

Activation of the Left Planum Temporale in Pitch Processing Is Shaped by Language Experience

Yisheng Xu,¹ Jackson Gandour,¹ Thomas Talavage,² Donald Wong,³
Mario Dzemidzic,⁴ Yunxia Tong,¹ Xiaojian Li,⁵ and Mark Lowe⁴

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

²School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana

³Department of Anatomy and Cell Biology, Indiana University School of Medicine,
Indianapolis, Indiana

⁴Cleveland Clinic Foundation, Cleveland, Ohio

⁵South China Normal University, Guangzhou, People's Republic of China

Abstract: Implicit, abstract knowledge acquired through language experience can alter cortical processing of complex auditory signals. To isolate prelexical processing of linguistic *tones* (i.e., pitch variations that convey part of word meaning), a novel design was used in which hybrid stimuli were created by superimposing Thai tones onto Chinese syllables (*tonal chimeras*) and Chinese tones onto the same syllables (*Chinese words*). Native speakers of tone languages (Chinese, Thai) underwent fMRI scans as they judged tones from both stimulus sets. In a comparison of native vs. non-native tones, overlapping activity was identified in the left planum temporale (PT). In this area a double dissociation between language experience and neural representation of pitch occurred such that stronger activity was elicited in response to native as compared to non-native tones. This finding suggests that cortical processing of pitch information can be shaped by language experience and, moreover, that lateralized PT activation can be driven by top-down cognitive processing. *Hum Brain Mapp* 27:173–183, 2006. © 2005 Wiley-Liss, Inc.

Key words: auditory; functional magnetic resonance imaging (fMRI); experience-dependent plasticity; planum temporale; prosody; tone; Mandarin; Thai

INTRODUCTION

A recent topic of controversy focuses on hemispheric laterality of pitch processing. Traditionally, pitch processing of complex auditory signals has been viewed as a right hemisphere (RH) function [Zatorre et al., 2002]. Yet dichotic listening and lesion deficit data clearly show that the left hemisphere (LH) is involved in pitch perception in *tone languages*, i.e., languages in which pitch variations at the syllable level are used to distinguish word meaning (e.g., Mandarin Chinese) [Gandour, 1998, 2005]. It has been proposed that auditory cortical areas in the two hemispheres differ in their relative sensitivity to temporal and spectral cues of sounds [Ivry and Robertson, 1998; Poeppel, 2003; Zatorre and Belin, 2001]. By such accounts, the LH is spe-

Contract grant sponsor: National Institutes of Health; Contract grant number: R01 DC04584-04; Contract grant sponsor: Purdue Research Foundation; Contract grant sponsor: National Institutes of Health; Contract grant number: T32 DC00030-14.

*Correspondence to: Dr. Jackson T. Gandour, Purdue University, Department of Speech, Language, and Hearing Sciences, 1353 Heavilon Hall, 500 Oval Dr., West Lafayette, IN 47907-2038.

E-mail: gandour@purdue.edu

DOI: 10.1002/hbm.20176

Published online 20 July 2005 in Wiley InterScience (www.interscience.wiley.com).

cialized for rapid temporal processing, the RH for slow spectral processing. Because pitch perception at the syllable level necessarily involves processing of slow changes in frequency, tone languages afford us an optimal window for testing competing hypotheses about hemispheric laterality of auditory signal processing. A cue-dependent hypothesis predicts RH tonal processing, whereas a function-based hypothesis predicts at least some aspects of tone processing are lateralized to the LH. By comparing tone and non-tone language speakers, or speakers of different tone languages, we are able to evaluate whether the phonological relevance of pitch information influences hemispheric laterality. The challenge for clarifying this phenomenon, however, is to isolate prelexical tonal processing from other cognitive processes involved in speech perception, especially lexical-semantic processing [Wong, 2002].

In a functional magnetic resonance imaging (fMRI) study we employ a novel experimental design involving hybrid speech stimuli to investigate whether common areas of brain activity between tone language groups (Chinese, Thai) exhibit differential responses to pitch contours associated with native and non-native tones that vary depending on a listener's language experience. The hybrid speech stimuli are *tonal chimeras* created by superimposing the tones of one tone language (Thai) onto the syllables of another tone language (Chinese). Similarly, synthetic Chinese stimuli are created by superimposing Chinese tones onto the same syllables. This procedure yields two sets of syllables that are acoustically identical except for tonal contours, thus eliminating any potential confound from concomitant temporal or intensity cues associated with Chinese and Thai tones. Moreover, tonal chimeras eliminate a potential lexical confound since these stimuli are hybrids which do not occur in the Thai lexicon.

