1 feng^1 le (S/Q)	The wind's blowing	Different
	ta^1 zai^4 pao^3 bu^4 (A/Q)	ta^1 zai^4 pao^3 bu^4 (S/Q)	He's running	Same
	xia^4 yu^3 le (H/Q)	xia^4 yu^3 le (H/S)	It's raining	Different
Emotion	ta^1 zai^4 jia^1 (A/S)	ta^1 zai^4 jia^1 (A/Q)	He's at home	Same
	ta^1 lai^2 wan^2 (A/S)	ta^1 lai^2 wan^2 (H/S)	He's visiting	Different
	gua^1 feng^1 le (S/S)	gua^1 feng^1 le (S/Q)	The wind's blowing	Same
	ta^1 zai^4 pao^3 bu^4 (A/Q)	ta^1 zai^4 pao^3 bu^4 (S/Q)	He's running	Different
	xia^4 yu^3 le (H/Q)	xia^4 yu^3 le (H/S)	It's raining	Same

* Emotion/Intonation: A = angry; H = happy; S = sad / S = statement; Q = question.

to compare directly focused attention of intonation to that of emotion.

All stimuli were digitally edited to have equal maximum energy level in dB SPL. Auditory stimuli were presented binaurally using a computer playback system (E-Prime) and a pneumatic-based audio system (Avotec). The plastic sound conduction tubes were threaded through tightly occlusive foam eartips inside the earmuffs that attenuated the average sound pressure level of the continuous scanner noise by ~30 dB. Average intensity of all experimental stimuli was 92 dB SPL, and was matched across all conditions. Average intensity of the scanner noise was 80 dB sound pressure level after attenuation by the earmuffs.

A scanning sequence consisted of two tasks presented in blocked format (32 sec) alternating with 16-sec rest periods. Each block contained nine trials. The order of scanning sequences and trials within blocks was randomized for each subject. Instructions were delivered to subjects in their native language via headphones during rest periods immediately preceding each task: "listen" for passive listening to speech stimuli (S); "intonation" for same–different judgments on Chinese intonation (I); and "emotion" for same–different judgments on Chinese emotion (E). Average trial duration was about 3.5 sec including a response interval of 1,800 msec. Average sentence duration was 720 msec.

Accuracy/reaction time and subjective ratings of task difficulty were used to measure task performance online and offline, respectively. Each task was self-rated by listeners on a 1- to 5-point graded scale of difficulty (1 = easy, 3 = medium, 5 = hard). Prior to scanning, subjects were trained to a high level of accuracy using stimuli different from those used during the scanning runs: emotion (Chinese, $M = 0.98$; En-

glish, $M = 0.97$); intonation (Chinese, $M = 0.94$; English, $M = 0.82$).

Imaging protocol

Scanning was done on a 1.5T Signa GE LX Horizon scanner (Waukesha, WI) equipped with birdcage transmit-receive radio frequency head coils. Each of three 200-volume echo-planar imaging (EPI) series was begun with a rest interval consisting of 8 baseline volumes (16 sec), followed by 184 volumes during which the two comparison conditions (32 sec) alternated with intervening 16-sec rest intervals, and ended with a rest interval of 8 baseline volumes (16 sec). Gradient-echo EPI images were acquired with the following parameters: repetition time/echo time (TR/TE) 2 sec/50 msec; matrix 64 × 64; flip angle (FA) 90 degrees; 24 × 24 cm field-of-view (FOV); receiver bandwidth 125 kHz. Fifteen 7-mm–thick axial slices with a 2-mm interslice gap were required to image the entire brain.

