

The Role of the Insular Cortex in Pitch Pattern Perception: The Effect of Linguistic Contexts

Patrick C. M. Wong,¹ Lawrence M. Parsons,² Michael Martinez,² and Randy L. Diehl¹

¹Department of Psychology, University of Texas, Austin, Texas 78712, and ²Research Imaging Center, University of Texas Health Science Center, San Antonio, Texas 78284

Auditory pitch patterns are significant ecological features to which nervous systems have exquisitely adapted. Pitch patterns are found embedded in many contexts, enabling different information-processing goals. Do the psychological functions of pitch patterns determine the neural mechanisms supporting their perception, or do all pitch patterns, regardless of function, engage the same mechanisms? This issue is pursued in the present study by using ¹⁵O-water positron emission tomography to study brain activations when two subject groups discriminate pitch patterns in their respective native languages, one of which is a tonal language and the other of which is not. In a tonal language, pitch patterns signal lexical meaning. Native Mandarin-speaking and English-speaking listeners discriminated pitch patterns embedded in Mandarin and English words and also passively listened to the same stimuli. When Mandarin listeners discriminated pitch embedded in Mandarin lexical tones, the left anterior insular cortex was the most active. When they discriminated pitch patterns embedded in English words, the homologous area in the right hemisphere activated as it did in English-speaking listeners discriminating pitch patterns embedded in either Mandarin or English words. These results support the view that neural responses to physical acoustic stimuli depend on the function of those stimuli and implicate anterior insular cortex in auditory processing, with the left insular cortex especially responsive to linguistic stimuli.

Key words: language; speech; pitch perception; anterior insula; PET; prosody

Introduction

In addition to being a major component of music and vocalizations in many species, pitch conveys information about several aspects of spoken human language, including speaker identity, affect, intonation, phonemic stress, and word meaning. The present study evaluated the following two hypotheses on the neural basis of human pitch perception. The “functional” hypothesis proposes that the psychological function of fundamental frequency (F0) patterns determines which neural mechanisms are engaged during pitch perception. The “acoustic” hypothesis proposes that, regardless of psychological functions, all F0 patterns elicit responses from the same neural mechanisms. Empirical support is available for both functional (Ross et al., 1992) and acoustic hypotheses (Zatorre and Belin, 2001; Warrier and Zatorre, 2004) (for review, see Wong, 2002). In the present study, we test these hypotheses by examining the perception of pitch patterns that are used to contrast word meaning. Such pitch patterns, called lexical tones, are present in most languages of the world (Fromkin, 2000). For example, in Mandarin Chinese,

the syllable /ma/ spoken in a high pitch means “mother,” but the same syllable spoken with a falling pitch pattern means “to scold.”

Lexical tone perception research has mainly focused on psychological processes (Xu, 1994; Wong and Diehl, 2003) and gross hemispheric specialization (Gandour et al., 1997). Recently, several neuroimaging studies have been conducted (Gandour et al., 2000) and reported activations in left inferior frontal regions. In contrast, nonlexical pitch processing studies (Zatorre et al., 1992) found homologous activations in the right hemisphere. Together, these studies provide evidence for the functional hypothesis, because the brain area engaged is determined by the psychological function of pitch.

However, all previous imaging studies of lexical tone perception possess a potentially serious confound of lexical tone processing and semantic processing. Specifically, the lexical tones were discriminated by native speakers of the tone language and by a control group unfamiliar with that language. Thus, for the former group but not the latter, the F0 patterns were lexically relevant and were embedded in meaningful words. This leaves open the possibility that the left hemisphere responses were attributable to semantic processing rather than to the perception of pitch as such.

The present study uses positron emission tomography (PET), an optimal acoustic environment for auditory studies, to compare native Mandarin- and English-speaking listeners discriminating pitch patterns embedded in Mandarin and English word pairs and passively listening to the same sets of stimuli. This

Received Dec. 1, 2003; revised Sept. 1, 2004; accepted Sept. 1, 2004.

This work was supported by National Science Foundation Grant BCS-9986246 (R.D., P.W.), National Institutes of Health Grant R01 DC00427-13, 14 (R.D.), and a grant from the Research Imaging Center (L.P.). We thank Jack Gandour, Ani Patel, and Kai Alter for their comments on a previous version of this manuscript.

Correspondence should be addressed to Patrick C. M. Wong, Speech Research Laboratory, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208. E-mail: pwong@northwestern.edu.

L. M. Parsons's and M. Martinez's present address: Department of Psychology, University of Sheffield, S10 2TP, UK.

DOI:10.1523/JNEUROSCI.2225-04.2004

Copyright © 2004 Society for Neuroscience 0270-6474/04/249153-08\$15.00/0

design permitted a comparison of lexical tone processing and nonlexical pitch processing. Previous related studies led to the hypothesis that the frontal operculum and anterior insula would be intensely activated by these tasks (Zatorre et al., 1992). The functional hypothesis predicts the activation of those areas in the left hemisphere (relative to passive listening) in listeners for whom the pitch patterns are lexically meaningful (i.e., in the condition where Mandarin listeners discriminate pitch embedded in Mandarin words). The acoustic hypothesis predicts that right hemisphere homologues should be active for both groups of listeners regardless of the linguistic status of the stimuli.

Materials and Methods

Subjects. The subjects were seven adult male native Mandarin speakers who also speak English and seven adult male native speakers of American English who had no previous exposure to Mandarin. All Mandarin-speaking subjects ranked Mandarin as their dominant language relative to English and spoke mostly Mandarin (or in some cases, Mandarin plus another Chinese dialect) in their childhood. The Mandarin speakers ranged from 18 to 32 years of age, with a mean of 25 and SD of 5.1; the English speakers ranged from 18 to 27 years of age, with a mean of 21 and SD of 3.2. The self-rated proficiency in Mandarin for all of the Mandarin-speaking subjects was 10, on a scale of 1 (not proficient) to 10 (extremely proficient); their self-rated proficiency in English ranged from 6 to 9, with a mean of 7 and SD of 1.2. All subjects were right handed, musically untrained, and had normal hearing. Right-handedness was assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). Musical ability was assessed by a simple music-reading task, which involves sight singing or playing, in relative pitch, a six-note melody in C major with no accidentals (sharps or flats). A traditional Western music score and a conversion of the score into numerals (e.g., 1, C; 3, E) were provided to the subjects. All subjects failed this music test.