Using fMRI it is possible to isolate the neural substrates underlying phonological processing of pitch information by identifying overlapping regions of hemodynamic activity between two tone language groups. We asked Chinese and Thai participants to make same/different judgments of paired tonal contours in the tonal chimeras and the synthetic Chinese stimuli. For Thai listeners, none of the stimuli is identifiable as a Thai word. Only the tone from a tonal chimera is native to the Thai language. Thus, we are able to demonstrate neural activity related to phonological processing in the absence of lexical-semantic processing by comparing listeners' judgments of native tones in tonal chimeras to non-native tones in synthetic Chinese stimuli. For Chinese listeners, synthetic Chinese stimuli are identifiable as real words. Only the segmental string from a tonal chimera is native to the Chinese language. In this group we elicit phonological processing, and possibly other automatic lexical-semantic processes, by comparing their judgment of native tones in synthetic Chinese stimuli relative to non-native tones in tonal chimeras. We hypothesize that conjoint brain activity [Price and Friston, 1997; Price et al., 1997] between Chinese and Thai listeners reflects a common cognitive com-

ponent underlying tonal processing independent of lexical semantics.

SUBJECTS AND METHODS

Subjects

Ten adult Chinese (five male; five female) and ten adult Thai (five male; five female) speakers were closely matched in age (Thai: $M = 25$, $SD = 4.4$; Chinese: $M = 28$, $SD = 3$) and years of formal education (Thai: $M = 19$, $SD = 1.5$; Chinese: $M = 18$, $SD = 2.5$). All participants 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. Chinese participants were from mainland China and native speakers of Mandarin; Thai participants were from Thailand and native speakers of standard Thai. 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.

Languages

Mandarin Chinese has four tonal categories [Chao, 1968]: e.g., ma^1 "mother," high level; ma^2 "hemp," high rising; ma^3 "horse," low falling rising; ma^4 "scold," high falling. Thai has five [Tingsabath and Abramson, 1993]: $khaa^M$ "stuck," mid; $khaa^L$ "galangal," low; $khaa^F$ "kill," falling; $khaa^H$ "trade," high; $khaa^R$ "leg," rising.

The primary acoustic correlate of Chinese [Howie, 1976] and Thai tones [Abramson, 1997] is F_0 . Both Chinese and Thai are *contour* tone languages [Pike, 1948], meaning that at least some of the tones must be specified as a gliding pitch movements within a tone space. Their tonal spaces are similar not only in the number of tones involving gliding pitch movements (Chinese (3): Tone 2, Tone 3, and Tone 4; Thai (3): H, R, and F), but also in their F_0 trajectories (Tone 2 \approx H; Tone 3 \approx R; Tone 4 \approx F) [Ladefoged, 2001]. Duration is a concomitant, subsidiary feature of tones in both languages [Gandour, 1994]. The shortest tone is the high falling tone (Chinese, Tone 4; Thai, F); the longest tone is the low falling rising tone (Chinese, Tone 3; Thai, R). From a perceptual standpoint, both Chinese and Thai listeners, relative to non-tone language listeners (English), place relatively greater importance on changes in pitch movement than on pitch level or height [Gandour, 1983]. Therefore, it is unlikely that any differences in brain activity between Chinese and Thai listeners can be attributed to differences in tonal typology, concomitant acoustic features, or perceptual dimensions underlying Chinese and Thai tonal spaces.

Stimuli

Recording of natural speech

A 26-year-old male native Mandarin speaker was instructed to produce a list of 428 Chinese monosyllables at a normal speaking rate. This list included 107 different onset-