Subjects were scanned with eyes closed and room lights dimmed. The effects of head motion were minimized using a head–neck pad and dental bite bar. Data for 2 of 22 subjects were excluded because maximal peak-to-peak displacement exceeded 0.15 mm [Jiang et al., 1995]. 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 3-D Spoiled-Grass (3-D SPGR) sequence (slice thickness 1.1–1.2 mm; TR/TE 35/8 msec; 1 excitation; 30 degrees 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 16-sec blocks, I in 32-sec blocks, and E in 32-sec blocks. The reference function of interest (state 1 – state 2) was defined as:

$$reference = 1 \quad i = [12,27],[60,75],[108,123],[156,171]$$

$$= -1 \, i = [36,51],[84,99],[132,147],[180,195]$$

where i is the index of the ith-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], summed pixel-by-pixel, combined into within-group activation maps, and displayed on anatomic images from a representative subject. Stereotaxic location of activation peaks and extent of activation were identified by drawing regions of interest around activation foci. For within-group comparisons, t-statistic thresholds (one-tailed, uncorrected) were $t(\alpha) = 6.3$, $P < 1.4 \times 10^{-10}$ for scanning runs involving a passive speech baseline (I vs. S, E vs. S); $t(\alpha) = 3.3$, $P < 4.5 \times 10^{-4}$ for scanning runs involving a direct comparison between intonation and emotion (I vs. E, E vs. I). For between-group comparisons, t-statistic thresholds were $t(\alpha) = 5.6$, $P < 1.2 \times 10^{-8}$ (I vs. S, E vs. S) and $t(\alpha) = 3.4$, $P < 4.1 \times 10^{-4}$ (I vs. E, E vs. I).

RESULTS

Task performance

A two-way analysis of variance on response accuracy revealed a significant two-way interaction be-tween task and group [$F(1,36) = 7.17$, $P < 0.0111$]. Tukey multiple comparisons showed that the Chinese (96%) was more accurate than the English group (80%) at judging Chinese intonation. Subjective ratings of task difficulty similarly showed a two-way interaction between task and group [$F(1,36) = 18.30$, $P < 0.0001$]. The intonation task was easier for the Chinese (1.6) than for the English group (3.4). Response times were longer for the Chinese (I = 590 msec; E = 527 msec) than for the English group (I = 450 msec; E = 376 msec) across tasks [$F(1,36) = 15.49$, $P < 0.0004$].

Within-group comparisons

A comparison of intonation judgments of Chinese speech relative to passive listening (I vs. S) yielded frontal, parietal, temporal activation bilaterally for both Chinese and English groups (Table II). Frontal activation was extensive in the inferior prefrontal cortex encompassing the middle and inferior frontal gyri along the inferior frontal sulcus, and extending medially into the anterior insula. Parietal foci were in the supramarginal gyrus near the intraparietal sulcus, though activation was more extensive in the LH compared to the RH for the Chinese group only. Temporal foci were in the posterior superior temporal gyrus near the superior temporal sulcus (STS). Activation of the medial frontal cortex including the supplementary motor area (SMA) and pre-SMA regions was observed in both groups.

A comparison of emotion judgments of Chinese speech relative to passive listening (E vs. S) yielded activation patterns similar to those of I vs. S in frontal and parietal regions for both groups (Table II). Temporal lobe activity was restricted to the RH for the English group.

A direct comparison of intonation and emotion (I vs. E) revealed left-sided activation foci in posterior prefrontal, inferior parietal, and occipitotemporal regions for the Chinese group only (Table II). Bilateral activation of frontopolar and anterior insular regions was observed in the Chinese and English groups, respectively. Activation in posterior prefrontal areas was restricted to the RH for the English group. The reverse comparison (E vs. I) revealed activity in the right middle frontal gyrus for the Chinese group only. Areas of increased activation were observed bilaterally in the fusiform and inferior occipital gyri, as well as in the right angular gyrus and the left posterior cingulate for the English group.