Stimuli. Mandarin and English stimulus sets were resynthesized from naturally produced utterances of two male Mandarin speakers and two male English speakers. F0s of these spoken words did not differ significantly between the two language groups. Resynthesis was performed with the Analysis/Synthesis Laboratory module of Computerized Speech Laboratory (Kay Elemetrics, Lincoln Park, NJ). Mandarin speakers were asked to produce a list of Mandarin words, all of which carried the high level tone (Tone 1). English speakers were asked to produce a list of English words in a high pitch. For every word produced, three speech tokens were resynthesized with pitch patterns resembling the high level tone (Tone 1), the rising tone (Tone 2), and the falling tone (Tone 4) of Mandarin. Tone 3, the dipping tone, was not included because it has been shown perceptually to be the most confusable tone both to native Mandarin speakers (Chuang et al., 1972) and to second-language learners of Mandarin (Kiriloff, 1969). Pitch contours were interpolated linearly through the voiced portion of each stimulus. The starting and ending pitch points for Tone 1 were identical. This value was the mean F0 of the list of words produced by each speaker. The ending pitch point of Tone 2 was the same as Tone 1, and the starting pitch point of Tone 2 was 26% lower than its ending point. The starting pitch point of Tone 4 was 10% higher than Tone 1 and fell by 82%. These pitch contours were modeled on the values obtained by Shih (1988). Three native informants of each language group transcribed the resynthesized speech tokens of their native language with an accuracy of 98% for the English set and 96% for the Mandarin set. The difference in transcription accuracy was not statistically significant. Mandarin informants were required to write the Chinese character, not the phonetic transcription (pin-yin) associated with the word they heard. For the Mandarin speech token set, only tokens that are actual Mandarin words were included in this intelligibility task and in the actual stimulus set. For example, /min1/ was not included because it is not an actual Chinese word (although it can be transcribed and pronounced by Mandarin speakers based on its pin-yin).

The stimuli on each trial consisted of pairs of resynthesized words originally spoken by the same speaker. For example, the /fei2/ and /wei2/ stimuli were based on the /fei1/ and /wei1/ productions of the same speaker. Half of the stimulus pairs had the same tone (or pitch patterns

for the English stimuli) in each word of the pair, and the other half had different tones in each word of the pair. The frequency of occurrence for each tone category was equal, and the order of tone presentation was counterbalanced. Except for the aforementioned acoustic properties, all properties within a word pair (e.g., the vowel) were identical. Stimuli were 72 monosyllables (36 in each language) of ~350 msec in duration each. Because of ceiling effects on discrimination accuracy observed in pilot studies, white noise, with amplitude of 2 dB lower than the stimuli, was added to the stimuli to increase the difficulty of the task. The aforementioned intelligibility testing was performed before the addition of noise.

Experimental procedures. There were five conditions, including a rest condition, two passive listening conditions, and two active listening conditions. Except for the rest condition, each condition was repeated twice, and a PET scan was performed for each repetition. The order of presentation of conditions was pseudorandomized for each subject. The four passive listening conditions (two conditions, each presented twice) were always presented before the active listening ones to reduce the possibility of active discrimination of the stimulus pairs by the subjects even in the passive conditions. Anatomical magnetic resonance images (MRIs) were acquired of the brain of each subject for accurate spatial normalization and localization of regional cerebral blood flow (rCBF) foci.

To reduce head motion during scanning, each subject was fitted with a thermal plastic facemask. A computer keyboard, for task responses, was placed within comfortable reach of the subject. All responses were made with the middle and index fingers of the left hand, pressing the 1 and 4 keys on the number pad, respectively. During all conditions, subjects lay supine in the scanning instrument with their eyes closed. All stimuli were presented via Sony Fontopia-Earbud headphones (Sony, Tokyo, Japan).

During the discrimination tasks, subjects judged whether the tone patterns in Mandarin word pairs or the pitch patterns in English word pairs were the same or different and responded by key press. During the passive listening tasks, subjects were instructed to listen to, but not attempt to discriminate, either Mandarin or English word pairs and to press the response keys alternatively after each word pair had been presented. During the rest condition, no stimuli were presented.

During PET trials, subjects began performing the task ~30 sec before the onset of the 40 sec scan. On average, a word pair trial lasted ~1 sec, with an intra-pair interval of 300 msec. A response from the subjects triggered the presentation of the next stimulus pair. If subjects did not respond within 1250 msec, the next stimulus pair was presented. A relatively short inter-stimulus interval was used in an attempt to facilitate attention to the experimental task. Inter-scan interval was ~10 min.

Imaging procedures and analysis. PET scans were performed on a GE 4096 camera, which has a pixel spacing of 2.0 mm, an inter-plane, center-to-center distance of 6.5 mm, 15 scan planes, and a z-axis field of view of 10 cm. Correction for radiation attenuation was made by means of a transmission scan collected before the first scan using a ⁶⁸Ge/⁶⁸Ga pin source. Cerebral blood flow was measured with H₂¹⁵O (half-life, 123 sec) administered as an intravenous bolus of 8–10 ml of saline containing 60 mCi. At the start of a scanning session, an intravenous cannula was inserted into the subject's right forearm for injection of each tracer bolus. A 40 sec scan was triggered as the radioactive tracer was detected in the field of view (the brain) by increases in the coincidence-counting rate. During this scan, the subject was in the rest condition or performed a task in one of the four experimental conditions. Immediately after this, a 50 sec scan was acquired as the subject lay with his eyes closed without performing a task. The latter rest PET images, in which task-related rCBF changes were still occurring in specific brain areas, were combined with the task PET images to enhance detection of relevant activations. A 10 min inter-scan interval was sufficient for isotope decay (5 half-lives) and return to resting state levels of regional blood flow within activated regions. Images were reconstructed using a Hann filter, resulting in images with a spatial resolution of ~7 mm (full-width at half-maximum).

Significant changes in cerebral blood flow indicating neural activity were detected using a regions-of-interest (ROIs)-free image subtraction strategy. High-resolution MRI scans were acquired for each subject. Inter-scan, intra-subject movement was assessed and corrected using the Woods' algorithm (Woods et al., 1993). Semiautomatic registration of

PET to matched, spatially normalized MRI (in Talairach space) was performed using in-house spatial normalization software implementing an algorithm using a nine-parameter, affine transformation (Talairach and Tournoux, 1988; Lancaster et al., 1995).