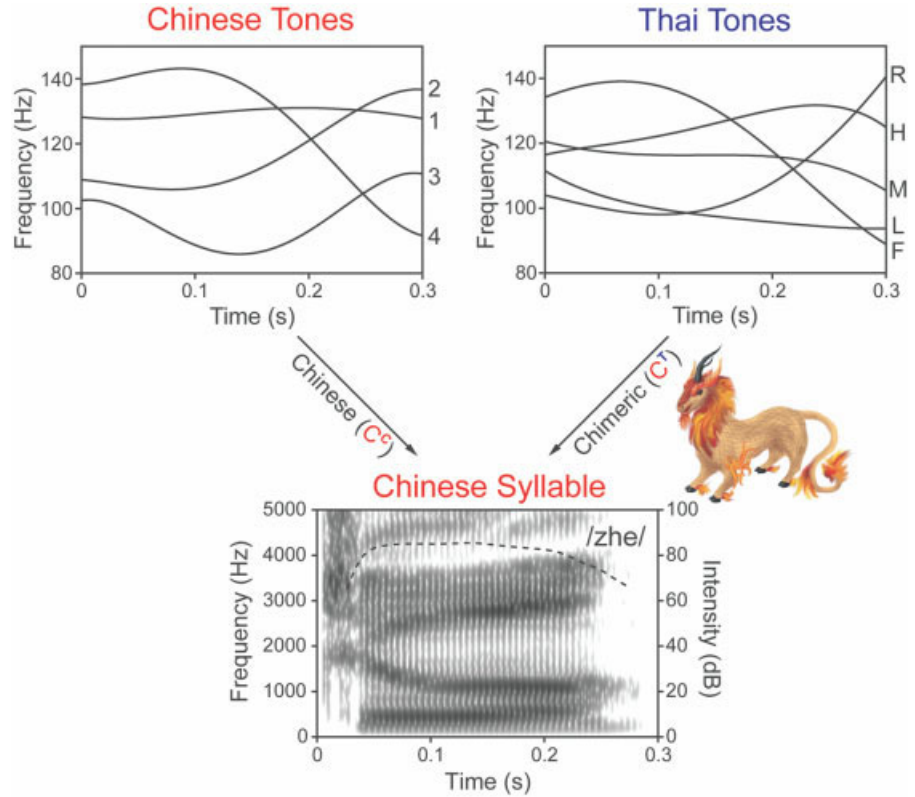


Figure 1.

Schematic diagram illustrating how Chinese (red) and Thai (blue) tones are superimposed on a Chinese syllable to create Chinese (C^C) and chimeric (C^T) stimuli, respectively. C^T stimuli are hybrids, analogous to a mythological creature (*Kilin*; middle right), and not recognizable as either a Chinese or Thai word. Upper panels display the four Chinese (top left) and five Thai (top right) tonal contours. Lower panel shows the broad-band spectrogram and intensity contour of the natural speech /zhe/ syllable.

rhyme combinations, comprising only those syllables that carry all four Chinese tones. Each syllable in the list was typed in *pinyin* followed by its corresponding Chinese character. A sufficient pause was provided between syllables in order to ensure that the talker maintained a uniform speaking rate. By controlling the pace of presentation, we maximized the likelihood of obtaining consistent, natural-sounding productions. To avoid list-reading effects, extra words were placed at the top and bottom of the list. Recordings were made in a double-walled soundproof booth using an AKG C410 headset type microphone and a Sony TCD-D8 digital audio tape recorder. The talker was seated and wore a custom-made headband that maintained the microphone at a fixed distance of 12 cm from the lips. All recorded utterances were digitized at a 16 bit/44.1 kHz sampling rate.

Creation of synthetic stimuli

Chinese (C^C) and chimeric (C^T) speech stimuli (Fig. 1) were created by superimposing Chinese (C) and Thai (T) tones on Chinese base syllables (C) using Praat software (Institute of Phonetic Sciences, University of Amsterdam), resulting in nine synthetic speech syllables (four C^C and five C^T) per onset-rhyme combination. Within each combination the recorded natural speech Tone 1 syllable was chosen as a template for analysis-resynthesis. Tone 1 is an ideal template as it exhibits less pitch movement than the other three tones and best approximates the average duration of all four tones. This enabled us to achieve superior synthetic speech quality

in terms of pitch and length manipulation (PSOLA algorithm). F_0 contours of Chinese and Thai tones were fitted by fourth-order polynomials (Appendix I) based on natural speech data [Abramson, 1962; Xu, 1997]. Stimulus duration was fixed as the average of each recorded natural speech quadruplet (e.g., /zhe/: its duration (≈ 0.3 s) was scaled to the average of four syllables across Chinese tones (/zhe¹/, /zhe²/, /zhe³/, /zhe⁴/). Intensity contour and formant frequencies, both derived from the Tone 1 template, were held constant per onset-rhyme combination.

Nonspeech counterparts (hums) of the synthetic speech stimuli were created in Praat using a flat intensity contour and six steady-state formants (bandwidths in Hz are enclosed in parentheses): 600(50), 1400(100), 2400(200) 3400(300), 4500(400), and 5500(500) Hz. Duration and F_0 contour of each speech stimulus were preserved in their nonspeech counterparts.

Prescreening of synthetic speech

All 428 synthetic C^C speech stimuli were presented individually in random order to five Chinese participants who were naive to the purposes of the experiment. Similarly, all 535 synthetic C^T speech stimuli were presented to five Thai participants. Participants were asked to indicate whether or not each stimulus corresponded to an existing word in their language. If so, they were required to write down the word corresponding to that stimulus. Thai subjects were also required to identify the Thai tone of each stimulus. Stimuli selected for experimental and training sessions had to satisfy

TABLE I. Sample trials in tone discrimination tasks

Chinese	Tone (C ^C)	Response ^a	Thai	Tone (C ^T)	Response
<i>bai</i> ¹	<i>yao</i> ¹	Same	<i>bai</i> ^H	<i>yao</i> ^L	Different
<i>zhu</i> ³	<i>xiao</i> ²	Different	<i>zhu</i> ^R	<i>xiao</i> ^R	Same
<i>tuo</i> ⁴	<i>po</i> ³	Different	<i>tuo</i> ^M	<i>po</i> ^F	Different

Segmental components of syllables, i.e., consonants and vowels, are presented in *pinyin*, a quasi-phonemic transcription system of Chinese characters. C^C = Chinese tones superimposed on Chinese syllables; C^T = Thai tones superimposed on the same Chinese syllables. Tones are represented in superscript notation: Chinese (^{1 2 3 4}); Thai (^{M L F H R}).