TABLE II. Significant activation foci for within-group comparisons*

Region	Chinese						English					
	BA	x	y	z	Peak t value	Extent	BA	x	y	z	Peak t value	Extent
Intonation vs. passive-listening-to-speech (I-S)												
Frontal												
L posterior MFG	9/44	−43	15	29	20.8	33.50	6/9	−37	4	34	17.1	29.93
R posterior MFG	9/44	46	15	31	18.4	22.35	9/44	50	15	28	19.5	40.03
M medial frontal gyrus	8	0	16	48	16.9	9.48	8	4	17	47	16.7	12.08
Parietal												
L IPS	40/7	−35	−42	42	15.3	19.78	7	−29	−52	39	12.2	5.04
R IPS	7	36	−52	46	11.0	5.59	7/40	36	−59	47	9.5	4.46
Temporal												
L posterior STG/STS	22	−52	−45	3	10.00	1.89	22	−48	−47	12	9.00	1.45
R posterior STG/STS	22	48	−25	3	8.5	1.81	22	47	−34	7	10.6	3.29
Occipital												
M calcarine sulcus	17/18	−2	−82	9	13.1	13.71	17	0	−82	7	7.5	0.26
Emotion vs. passive-listening-to-speech (E-S)												
Frontal												
L posterior MFG	9/44	−43	20	28	18.1	22.85	6/9	−36	6	33	13.6	11.55
R posterior MFG	9/44	45	15	29	18.1	19.10	9/44	50	17	30	15.5	26.26
M medial frontal gyrus	8	−1	16	48	17.3	6.62	8	1	13	49	13.4	6.07
Parietal												
L IPS	7/40	−34	−44	43	13.8	11.40	7	−28	−53	38	10.5	2.43
R IPS	7	36	−53	44	10.9	1.51	7/40	34	−56	47	9.7	4.24
Temporal												
L posterior STG/STS	22	−55	−47	5	8.5	0.68						
R posterior STG/STS	22	48	−26	3	11.8	2.78	22	47	−36	9	7.9	0.90
Occipital												
M calcarine sulcus							17	6	−78	13	7.2	0.37
Intonation vs. emotion (I-E)												
Frontal												
L posterior MFG	6	−29	0	51	5.0	4.28						
L anterior MFG	9/10	−26	47	24	4.7	2.29						
R anterior MFG	10	33	53	22	4.3	0.66						
R posterior MFG							9	52	19	36	4.8	2.10
L anterior INS								−42	14	3	5.1	1.63
R anterior INS								42	19	7	4.5	0.95
R medial frontal gyrus	6	9	3	63	4.5	1.15	8	9	32	40	4.0	0.69
Parietal												
L inferior parietal lobule	40/2	−58	−27	34	4.1	0.53						
Occipitotemporal												
L fusiform gyrus	19/37	−46	−67	3	4.0	0.36						
Emotion vs. intonation (E-I)												
Frontal												
R anterior MFG	46/45	46	34	17	5.3	0.86						
Temporal												
R posterior STS/MTG							37/39	54	−60	16	4.1	0.61
Occipitotemporal												
L IOG, fusiform gyrus							19/37	−42	−70	7	4.2	0.39
R IOG, fusiform gyrus							19/37	25	−57	0	5.2	3.57
Occipital												
L inferior occipital gyrus							18	−21	−89	10	5.3	3.71
R inferior occipital gyrus							18	35	−82	0	5.1	4.94
Other												
L posterior cingulate							30/31	−10	−63	12	4.1	0.56

* M = medial; IFG = inferior frontal gyrus, IOG = inferior occipital gyrus, IPS = intraparietal sulcus, LG = lingual gyrus, MFG = middle frontal gyrus, MOG = middle occipital gyrus, MTG = middle temporal gyrus, STG = superior temporal gyrus, STS = superior temporal 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 (> 0.10 ml) refers to the size of activation (volume size > threshold). Designation of Brodmann's areas (BA) are referenced to this atlas, and are approximate only.