The data were smoothed with an isotropic $3 \times 3 \times 3$ Gaussian kernel to yield a final image resolution of 8 mm. By use of moderate smoothing, resolution was not lost; control for false positives was provided by the p value criteria used in later stages of our analysis stream. The data were then analyzed using the Fox et al. (1988) algorithm as implemented in Multiple Image Processing Station (MIPS) (Research Imaging Center, University of Texas Health Science Center San Antonio, San Antonio, TX, USA). Intra-subject image averaging was performed within conditions, and the resulting image data were analyzed by an omnibus (whole-brain) test.

For this analysis, local extrema (positive and negative) were identified within each image using a three-dimensional search algorithm (Mintun et al., 1989). Each set of local extrema data was plotted as a frequency histogram (for visual inspection). Then, a β -2 statistic measuring kurtosis and a β -1 statistic measuring skewness of the histogram of the difference images (change distribution curve) (Fox and Mintun, 1989) were used as omnibus tests to assess overall significance (D'Agostino et al., 1990). Critical values for β statistics were chosen at $p < 0.01$. The β -1 and β -2 tests, which are implemented in the MIPS software (Research Imaging Center, University of Texas Health Science Center at San Antonio, San Antonio, TX) in a manner similar to the use of the gamma-1 and gamma-2 statistics (Fox and Mintun, 1989), improve on the gamma statistic by using a better estimate of the degrees of freedom (Worsley et al., 1992).

The null hypothesis of omnibus significance was rejected, so a *post hoc* (regional) test was done (Fox et al., 1988; Fox and Mintun, 1989). In this algorithm, the pooled variance of all brain voxels was used as the reference for computing significance. This method is distinct from methods that compute the variance at each voxel but is more sensitive (Strother et al., 1997), particularly for small samples, than the voxel-wise variance methods of Friston et al. (1991) and others. In this analysis, a maxima and minima search was conducted to identify local extrema within a search volume of 125 mm^3 (Mintun et al., 1989). Cluster size was determined based on the number of significant, contiguous voxels within the search cube of 125 mm^3 . The statistical parameter images were converted to z -values by dividing each image voxel by the average SD of the null distribution. p values were assigned from the Z distribution. Only Z values > 2.96 ($p < 0.001$) were reported. The critical value threshold for regional effects ($Z > 2.96$; $p < 0.001$) was not raised to correct for multiple comparisons (e.g., the number of image resolution elements). This is because omnibus statistics were established before *post hoc* analysis. The scanning methods used at the Research Imaging Center have been described previously (Raichle et al., 1983; Fox et al., 1985; Fox and Mintun, 1989; Lancaster et al., 1995).

Gross anatomical labels were applied to the detected local maxima using a volume occupancy-based, anatomical-labeling strategy as implemented in the Talairach Daemon (Lancaster et al., 2000), except for activations in cerebellum, which were labeled manually with reference to an atlas of the cerebellum (Schmahmann et al., 1999).

Anatomical MRI scans were performed on an Elscint 1.9 T Prestige system. The scans used three-dimensional gradient recalled acquisitions in the steady state (3D GRASS), with a repetition time of 33 msec, an echo time of 12 msec, and a flip angle of 60° to obtain a $256 \times 192 \times 192$ volume of data at a spatial resolution of 1 mm^3 .

Results

Behavioral testing

A two-way, repeated measures ANOVA performed on response accuracy (Fig. 1) revealed a main effect for subject group, with Mandarin-speaking subjects showing greater accuracy in discriminating tone and pitch patterns than English-speaking ones ($F_{(1,12)} = 8.29$; $p < 0.05$). A two-way, repeated measures ANOVA performed on response times (Fig. 2) yielded a significant main effect of task ($F_{(1,12)} = 11.14$; $p < 0.01$), with subjects slower on

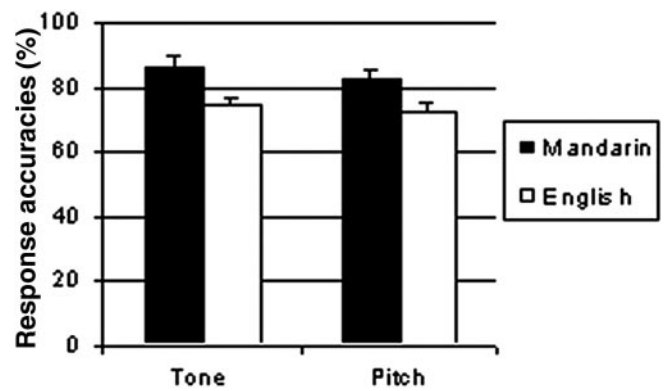


Figure 1. Response accuracies for the discrimination of pitch pattern embedded in Mandarin by Mandarin-speaking and English-speaking subjects.

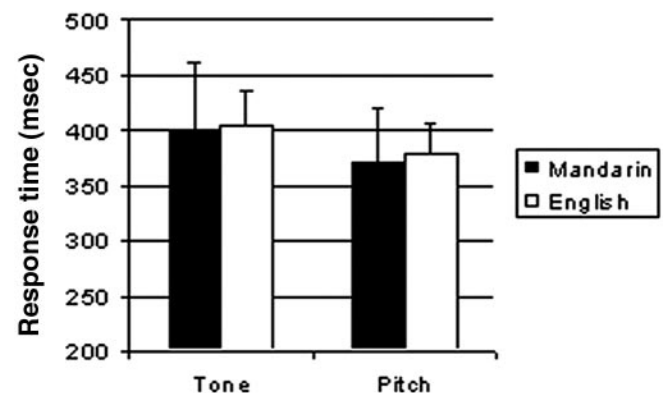


Figure 2. Response times for the discrimination of pitch pattern embedded in Mandarin by Mandarin-speaking and English-speaking subjects.

the tone task than the pitch task. Within subject groups, there appeared to be a speed-accuracy trade off such that both subject groups were faster but less accurate in the pitch task relative to the tone task.

PET results: cerebral blood flow increases

Distinct patterns of activation were observed for each task and each group. Tables 1 and 2 itemize the regions of cerebral blood activity for the tone and pitch tasks relative to their corresponding passive listening tasks. Only z values > 2.96 ($p < 0.001$) and forming contiguous clusters $> 60 \text{ mm}^3$ are reported. In the descriptions, figures, and tables that follow, activations that resulted spuriously from subtractions of deactivations are excluded. Thus, all activations were verified by subtracting the values in the rest condition from those during the relevant active tasks, and false-positive activations caused by subtractions of deactivations were eliminated.