^a Response = correct response; subjects were required to make discrimination judgments of tones by pressing a mouse button (left button = same; right button = different).

three criteria: all C^C stimuli were identified as real Chinese words; none of the C^T stimuli was identified as a Thai word; all C^T stimuli carried clearly identifiable Thai tones (^T). These criteria were met through the consensus of at least four evaluators in each language group. Of the 80 C^C stimuli used in the fMRI experiment, 77 were identified by 5/5 Chinese listeners as Chinese words, and 3 C^C stimuli by 4/5 Chinese listeners. Of the 80 C^T stimuli in the experiment, 54 were not recognized as Thai words by 5/5 Thai listeners, and 26 C^T stimuli by 4/5 Thai listeners. Of the 80 C^T stimuli, 53 were judged to carry Thai tones by 5/5 Thai listeners, and 27 by 4/5 Thai listeners. This prescreening procedure effectively eliminated the potential confound of lexical-semantic processing in the cross-linguistic comparisons of functional imaging data. Moreover, 68% of the syllables were judged to be phonetically impermissible in the Thai language (second author). This procedure minimized any potential confound between syllable and tone in accounting for brain activity in the Thai group. When presented with the C^T stimuli, we can be confident that Thai listeners were judging native tones on non-native syllables.

Task Procedures

The experiment consisted of two tonal discrimination tasks (C^C, C^T) and a passive listening control task (L). In the discrimination tasks, subjects were required to make speeded-response, same-different judgments on pairs of Chinese (C) and Thai (T) tones (Table I) and respond by pressing the left (“same”) or right (“different”) mouse button. There were 40 pairs of syllables with 80 unique onset-rhyme combinations in each task (Appendix II). The two tasks (C^C, C^T) were made up of the same set of onset-rhyme combinations; they differed only in tone (C vs. T). In addition, each pair of syllables comprised a *nonminimal* pair. That is, onsets and rhymes always differed between the two syllables; tones were either the same or different. In nonminimal pairs subjects must parse the stimuli and extract tonal information in order to make a same-different judgment. If, however, only the tones differ between the two syllables (*minimal* pair; e.g., *ma*¹ vs. *ma*³), a subject’s judgment can be based exclusively on the holistic impression of the two sounds [Cowan, 1984].

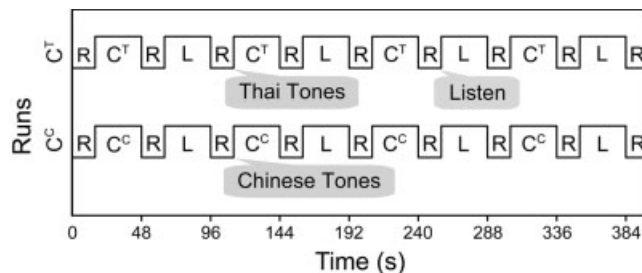


Figure 2.

A functional imaging run consisted of two tasks (active tone discrimination and passive listening) presented in blocked format (32 s) alternating with 16-s rest periods. C^C and C^T denote Chinese and chimeric synthetic speech stimuli, respectively. R = rest interval, L = passive listening to nonspeech hums. Instructions displayed in callouts were delivered to participants in their native language during the rest intervals immediately preceding their corresponding task blocks. For active task blocks, a Chinese or Thai participant made either native or non-native tone judgments depending on whether Chinese or Thai tones were presented.

These two pair types appear to recruit different neural mechanisms in phonological processing tasks [Burton et al., 2000; Gandour et al., 2003]. By using nonminimal pairs subjects are encouraged to analyze the stimuli, thus enhancing access to long-term phonological representations. None of the Chinese tone pairs contained Tone 3 on both syllables in order to avoid tone sandhi effects [Wang and Li, 1967]. Only three of 40 pairs of Thai tones involved a contrast between the mid and low tones because of their relative confusability when presented in isolation as compared to other pairs of tones [Abramson, 1975, 1997]. No paired monosyllables formed meaningful bisyllabic words in either language.