TABLE III. Significant activation foci for between-group comparisons

| Region | Chinese–English | | | | | |
	BA	*x*	*y*	*z*	Peak *t* value	Extent
I-S						
Frontal						
L posterior MFG	6/9	−36	10	42	10.0	5.94
Parietal						
L IPS	7/40	−30	−53	56	8.8	3.23
L precuneus	7	−8	−62	46	7.2	0.37
Occipital						
R cuneus	18	11	−84	17	7.3	0.42
Other						
R posterior cingulate	30/31	14	−60	18	9.4	2.57
E-S						
Frontal						
L posterior MFG	6	−34	9	43	8.8	3.74
Parietal						
L IPS	7/40	−34	−51	53	6.7	0.78
I-E						
Frontal						
L MFG	6	−21	11	50	4.5	0.77
L precentral sulcus	6	−50	−2	37	4.4	0.21
R MFG	6	24	11	51	4.2	0.26
R anterior MFG	10	27	51	16	4.5	0.41
Parietal						
L inferior parietal lobule	40/2	−51	−22	44	4.4	0.59
Occipitotemporal						
L MTG/MOG	37/19	−46	−67	5	5.2	0.81
R MTG/MOG	37/19	44	−71	5	4.4	0.80
Occipital						
L LG	18	−24	−76	−8	6.0	6.64
R LG	19/18	24	−71	−4	5.5	9.36
E-I						
Frontal						
R posterior MFG	9/44	51	19	33	4.5	0.49
R anterior MFG/IFG	10/46	49	34	17	5.6	2.80

Between-group comparisons

A comparison between Chinese and English groups of intonation and emotion relative to passive listening revealed significant activation of left posterior frontal and inferior parietal regions (I vs. S; E vs. S) (Fig. 1, Table III).

A comparison of intonation relative to emotion (I vs. E) between groups yielded bilateral activity in posterior prefrontal cortex, occipitotemporal, and occipital areas, left-sided activity in the inferior parietal lobule, and right-sided activity in the frontopolar region for the Chinese group (Fig. 2, Table III). In the reverse comparison (E vs. I), the Chinese group showed activation in anterior and posterior prefrontal areas of the RH.

DISCUSSION

Our major findings demonstrate that the RH plays an instrumental role in assessing the emotional significance of prosodic cues, and moreover, that at least some aspects of emotional prosody are dissociable from linguistic prosody. The crucial comparison between the Chinese and English groups, E vs. I, reveals activity circumscribed to anterior and posterior prefrontal regions in the RH. This laterality effect cannot be ascribed to the size of the temporal or prosodic domain since our sentence stimuli were identical in duration and hierarchical structure across tasks [see Gandour et al., in press]. Moreover, subjects had to integrate low-level pitch and duration information
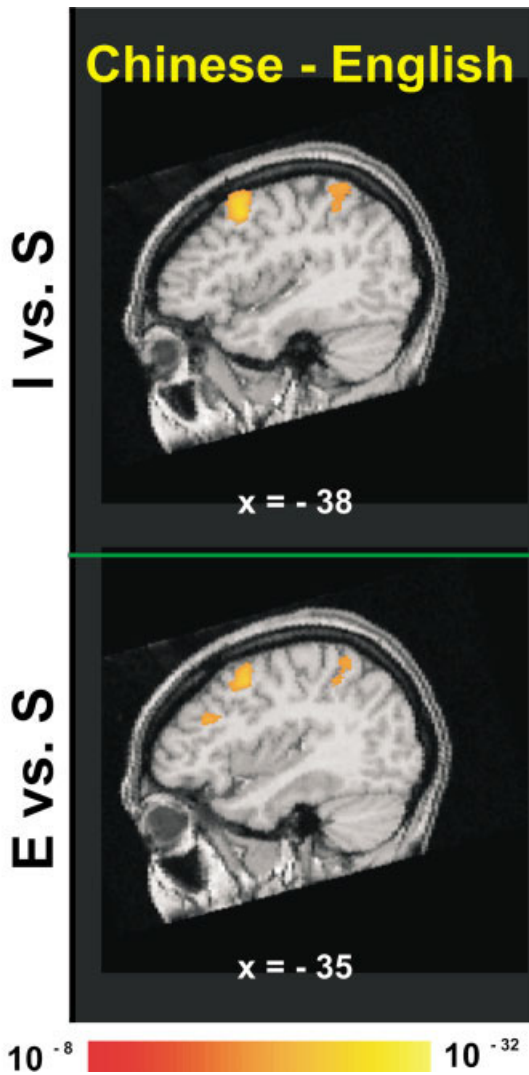
Figure 1.