Comparison of the Mandarin tone task to Mandarin passive listening for the Mandarin speakers yielded significant rCBF increases predominately in the left hemisphere. The strongest activation was in the left anterior insula (Fig. 3, left panel). Another very strong activation was present in the left basal ganglia (putamen). Other left hemispheric areas activated were fusiform gyrus [Brodmann's Area (BA) 36, 37], posterior insula, inferior supplementary motor area (SMA; medial BA 6), and dorsolateral premotor cortex (BA 6). There were right hemispheric activations in somatosensory cortex (BA 3) and dorsolateral premotor cortex

Table 1. Active Mandarin tone discrimination versus passive Mandarin listening

Gyrus or region	BA	Coordinates			Z	Score
		x	y	z		
Mandarin speakers						
Left hemisphere						
Anterior insula	45	-27	24	4	5.30	
Putamen		-20	4	4	4.12	
Thalamus		-10	-22	14	4.27	
Inferior fusiform gyrus	36	-44	-48	-18	3.77	
Precentral gyrus	6	-49	-6	40	3.74	
Insula		-32	14	8	4.06	
Fusiform gyrus	37	-34	-60	-12	3.77	
Medial frontal gyrus	6	0	16	44	3.62	
Left cerebellum						
Crus I		-14	-80	-24	3.53	
Culmen (IV)		-4	-46	-4	4.30	
Right hemisphere						
Postcentral gyrus	3	56	-20	38	3.92	
Middle frontal gyrus	6	44	-2	48	3.42	
Right cerebellum						
Declive (vermis) (VI)		6	-70	-12	3.36	
Declive (vermis) (VI)		10	-66	-24	3.59	
English speakers						
Left hemisphere						
Middle frontal gyrus	10	-22	54	-8	4.25	
Anterior cingulate	24	-2	18	22	3.61	
Supramarginal gyrus	40	-38	-40	32	3.56	
Frontal subgyral	6	-20	-12	48	4.94	
Anterior cingulate gyrus	24	-8	0	46	4.45	
Frontal subgyral	2	-30	-22	38	3.36	
Anterior cingulate gyrus	24	-2	12	34	3.18	
Middle temporal gyrus	39	-28	-54	28	3.21	
Superior temporal gyrus	22	-48	6	-4	4.02	
Caudate nucleus		-3	18	13	3.30	
Parietal subgyral	7	-22	-50	39	3.67	
Frontal subgyral	44	-42	6	20	3.27	
Frontal subgyral	6	-34	-4	32	2.98	
Right hemisphere						
Precentral gyrus	6	42	-2	28	4.36	
Insula		38	16	2	5.77	
Middle frontal gyrus	6	24	2	44	5.02	
Frontal subgyral	4	36	-12	38	3.90	
Insula		30	9	-4	4.51	
Inferior frontal gyrus	47	28	20	-8	4.28	
Anterior cingulate gyrus	32	6	14	40	3.67	
Anterior cingulate gyrus	24	20	-14	46	4.07	
Frontal subgyral	2	36	-24	32	3.70	
Midbrain		12	-20	-12	4.42	
Frontal subgyral	3	30	-22	46	4.25	
Anterior cingulate gyrus	32	10	20	36	3.64	

Local maxima in regions demonstrating rCBF increases ($p < 0.001$). Brain atlas coordinates (Talairach and Tournoux, 1998) are in millimeters along left-right (x), anteroposterior (y), and superior-inferior (z) axes.

(BA 6). In addition, there were bilateral activations in vermal or midline regions of cerebellum, mostly in the posterior hemisphere.

The same comparison for the English speakers yielded similar activations in each hemisphere but no cerebellar activity. Of particular note is that English speakers intensely activated anterior insula in the right hemisphere, homologous to that activated by the Mandarin speakers (Fig. 3, right panel). Although activations were strong in both hemispheres when English subjects discriminated Mandarin stimuli, left hemisphere activity was more diffuse, compared with right hemisphere activity, which clustered around the insular and the frontal cortices. In the left hemisphere, there were activations in middle frontal cortex (BA 10), anterior cingulate (BA 24), supramarginal gyrus (BA 40), middle and superior temporal gyrus (BA 39, 22), basal ganglia (caudate nucleus), and somatosensory and premotor frontal cortex (BA 2,

Table 2. Active English pitch discrimination versus passive English listening

Gyrus or region	BA	Coordinates			Z	Score
		x	y	z		
Mandarin speakers						
Left hemisphere						
Anterior cingulate gyrus	32	-5	8	44	4.07	
Precentral gyrus	6	-48	2	38	4.76	
Right hemisphere						
Anterior cingulate gyrus	32	8	14	36	4.27	
Pons		2	-20	-26	4.33	
Medial frontal gyrus	6	4	4	52	5.10	
Insula		24	17	4	3.43	
Extra-nuclear		28	-32	6	3.35	
Inferior frontal gyrus	44/6	42	4	30	3.78	
Midbrain		12	-18	-6	3.89	
Lentiform nucleus		20	-2	0	4.30	
Insula		40	-22	16	3.63	
Precentral gyrus	6	51	-2	32	3.00	
Thalamus		8	-16	9	4.15	
Right cerebellum						
Declive (vermis) (VI)		10	-70	-22	3.58	
Simplex (quadrangular) (VI)		31	-68	-26	4.09	
English speakers						
Left hemisphere						
Anterior cingulate gyrus	24	-12	-8	36	5.09	
Anterior cingulate gyrus	24	-7	8	35	3.61	
Anterior cingulate gyrus	24	-22	-6	44	4.56	
Anterior cingulate gyrus	24	-4	4	44	4.70	
Anterior cingulate gyrus	32	-4	22	40	3.83	
Left cerebellum						
Culmen (V)		-4	-58	0	3.42	
Right hemisphere						
Frontal subgyral	4	26	-10	44	5.76	
Anterior cingulate gyrus	24	24	0	42	5.15	
Inferior parietal gyrus	40	32	-30	36	4.20	
Insula		30	22	0	4.28	
Inferior frontal gyrus	47	30	44	-10	4.98	
Frontal subgyral	44/9	34	14	30	3.89	
Parietal subgyral	40	36	-44	34	3.53	
Inferior frontal gyrus	44/6	46	12	24	4.06	
Superior frontal gyrus	10	28	54	-7	4.23	
Frontal subgyral	2	34	-20	36	3.61	
Midbrain		6	-18	-10	4.34	
Frontal subgyral	46	44	24	22	3.36	
Frontal subgyral	3	26	-24	46	3.44	
Cuneus	18	23	-79	18	4.11	
Precentral gyrus	4	51	-8	30	4.42	
Middle temporal sulcus	39	31	-54	28	4.42	
Medial occipital gyrus	19	32	-72	14	3.11	
Precentral gyrus	6	38	-3	42	3.92	
Anterior cingulate gyrus	32	20	12	38	3.30	
Corpus callosum		4	4	20	3.28	
Caudate		14	14	0	3.72	
Postcentral gyrus	2	42	-18	30	3.36	
Insula		38	12	-2	3.86	

Local maxima in regions demonstrating rCBF increases ($p < 0.001$). Brain atlas coordinates (Talairach and Tournoux, 1998) are in millimeters along left-right (x), anteroposterior (y), and superior-inferior (z) axes.