The control task involved passive listening to 40 pairs of nonspeech hums derived from the synthetic speech stimuli in the discrimination tasks. In this task, subjects were instructed to listen passively and to respond by alternately pressing the left and right mouse buttons after hearing each pair. No overt judgment was required. Half of the hums carried Chinese tones, the other half Thai tones. A hum derived from a Chinese tone was always paired with a hum derived from a Thai tone in a random order. This mixing of hums was designed to minimize the possibility of listeners recalling actual pairs of C^C or C^T speech stimuli presented in the discrimination tasks.

Prior to scanning, subjects were trained to a high level of accuracy using stimuli (C^C and C^T, 20 pairs each) different from those presented during the scanning runs: C^C (Chinese, 96.5% correct; Thai, 90.0%); C^T (Chinese, 90.0%; Thai, 92.0%).

Each functional imaging run consisted of four 32-s blocks of one discrimination task (C^C or C^T) and four 32-s blocks of the control task (L), organized in an alternating boxcar design with intervening 16-s rest periods (Fig. 2). Each block contained 10 trials. The order of scanning runs, blocks

within tasks, and trials within blocks were randomized for each subject. Instructions were delivered to subjects in their native language via headphones at the end of the rest periods immediately preceding each task: “listen” for passive listening to nonspeech stimuli; “Chinese tones” for same-different judgments on C^C speech stimuli; and “Thai tones” for same-different judgments on C^T speech stimuli. Average trial duration was 3.2 s, including an interstimulus interval of 200 ms and a response interval of 2.15 s.

All stimuli, speech and nonspeech, were digitally normalized to a common peak intensity. They were presented binaurally by means of computer playback (E-Prime) through a pair of MRI-compatible headphones (Avotec). Plastic sound conduction tubes were threaded through tightly occlusive foam ear tips inside earmuffs that attenuated the average sound pressure level of the continuous scanner noise by approximately 30 dB. Average intensity of all experimental stimuli was about 90 dB SPL against a background of scanner noise at 80 dB SPL after attenuation.

Image Acquisition

Scanning was performed on a 1.5 T Signa GE LX Horizon scanner (Waukesha, WI) equipped with a birdcage head coil. Each of two 200-volume echo-planar imaging (EPI) series was acquired using gradient-echo EPI with the following parameters: repetition time (TR) 2 s; echo time (TE) 50 ms; matrix 64×64 ; flip angle (FA) 90° ; field of view (FOV) 24×24 cm. Sixteen 7.5-mm-thick, contiguous axial slices were used to image the entire cerebrum and superior aspects of the cerebellum. Prior to functional imaging runs high-resolution, anatomic images were acquired in 124 contiguous axial slices using a 3D Spoiled-Grass (3D SPGR) sequence (slice thickness 1.2–1.3 mm; TR 35 ms; TE 8 ms; 1 excitation; FA 30° ; matrix 256×128 ; FOV 24×24 cm) for purposes of anatomic localization and coregistration to a standard stereotaxic system [Talairach and Tournoux, 1988]. Subjects were scanned with eyes closed and room lights dimmed. The effects of head motion were minimized by a head-neck pad and dental bite bar.

Image Analysis

Image analysis was conducted using the AFNI software package [Cox, 1996]. For each subject, all image volumes in a functional imaging run were motion-corrected to the fourth acquired volume of the first functional imaging run. The time series of each voxel was detrended using a second-order polynomial fit to account for temporal variation in the signal baseline, then normalized to its mean intensity. The hemodynamic responses for discrimination (C^C , C^T) and control (L) tasks were deconvolved from the baseline (rest) for each voxel in the imaging volume and evaluated by a general linear model approach. This analysis yielded differential responses, illustrated by Student’s t -statistic maps (C^C vs. L, C^T vs. L). These activation datasets, and the anatomic datasets for the corresponding subjects, were then interpolated to isotropic 1 mm^3 voxel volumes, coregistered, and transformed to the Talairach coordinate system. The activa-

tion datasets were then spatially smoothed by a 5.2 mm full-width half-maximum (FWHM) Gaussian filter chosen to reflect functional image resolution and minimize the effects of intersubject anatomic variability. For convenience of group analysis, the t -value of each voxel on the resulting statistical maps was converted to a z -score via the corresponding P value. Voxel-wise analysis of variances (ANOVAs) were conducted on z -scores (subjects as a random effect) to generate group activity maps comparing the two discrimination tasks (C^C vs. C^T). When computing these ANOVAs, only positive activations (C^C vs. L, C^T vs. L) in each individual map were retained. In order to identify commonly activated areas shared by the two groups ($C^C > C^T$ for Chinese; $C^T > C^C$ for Thai), an exploratory whole brain cluster analysis with Student’s t -statistic threshold of $t = 2.262$ (equivalent to one-tailed $P = 0.025$ per group) and volume threshold of 422 mm^3 (equivalent to the size of four original resolution voxels) were performed on each ANOVA group map. The resulting two cluster maps of the Chinese and Thai groups were then overlaid to construct a map that shows the overlapping activations between them.