Averaged fMRI activation maps obtained from direct comparison of intonation (**top**) and emotion (**bottom**) judgments relative to a passive listening baseline between the two language groups (Chinese minus English). Both panels show a left sagittal section through stereotaxic space superimposed onto a representative brain anatomy. Stereotaxic coordinates (mm) are derived from the human brain atlas of Talairach and Tournoux [1988]. I = discrimination judgments of intonation; E = discrimination judgments of emotion; S = passive listening to Chinese speech. Both panels show activation foci in left-sided frontal and parietal regions.

Figure 2.

Averaged fMRI activation maps obtained from direct comparison of intonation (**top**) and emotion (**bottom**) judgments relative to each other between the two language groups (Chinese minus English). Top and bottom, respectively, show axial and right sagittal sections through stereotaxic space superimposed onto a representative brain anatomy. Stereotaxic coordinates (mm) are derived from the human brain atlas of Talairach and Tournoux [1988]. I = discrimination judgments of intonation; E = discrimination judgments of emotion. Top: Activation foci bilaterally in frontal regions and in the left inferior parietal lobule. Bottom: In contrast, shown is activation in right frontal areas only.

throughout the entire sentence in order to make their discrimination judgments. Pitch and duration cues are associated with both linguistic and affective prosody in Chinese. Therefore, it is unlikely that the differential involvement of the two hemispheres can be attributed to hemispheric specialization for low-level auditory cu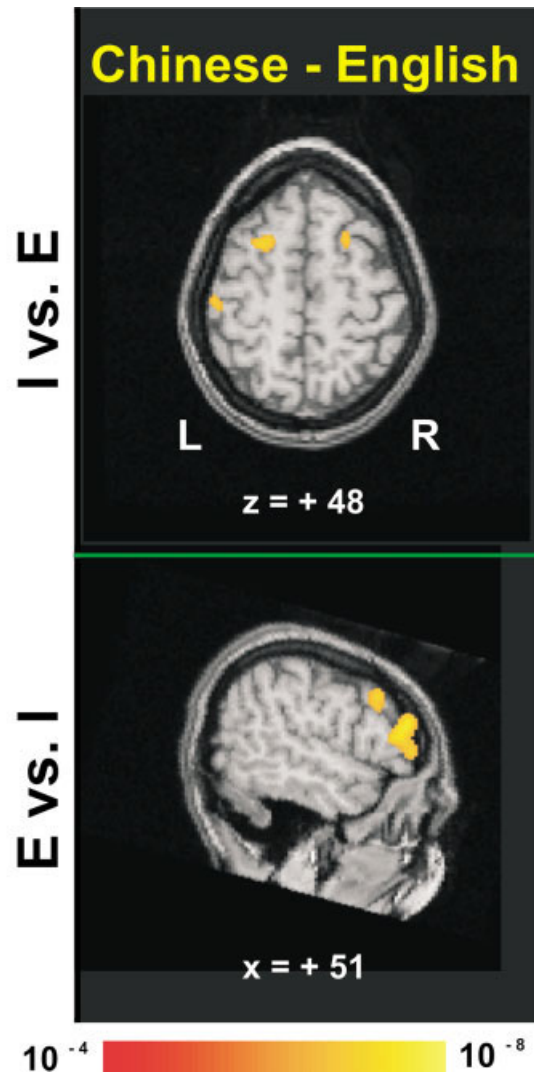es [see Buchanan et al., 2000]. Neither is the lateral-ization of emotion perception to the RH due to a larger-sized prosodic domain. In this study, both emotion and intonation are sentence-level constituents. Instead, it appears that attention-modulated, top-down effects of functional attributes of prosodic stimuli (affective vs. linguistic) may influence the direction and extent of hemispheric specialization.