6, 44). In the right hemisphere, apart from the insula activation described earlier, there was strong activation in dorsolateral premotor frontal cortex (BA 6) as well as considerable activation in other frontal areas, including primary motor (BA 4), inferior frontal (BA 47), and somatosensory cortices (BA 2, 3). In addition, there were right hemispheric activations in anterior cingulate cortex.

Interestingly, when the active English pitch task was contrasted with its passive counterpart, both Mandarin and English speakers showed primarily right hemispheric activations. For the Mandarin speakers, activity was observed in right hemisphere in

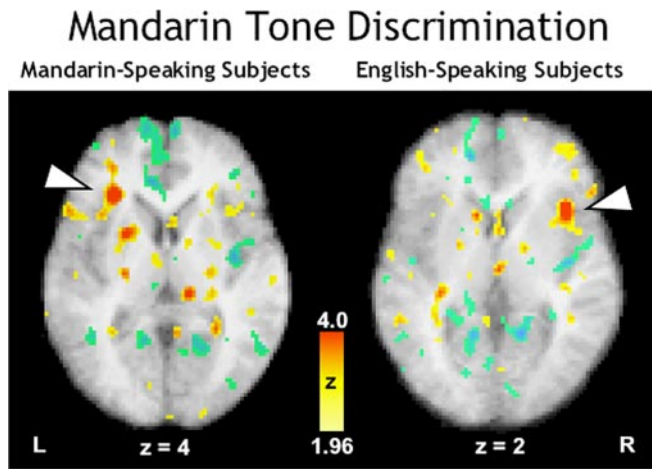


Figure 3. PET-rCBF activations for the discrimination of pitch patterns embedded in Mandarin (in yellow-red Z-score scale) overlaid onto subjects' mean anatomical MR images (grayscale) in transverse planes. L, Left; R, right.

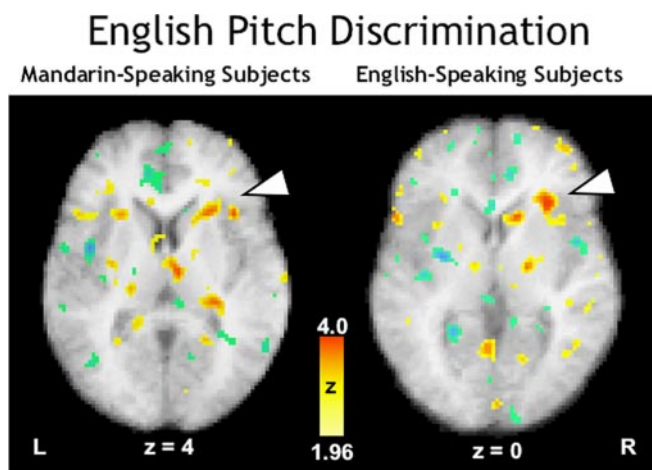


Figure 4. PET-rCBF activations for the discrimination of pitch patterns embedded in Mandarin (in yellow-red Z-score scale) overlaid onto subjects' mean anatomical MR images (grayscale) in transverse planes. L, Left; R, right.

insula (Fig. 4, left panel) in a region approximately homologous (24, 17, 4) to that activated in the left hemisphere during the active Mandarin tone task (−27, 24, 4). In addition, there was right hemispheric activity in SMA (medial BA 6), pons, midbrain, anterior cingulate cortex, inferior frontal gyrus (BA 44/6), posterior insula, dorsolateral frontal cortex (BA 6), lentiform nucleus, and thalamus. In the left hemisphere, there were two foci only, one each in anterior cingulate cortex (BA 32) and dorsolateral premotor frontal cortex (BA 6). The right cerebellum was activated in midline and lateral areas in the posterior hemisphere.

For the English speakers, there was strong activation in right insula (30, 22, 0) near that observed for the Mandarin speakers (Fig. 4, right panel). In addition, there was right hemispheric activation in superior (BA 10) and inferior frontal cortex (BA 47, 44/6, 46) as well as somatosensory (BA 2, 3) and primary motor and premotor cortex (BA 4, 6). Other right areas activated included anterior cingulate (BA 24, 32), prefrontal cortex (BA 44/9), middle temporal cortex (BA 39), insula (posterior to the focus described earlier), cuneus and extrastriate occipital cortex (BA 18, 19), and basal ganglia (caudate nucleus). In the left hemisphere, there were several foci in anterior cingulate (BA 24, 32). There was one activation in the left midline cerebellum.

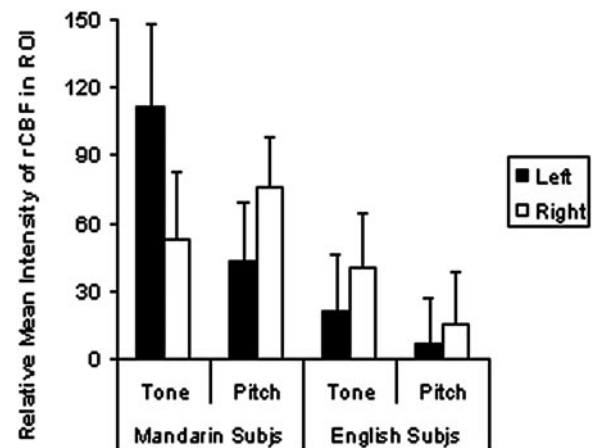


Figure 5. PET-rCBF activations in the anterior insular cortex averaged across subjects (Subjs).

English speakers only, during both Mandarin and English pitch discrimination tasks, exhibited activations in bilateral parietal cortex [left angular gyrus (BA 39), left supramarginal gyrus (BA 40), right inferior parietal regions (BA 39 and 40)] as well as in right visual cortex [medial occipital gyrus (BA 19) and cuneus (BA 18)]. These areas are not typically viewed as related to speech and pitch processing, although activations in these areas have been reported in speech (Zatorre et al., 1992) and melody perception (Zatorre et al., 1994). These activations may be attributable to subjects' use of resources outside the regular speech and pitch processing domains. English-speaking subjects, for whom identifying pitch patterns at the word level is an unfamiliar task, could be using strategies such as visualizing the written forms of the stimuli they heard or mentally drawing the pitch patterns of the stimuli.