Region of Interest (ROI) Analysis

Statistical validation and localization of conjoint functional activations between Chinese and Thai groups were achieved by computing the mean activation (z -score) within macroanatomically defined ROIs per individual [Brett et al., 2002]. Of two candidate areas identified in the exploratory analysis (see Results), we selected the planum temporale (PT) for ROI analysis because of its well-known role in auditory processing [Griffiths and Warren, 2002] and its potential significance for addressing theoretical issues of hemispheric laterality in relation to physical vs. functional characteristics of auditory signals [Zatorre et al., 2002].

Our ROIs represent the medial portion of the left/right PT including both gray and white matter. They were created manually slice-by-slice using a series of fixed-radius circles on the coronal planes. A 6 mm radius was used as this was the minimum size required to enclose the superior–inferior extent of the superior temporal gyrus (STG) in all subjects. Circles were placed laterally tangential to an anterior–posterior line drawn from the medialmost point of the superior temporal sulcus (STS). Each circle was individually positioned inferiorly tangential to the superior aspect STG. These circles formed a tilted 3D pipe following the anterior–posterior course of STG, bounded anteriorly by Heschl’s sulcus (HS) and posteriorly by the end of the horizontal stem of the Sylvian fissure [Steinmetz et al., 1989].

Mean z -score values for all voxels within each ROI were computed for each subject. These ROI means were analyzed using mixed-model ANOVAs with subject as a random effect nested within language group (Chinese, Thai), and task (C^C , C^T) and hemisphere (LH, RH) as fixed, repeated measure effects. To obtain a measure of the size of PT in both hemispheres the total number of ROI voxels per hemisphere was also computed for each subject. Pooling across language groups (Chinese, Thai), size of the PT was analyzed using a

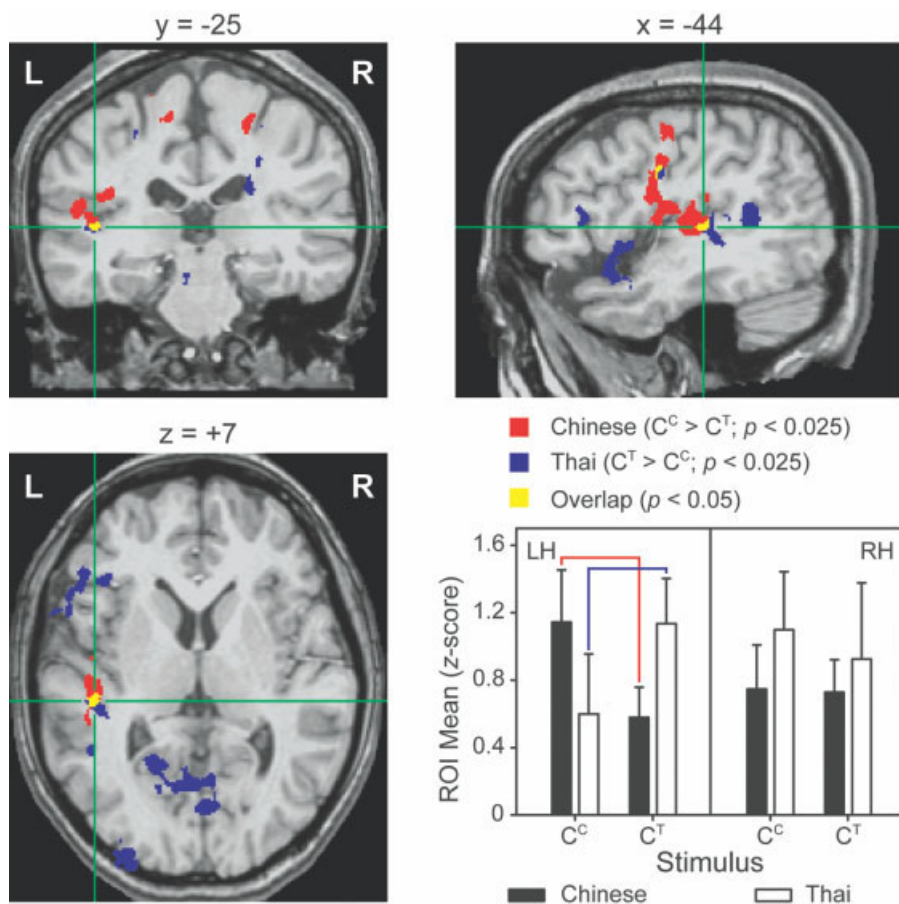


Figure 3.