Even though emotional states are language-independent, they co-occur with the verbal content of a particular language. Thus, it is not surprising that within-group comparisons of intonation and emotion relative to passive listening (I vs. S, E vs. S) elicit bilateral activation in frontotemporoparietal regions for the Chinese group. The effects of language-dependent, top-down mechanisms on speech perception emerge in within-group comparisons of intonation relative to emotion (I vs. E, E vs. I). In the Chinese group, activation foci occur predominantly in the LH for I vs. E, whereas a single focus occurs in the right dorsolateral prefrontal cortex (DLPFC) for E vs. I. In contrast, for the English group, except for activity in the supplementary motor area, different, non-overlapping regions were called into play in the I vs. E and E vs. I comparisons. Especially noteworthy is the bilateral activity in anterior insular, occipitotemporal, and occipital areas, and right-sided activation in posterior frontal and temporal areas. The between-group comparisons (Chinese minus English) corroborate the effects of language-dependent top-down mechanisms. For both I vs. S and E vs. S, the between-group maps show that frontoparietal activity is circumscribed to the LH only. For I vs. E, inferior parietal activity is confined to the LH; for E vs. I, prefrontal activity is confined to the RH.

Taken together, our findings indicate that long-term memory of verbal and emotional auditory information shares a common neuronal system comprising subregions in frontal, parietal, temporal, and occipital areas [see Rama et al., 2001]. The novel results in this study (Chinese: I vs. E & E vs. I; Chinese minus English: I vs. E & E vs. I) demonstrate that activation of frontoparietal subregions are preferentially-activated in the LH or RH depending on functionally-relevant properties of the complex auditory signal. Moreover, our results are consistent with the notion that inferior frontal and parietal regions mediate rehearsal and storage components of verbal working memory, respectively [Awh et al., 1996; Smith and Jonides, 1999], as well as the DLPFC and the region in and around the intraparietal sulcus that appear to mediate the manipulation and maintenance of information in short-term storage [D'Esposito et al., 2000; Newman et al., 2002].

It remains a matter of controversy what the specific contributions are of the left and right anterior DLPFC during memory processing. The traditional view is that the RH is involved in spatial working memory. But the contents of working memory elicit bilateral activation of DLPFC, with a RH preference, regardless of whether the stimuli are spatial or nonspatial [D'Esposito et al., 1999]. It has been suggested that the right DLPFC integrates and maintains information [Prabhakaran et al., 2000] or plays a strategic planning role in memory processing [Newman et al., 2002]. The left DLPFC, on the other hand, has been implicated in representing and maintaining task demands needed for top-down control [MacDonald et al., 2000]. In the current study, emotion and intonation carry the same task demands, involve top-down control, but differ in their functional significance. Our findings suggest that the right as well as left DLPFC may be engaged in maintaining task demands needed for top-down control. Moreover, hemispheric lateralization of the DLPFC appears to vary depending on the nature of the stimulus (e.g., linguistic vs. affective).

In conclusion, it is the differential brain activation patterns between language groups that support the view that functionally relevant properties of complex auditory stimuli are critical in determining which neural mechanisms are engaged in the perception of speech prosody. These findings also support the view that functionally relevant parameters of the speech signal are not encoded wholly on the basis of their complex acoustic properties. Rather, it is likely that functional specialization emerges at multiple levels and interacts in complex ways [Zatorre et al., 2002]. Speech perception itself is likely to emerge from an interaction between low-level perceptual input processes and higher-level, language-specific mechanisms.