To further assess the activity in insular cortex for groups and conditions, a ROI analysis was performed. Mean activity within a ROI of $5 \times 5 \times 5$ voxels (voxel, 2 mm^3) was measured at the extrema point within both the left and right anterior insular cortex in both subject groups for both experimental conditions. A repeated measures ANOVA was performed on the resulting ROI data. The analysis revealed a group \times condition \times hemisphere interaction ($F_{(1, 13)} = 6.01$; $p < 0.05$); the left anterior insula was activated to a greater degree in the tone condition in the Mandarin-speaking subjects (Fig. 5). Other than a marginal main effect of subject group ($F_{(1, 13)} = 4.19$; $p < 0.07$), no other main effects or interactions were found. The group \times condition \times hemisphere interaction remained when the search volume was decreased from 1000 mm^3 ($5 \times 5 \times 5$ voxel ROI) to 216 mm^3 ($3 \times 3 \times 3$ voxel ROI) and 8 mm^3 ($1 \times 1 \times 1$ voxel ROI).

Discussion

These findings indicate that the brain areas engaged during pitch pattern perception depend on the psychological functions for those patterns. Unlike previous neuroimaging studies of lexical tone processing, our subject groups discriminated pitch patterns of words in their respective native languages, and the patterns of brain activations were different depending on the functions of these pitch patterns. When pitch patterns had a lexical function, activations were predominately in the left hemisphere, including areas that have been shown to be active during language tasks. When Mandarin-speaking subjects discriminated the same pitch patterns embedded in English words, activations were primarily in the right hemisphere, including the inferior frontal gyrus,

which has been shown to be active during pitch processing. However, when English-speaking subjects discriminated pitch patterns of English word pairs without a lexical function, the right hemisphere, including the right inferior gyrus, was strongly activated.

These results are inconsistent with an acoustic hypothesis predicting that the same brain areas subserved a particular stimulus category independent of its function. However, these data are not necessarily inconsistent with other hypotheses that propose localized, specialized, acoustic processing. Thus, one hypothesis proposes that left primary auditory cortex is specialized for temporal cues, whereas that in the right hemisphere is specialized for spectral cues (Zatorre and Belin, 2001). Our findings may be consistent with this hypothesis if the activity we observed occurs downstream from such acoustically specialized processing.

Our findings provide clear evidence that there is a region of anterior insula that supports aspects of pitch pattern perception. Furthermore, our results imply that on the left side, this insular region supports pitch pattern perception when the patterns carry lexical information, whereas on the right side, it supports the pitch pattern perception when the patterns are not lexically informative. The insula spans essentially the length of primary language areas in the brain and connects to orbital, frontal opercular, lateral premotor, ventral granular, and superior temporal cortices among other areas (Mesulam and Mufson, 1985; Augustine, 1996). Regions of the insula have been implicated by studies in aspects of language-related functions (for review, see Ardila, 1999), including aphasia (Damasio and Damasio, 1980), dyslexia (Paulesu et al., 1996), affective and nonaffective prosody (Borod, 2000), and auditory agnosia (Habib et al., 1995).

A recent cross-language functional MRI (fMRI) study found bilateral insular activity during phonological and nonphonological speech processing (Jacquemot et al., 2003). Subjects discriminated vowel duration and epenthetic vowels embedded in pseudowords by French and Japanese speakers; vowel duration is phonologically relevant to Japanese but not French speakers, but the reverse is true for epenthetic vowel. This use of pseudowords is likely reflected in the lack of laterality effect seen here with lexical and nonlexical pitch processing in real words. Their results further provide evidence supporting the functional account of pitch perception that emphasizes the importance of context (i.e., lexical tones are most saliently lexical when they are embedded in real words). In an fMRI study of phonetic learning, Golestani and Zatorre (2004) examined the discrimination of native and non-native phonemes as monolingual English-speaking subjects learned to perceive Hindi consonants. For both native and acquired non-native consonant discriminations, activity was detected in regions of bilateral anterior insula slightly lateral to those activated here in the context of linguistic (left) or acoustic-only (right) features. However, because subjects were required to attend to individual phonemes (e.g., /ta/) isolated from lexical or broader linguistic contexts (e.g., an actual word “tea”), it is likely that they closely attended to the acoustic characteristics of the stimuli, because the spectrotemporal features of the stimuli were being varied. Thus, it is unclear whether a pure psychoacoustic strategy was used, or whether the potential lexical–linguistic relevance of these stimuli played any role. The bilateral insular activation (as opposed to a left-lateralized activation) may reflect a combination of these strategies.

The insula is often active in neurological studies of pitch processing and vocalization (Table 3) (for review, see Bamiou et al., 2003). Neuroimaging studies of sentence intonation processing have found activation in the frontal operculum extending to the

Table 3. Insular activation in relevant past studies

	Region	x	y	z
Lexical tone discrimination studies				
Gandour et al. (2000)				
Thai tone — rest (English listeners)	L anterior	−28	21	2
Pitch — rest (Chinese listeners)	R anterior	30	12	4
Gandour et al. (1998)	None			
Hsieh et al. (2001)				
Mandarin tone — pitch (English listeners)	L posterior	−39	−17	−4
Klein et al. (2001)				
Mandarin tone — rest (English listeners)	L anterior	−36	18	3
Other related studies				
Jacquemot et al. (2003)				
Phonological change versus no change (pseudowords)	L	−40	−8	12
	R anterior	40	16	8
Nonphonological change versus no change (pseudowords)	L anterior	−28	20	4
	R anterior	24	24	4
Brown et al. (2004)				
Monopitch singing	L	−38	−2	−4
Melody singing	L anterior	−42	12	2
Harmony singing	L anterior	−42	20	0
Perry et al. (1999)				
Singing — passive pitch perception	L	−42	−1	8
	R posterior	39	−6	−6
Riecker et al. (2000)				
Overt speech	L anterior	−35	18	−4
Overt nonlyrical singing	R anterior	32	16	−6
Rumsey et al. (1997)				
Phonological decision making	L anterior	−32	16	4
Zatorre et al. (1994)				
First/last note judgment — passive melodies	L anterior/45	−31	22	8
	R anterior/45	38	20	5

Local maxima in regions demonstrating rCBF increases ($p < 0.001$). Brain atlas coordinates (Talairach and Tournoux, 1998) are in millimeters along left-right (x), anteroposterior (y), and superior-inferior (z) axes.

anterior insula (Kotz et al., 2003; Meyer et al., 2003). Magnetoencephalography studies have observed a right-lateralized frontal region to be associated with nonlexical intonation processing (Herrmann et al., 2003). An event-related potential study investigating the processing of nonlexical pitch patterns in German words failed to show neural distinctions between the detection of correct and incorrect pitch contours (Friedrich et al., 2001), implying that the lexical status of pitch is instrumental in localizing the associated neural substrates. When subjects process lexical tones, left insula is activated when a linguistic component (whether native or non-native) is present (i.e., words minus rest, or words minus pitch), whereas the right insula is activated when nonlexical pitch is present (i.e., pitch minus rest). These activations have sometimes been attributed to subvocal rehearsal (Paulesu et al., 1996). As discussed later, evidence does not favor a subvocal account of our observed insular activations but an account more consistent with speech perception and phonological processing.