Cortical areas activated in response to discrimination of Chinese and Thai tones. A common focus of activation is indicated by the overlap (yellow) between Chinese and Thai groups in the functional activation maps. Green cross-hair lines mark the stereotactic center coordinates for the overlapping region in the left planum temporale (top left: coronal section; bottom left: axial section; top right: sagittal section). A *double dissociation* (bar charts, bottom right) between tonal processing (C^C , C^T) and language experience (Chinese, Thai) reveals that for the Thai group Thai tones elicit stronger activity relative to Chinese tones, whereas for the Chinese group stronger activity is elicited by Chinese relative to Thai tones. A second focus of common activity is observed in the ventral precentral gyrus (sagittal section). L/LH = left hemisphere; R/RH = right hemisphere; ROI = region of interest.

mixed-model ANOVA for the effect of hemisphere. Pooling across groups, tasks, and hemispheres, Pearson correlation coefficients were then computed between PT sizes and functional activation.

Task Performance Analysis

Accuracy, reaction time (RT), and subjective ratings of task difficulty were used to measure task performance. Each task was self-rated by listeners on a 1–5-point graded scale of difficulty (1 = easy, 3 = medium, 5 = hard) at the end of the scanning session. To satisfy assumptions of normality and homogeneity of variance, accuracy (i.e., proportion of correct responses) and reaction time were arcsine- and log-transformed, respectively [Winer et al., 1991, pp. 354–358]. In addition, raw values exceeding two standard deviations were discarded.

To analyze task performance data in relation to functional activation, Pearson correlation coefficients were computed between the three behavioral measures and the activation (mean z-scores) of the left/right PT ROIs for all subjects pooled across tasks (C^C , C^T) and language groups (Chinese, Thai). The purpose of this pooled analysis was to tease apart the effect of language experience from more general sensory and decision-making processes in auditory perception [Binder et al., 2004].

RESULTS

Conjoint Activations

Two conjoint areas of activation were identified in the overlaid map between the Chinese ($C^C > C^T$) and the Thai ($C^T > C^C$) groups, located at the anterior-medial portion of the left PT (center of mass (CM) coordinates: $-44, -25, 7$) and the ventral part of the left PrCG (CM: $-43, -6, 29$) (Fig. 3, top and bottom left). These two conjoint areas were also confirmed by evaluating the volume size and CM coordinates of each cluster from the group-wise activations (Table II).

ROI: Planum Temporale

Results of a mixed-model ANOVA showed a significant three-way interaction (group \times task \times hemisphere) in ROI means ($P = 0.0487$) for the PT. A breakdown of the model revealed a group \times task interaction (Fig. 3, bottom right) that occurred in the left PT only ($P = 0.0014$). Tests for simple main effects indicated significantly higher ROI means in the C^C than C^T task for the Chinese group ($P = 0.0173$), and a reverse pattern for the Thai group ($C^T > C^C$; $P = 0.0132$). No significant hemispheric laterality effects were observed for either language group when examining

TABLE II. Neural regions showing effects of comparisons between Chinese and chimeric stimuli by language group ($P < .025$)

Region	BA	Chinese group ($C^C > C^T$)			Thai group ($C^T > C^C$)		
		Volume	CM (x, y, z)	PA (x, y, z)	Volume	CM (x, y, z)	PA (x, y, z)
<i>Frontal</i>							
L posterior MFG	6				1352	-35, 0, 36	-35, 6, 43
L anterior INS					2760	-48, 19, 3	-40, 9, 1
L mdFG (SMA)	6	889	-18, -10, 65	-12, -4, 65			
R mdFG (SMA)	6				666	-2, 15, 56	4, 14, 60
L PrCG	6/4	466	-43, -11, 49	-47, -4, 53	615	-30, -20, 49	-22, -19, 52
R PrCS	6	1398	24, -17, 56	25, -12, 57			
R ACC	32				447	17, 17, 32	22, 18, 30
<i>Parietal</i>							
L PoCG	3	735	-17, -31, 59	-15, -34, 74			
R PoCG	43	2110	49, -13, 20	58, -14, 18			
L precuneus	7	534	-15, -71, 42	-12, -66, 44			
R SPL	7	512	18, -50, 60	18, -53, 59			
L PCG/POS	31				3400	-5, -63, 10	-6, -63, 15
<i>Temporal</i>							
L PT	42/22	2548	-45, -18, 16	-52, -25, 15	490	-43, -29, 6	-43, -27, 7
L posterior MTG	21				717	-47, -48, 13	-44, -45, 11
<i>Occipital</i>							
L cuneus	19				834	-12, -84, 35	-14, -83, 35
L IOG	18				1679	-31, -90, 4	-31, -90, -2
<i>Cerebellum</i>							
Vermis					920	2, -69, -25	4, -67, -26
R lateral		927	25, -63, -20	41, -52, -19			