ACKNOWLEDGMENTS

REFERENCES

Awh E, Jonides J, Smith EE, Schumacher EH, Koeppe RA, Katz S (1996): Dissociation of storage and rehearsal in verbal working memory. Psycholog Sci 7:25–31.

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.

Baum S, Pell M (1999): The neural bases of prosody: Insights from lesion studies and neuroimaging. Aphasiology 13:581–608.

Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K, Jancke L (2000): Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Cogn Brain Res 9:227–238.

D'Esposito M, Postle BR, Ballard D, Lease J (1999): Maintenance versus manipulation of information held in working memory: An event-related fMRI study. Brain Cogn 41:66–86.

D'Esposito M, Postle BR, Rypma B (2000): Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. Exp Brain Res 133:3–11.

Edmondson JA, Chan JL, Seibert GB, Ross ED (1987): The effect of right-brain damage on acoustical measures of affective prosody in Taiwanese patients. J Phonet 15:219–233.

Gandour J, Larsen J, Dechongkit S, Ponglorpisit S, Khunadorn F (1995): Speech prosody in affective contexts in Thai patients with right hemisphere lesions. Brain Lang 51:422–443.

Gandour J, Dzemidzic M, Wong D, Lowe M, Tong Y, Hsieh L, Satthamnuwong N, Lurito J. Temporal integration of speech prosody is shaped by language experience: an fMRI study. Brain Lang (in press).

George MS, Parekh PI, Rosinsky N, Ketter TA, Kimbrell TA, Heilman KM, Herscovitch P, Post RM (1996): Understanding emotional prosody activates right hemisphere regions. Arch Neurol 53:665–670.

Ho AT (1977): Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative. Phonetica 34:446–457.

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.

Hughes CP, Chan JL, Su MS (1983): Aprosodia in Chinese patients with right cerebral hemisphere lesions. Arch Neurol 40:732–736.

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. Hum Brain Mapp 3:224–235.

Lowe M, Sorenson J (1997): Quantitative comparison of functional contrast from BOLD-weighted spin-echo and gradient-echoplanar imaging at 1.5 Tesla and $H_2{}^{15}O$ PET in the whole brain. Magn Reson Med 37:723–729.

Lowe MJ, Russell DP (1999): Treatment of baseline drifts in fMRI time series analysis. J Comput Assist Tomogr 23:463–473.

MacDonald AW, 3rd, Cohen JD, Stenger VA, Carter CS (2000): Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 288:1835–1838.

Newman SD, Just MA, Carpenter PA (2002): The synchronization of the human cortical working memory network. Neuroimage 15:810–822.

Oldfield RC (1971): The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Pell MD (1998): Recognition of prosody following unilateral brain lesion: influence of functional and structural attributes of prosodic contours. Neuropsychologia 36:701–715.

Prabhakaran V, Narayanan K, Zhao Z, Gabrieli JD (2000): Integration of diverse information in working memory within the frontal lobe. Nat Neurosci 3:85–90.

Rama P, Martinkauppi S, Linnankoski I, Koivisto J, Aronen HJ, Carlson S (2001): Working memory of identification of emotional vocal expressions: an fMRI study. Neuroimage 13:1090–1101.

Ross ED (2000): Affective prosody and the aprosodias. In: Mesulam M, editor. Principles of behavioral and cognitive neurology. New York: Oxford University Press. p 316–331.

Sackheim H, Greenberg M, Weiman A, Gur R, Hungerbuhler J, Geschwind N (1982): Hemispheric asymmetry in the expression of positive and negative emotions: Neurological evidence. Arch Neurol 39:210–218.

Shen X-N (1990): The prosody of Mandarin Chinese. Berkeley, CA: University of California Press.

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: 3-dimensional proportional system: an approach to cerebral imaging. New York: Thieme Medical Publishers. viii, 122.

Wildgruber D, Pihan H, Ackermann H, Erb M, Grodd W (2002): Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. Neuroimage 15:856–869.

Zatorre RJ, Belin P, Penhune VB (2002): Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37–46.