Insular cortex also has a role in overt vocal production. Perry et al. (1999) found that singing a single pitch on a vowel, relative to passive listening to complex tones, activated the insular cortex bilaterally. Their left anterior activation (−42, −1, 8) was linked to the production aspect of singing, and the right side (39, −6, −6) was associated with the self-regulation–pitch perception aspect of

singing. Likewise, Riecker et al. (2000) found that overt speaking and overt nonlyrical singing elicited bilateral insula activity. However, both activations were in the anterior insula ($-35, 18, 4$ for speech; $32, 16, -6$ for singing), unlike the results of the study by Perry et al. (1999). No such activations were observed in covert speaking and singing of the same stimuli. Left anterior insula activity was present when amateur musicians sang either a monotonic pitch, a melody, or a harmonization to a melody (Brown et al., 2004).

A unique feature of the present study is that the stimuli were embedded in noise (to eliminate a behavioral ceiling effect). Noise may increase subjects' tendency to rehearse the stimuli subvocally, and subvocal rehearsal led to the strong insular activations. Several aspects of the results suggest otherwise. Subvocal rehearsal is associated with activation on the lateral area of BA 44 and the inferior and superior aspects of BA 6 (Smith et al., 1998). When our subjects performed the nonlexical pitch tasks, only right insular activations were detected. Because the stimuli included consonants, vowels, and actual words in addition to the pitch patterns, subvocal rehearsal should induce left insular activations in addition to the right activations. Thus, the absence of activation in left insula appears inconsistent with the use of subvocal rehearsal. However, the addition of noise introduced an extra cognitive–perception demand (i.e., signal-to-noise extraction), which may be reflected in our activations.

There were a variety of other brain areas activated in these studies. The activations in the motor (BA 4), premotor (BA 6), and primary sensory (BA 2, 3) areas were likely to be related to subjects' left-handed button press responses. Bilateral anterior cingulate activations were observed, similar to other studies of language and pitch processing [e.g., orally producing words, Tan et al. (2001); word discrimination, Kiehl et al. (1999); identifying words from multiple talkers, Wong et al. (2004)]. Perry et al. (1999) found that relative to passive listening to complex tones, singing a single pitch activated the right anterior cingulate gyrus. Zatorre et al. (1994) found right and midline anterior cingulate activations in pitch discrimination. Thus, left and midline anterior cingulate cortex are implicated in the production and perception of ordinary language, whereas the right and midline anterior cingulate cortex appear to support pitch production.

Basal ganglia activations were observed in both tasks and both subject groups. Similar activations were observed in other language and pitch processing studies (Brown et al., 2004). Lesions in the thalamus are associated with language deficits (Nadeau and Crosson, 1997). Specifically, the pulvinar-lateral posterior complex of the thalamus is implicated in language because of its extensive connections with temporal and parietal language areas (Jones, 1985). Not surprisingly, we observed thalamic activations, especially in the pulvinar. Thalamic activations were observed in other functional imaging studies of tasks with lexical tones (Hsieh et al., 2001).

Investigations of lexical tone processing, like that here, report cerebellar activations in midline and right-sided lateral and posterior areas (Gandour et al., 2000; Li et al., 2003). It is not yet clear exactly what functions are supported by the cerebellar activity during auditory and nonmotor language tasks, although new findings suggest a role for cerebellum in supporting auditory sensory processing (Bower and Parsons, 2003) and working memory (Desmond and Fiez, 1998).

Overall, these findings imply that brain activations associated with pitch perception depend on the function of those stimuli and suggest that left anterior insula is important for lexically distinctive prosodic information. These results indicate that

speech and phonological processing may involve neural resources beyond those associated with basic auditory processing, regardless of signal complexity.

References

- Ardila A (1999) The role of insula in language: an unsettled question. *Aphasiology* 13:79–87.
- Augustine JR (1996) Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Res Brain Res Rev* 22:229–244.
- Bamiou D-E, Musiek FE, Luxon LM (2003) The insula (Island of Reil) and its role in auditory processing: literature review. *Brain Res Brain Res Rev* 42:143–154.
- Borod J (2000) *The neuropsychology of emotion*. New York: Oxford UP.
- Bower JM, Parsons LM (2003) Rethinking the lesser brain. *Sci Am* 289:50–57.
- Brown S, Martinez MJ, Hodges D, Fox PT, Parsons LM (2004) The song system of the human brain. *Cogn Brain Res* 20:363–375.
- Chuang C-K, Hiki S, Sone T, Nimura T (1972) The acoustical features and perceptual cues of the four tones of standard colloquial Chinese. *Proc Int Congr Acoust* 3:297–300.
- D'Agostino RB, Belatner A, D'Agostino Jr RB (1990) A suggestion for using powerful and informative tests of normality. *Am Stat* 44:316–321.
- Damasio H, Damasio A (1980) The anatomical basis of conduction aphasia. *Brain* 103:337–350.
- Desmond J, Fiez J (1998) Neuroimaging studies of the cerebellum: language, learning, and memory. *Trends Cogn Sci* 2:355–362.
- Fox PT, Mintun MA (1989) Noninvasive functional brain mapping by change-distribution analysis of averaged PET images of $H_2^{15}O$ tissue activity. *J Nucl Med* 30:141–149.
- Fox PT, Perlmutter JS, Raichle ME (1985) A stereotactic method of anatomical localization for positron emission tomography. *J Comp Assist Tomogr* 9:141–153.
- Fox PT, Mintun MA, Reiman EM, Raichle ME (1988) Enhanced detection of focal brain responses using inter-subject averaging and change-distribution analysis of subtracted PET images. *J Cereb Blood Flow Metab* 8:642–653.
- Friedrich CK, Alter K, Kotz SA (2001) An electrophysiological response to different pitch contours in words. *NeuroReport* 12:3189–3191.
- Friston KJ, Frith CD, Liddle PR, Frackowiak RSJ (1991) Comparing functional (PET) images: the assessment of significant change. *J Cereb Blood Flow Metab* 11:690–699.
- Fromkin VA (2000) *Linguistics: an introduction to linguistic theory*. Oxford: Blackwell.
- Gandour J, Ponglorpisit S, Potisuk S, Khunadorn F, Boongird P, Dechongkit S (1997) Interaction between tone and intonation in Thai after unilateral brain damage. *Brain Lang* 58:174–196.
- Gandour J, Wong D, Hsieh L, Weinzapfel B, Van Lancker D, Hutchins GD (2000) A cross-linguistic PET study of tone perception. *J Cogn Neurosci* 12:207–222.
- Golestani N, Zatorre RJ (2004) Learning new sounds of speech: reallocation of neural substrates. *NeuroImage* 21:494–506.
- Habib M, Daquin G, Milandre L, Royere M, Rey M, Lanteri A, Salamon G, Khalil R (1995) Mutism and auditory agnosia due to bilateral insular damage: role of the insula in human communication. *Neuropsychologia* 33:327–339.
- Herrmann CS, Friederici AD, Oertel U, Maess B, Hahne A, Alter K (2003) The brain generates its own sentence melody: a Gestalt phenomenon in speech perception. *Brain Lang* 85:396–401.
- 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.
- Jacquemot C, Pallier C, LeBihan D, Dehaene S, Dupoux E (2003) Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J Neurosci* 23:9541–9546.
- Jones EG (1985) *The thalamus*. New York: Plenum.
- Kiehl KA, Liddle PF, Smith AM, Mendrek A, Forster BB, Hare RD (1999) Neural pathways involved in the processing of concrete and abstract words. *Hum Brain Mapp* 7:225–233.
- Kirilloff C (1969) On the auditory perception of tones in Mandarin. *Phonetica* 20:63–67.
- Kotz SA, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD (2003)

- On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang* 86:366–376.
- Lancaster JL, Glass TG, Lankipalli BR, Downs H, Mayberg H, Fox PT (1995) A modality-independent approach to spatial normalization of tomographic images of the human brain. *Hum Brain Mapp* 3:209–223.
- Lancaster JL, Woldorff MG, Parsons LM, Liotti M, Fretas CS, Rainey L, Kochunov PV, Nickerson D, Mikiten SA, Fox PT (2000) Automated Talairach atlas labels for functional brain mapping. *Hum Brain Mapp* 10:120–131.
- Li X, Gandour J, Talavage T, Wong D, Dziedzic M, Lowe M, Tong Y (2003) Selective attention to lexical tones recruits left dorsal frontoparietal network. *NeuroReport* 14:2263–2266.
- Mesulam M-M, Mufson EJ (1985) The insula of Reil in man and monkey. In: *Cerebral cortex* (Peters A, Jones E, eds). New York: Plenum.
- Meyer M, Alter K, Friederici A (2003) Functional MR imaging exposes differential brain responses to syntax and prosody during auditory sentence comprehension. *J Neurolinguist* 16:277–3000.
- Mintun M, Fox PT, Raichle ME (1989) A highly accurate method of localizing regions of neuronal activity in the human brain with PET. *J Cereb Blood Flow Metab* 9:96–103.
- Nadeau SE, Crosson B (1997) Subcortical aphasia. *Brain Lang* 58:335–402.
- Oldfield R (1971) The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Paulesu E, Frith U, Snowling M, Gallagher A, Morton J, Frackowiak R, Frith C (1996) Is developmental dyslexia a disconnection syndrome? *Brain* 119:143–157.
- Perry DW, Zatorre RJ, Petrides M, Alivisatos B, Meyer E, Evans AC (1999) Localization of cerebral activity during simple singing. *NeuroReport* 10:3979–3984.
- Raichle ME, Martin WRW, Herscovitch P, Mintun MA, Markham J (1983) Brain blood flow measured with intravenous $H_2^{15}O$. Implementation and validation. *J Nucl Med* 24:790–798.
- Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W (2000) Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *NeuroReport* 11:1997–2000.
- Ross E, Edmondson J, Seibert Chan J-L (1992) Affective exploitation of tone in Taiwanese: an acoustical study of “tone latitude.” *J Phonetics* 20:441–456.
- Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS, Kabani N, Toga A, Evans A, Petrides M (1999) Three-dimension MRI atlas of the human cerebellum in proportional stereotaxic space. *NeuroImage* 10:233–260.
- Shih C-L (1988) Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory* 3:83–109.
- Smith EE, Jonides J, Marshuetz C, Koeppel RA (1998) Components of verbal working memory: evidence from neuroimaging. *Proc Natl Acad Sci USA* 95:876–882.
- Strother SC, Lang N, Anderson JR, Schaper KA, Rehm K, Hansen LK, Rotenberg DA (1997) Activation pattern reproducibility: measuring the effects of group size and data analysis models. *Hum Brain Mapp* 5:312–316.
- 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.
- Tan LH, Feng C-M, Fox PT, Gao J-H (2001) An fMRI study with written Chinese. *NeuroReport* 12:82–88.
- Warrier CM, Zatorre RJ (2004) Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain* 127:1616–1625.
- Wong PCM (2002) Hemispheric specialization of linguistic pitch contrasts. *Brain Res Bull* 59:83–95.
- Wong PCM, Diehl RL (2003) Perceptual normalization of inter- and intra-talker variation in Cantonese level tones. *J Speech Lang Hear Res* 46:413–421.
- Wong PCM, Nusbaum HC, Small SL (2004) Neural bases of talker normalization. *J Cogn Neurosci* 16:1173–1184.
- Woods RP, Mazziotta JC, Cherry SR (1993) MRI-PET registration with automated algorithm. *J Comp Assist Tomogr* 17:536–546.
- Worsley KJ, Evans AC, Marrett S, Neelin A (1992) A three dimensional statistical analysis for CBF activation studies in human brain. *J Cereb Blood Flow Metab* 12:900–918.
- Xu Y (1994) Production and perception of coarticulated tones. *J Acoust Soc Am* 95:2240–2253.
- Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953.
- Zatorre RJ, Evans EC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849.
- Zatorre RJ, Evans A, and Meyer E (1994) Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14:1908–1919.