Volume size of clusters is expressed in mm^3 . Stereotactic ordinates of the center of mass (CM) and peak activation (PA) of each foci, in millimeters, are measured in the Talairach space. The x -coordinate refers to medial-lateral position relative to midline (negative = left); y -coordinate refers to anterior-posterior position relative to the anterior commissure (positive = anterior); z -coordinate refers to superior-inferior position relative to the AC-PC (anterior commissure-posterior commissure) line (positive = superior). Designation of Brodmann's areas (BA) are referenced to the atlas of Talairach and Tournoux (1988), and are approximate only. Boldface type indicates a common focus of activation between language groups. L, left; R, right; ACC, anterior cingulate gyrus; INS, insula; IOG, inferior occipital gyrus; mdFG, medial frontal gyrus; PCG, posterior cingulate gyrus; PoCG, postcentral gyrus; POS, parietal occipital sulcus; PrCG, precentral gyrus; PrCS, precentral sulcus; PT, planum temporale; SMA, supplementary motor area; SPL, superior parietal lobule.

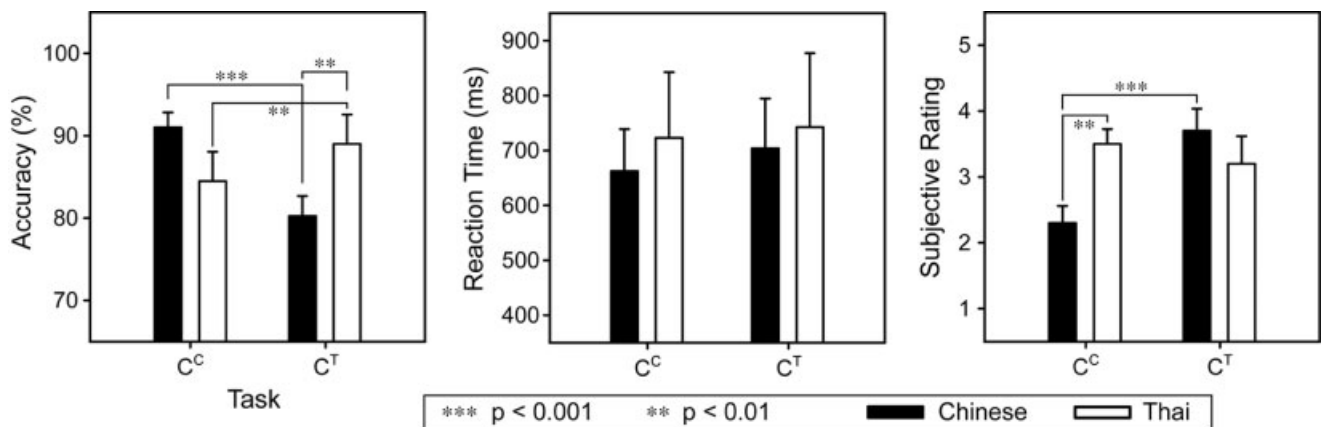


Figure 4.

Task performance of Chinese and Thai groups (M, SE) measured by accuracy (% correct), reaction time (ms) and subjective ratings of task difficulty. Significant simple main effects are marked by asterisks.

each task (C^T , C^C) separately, meaning that activation was bilateral.

The left lateralized difference between native and non-native tones cannot be explained by anatomic asymmetry. Although ROI sizes (mean \pm SD, left PT: $2040 \pm 441 \text{ mm}^3$; right PT: $1424 \pm 316 \text{ mm}^3$) showed a significant leftward asymmetry ($P < 0.0001$), there was no correlation between ROI sizes and means (Pearson $r = -0.0102$).

Task Performance

Results of a mixed-model ANOVA of the behavioral data revealed significant group \times task interactions in accuracy ($F_{1,16} = 34.16$; $P < 0.0001$) and subjective ratings ($F_{1,15} = 13.74$; $P = 0.0021$) (Fig. 4). Comparing tasks (C^C , C^T), the Chinese group was more accurate on C^C ($F_{1,16} = 25.54$; $P = 0.0001$), and subjectively rated the C^C task to be easier than C^T ($F_{1,15} = 18.08$; $P = 0.0007$). The Thai group, on the other hand, achieved a higher level of accuracy on C^T ($F_{1,16} = 10.31$; $P = 0.0054$), and rated the two tasks to be nearly equal in difficulty. Comparing groups (Chinese vs. Thai), the Thai group was more accurate than Chinese on the C^T task ($F_{1,16} = 9.58$; $P = 0.0069$), whereas the C^C task was easier for the Chinese than the Thai group ($F_{1,15} = 9.59$; $P = 0.0074$). No other between-group comparisons reached significance, neither were there any significant main or interaction effects for RT.

Pooling across tasks and language groups, there was a significant positive correlation between RT and the ROI means of the left PT only ($r = 0.4582$; $P = 0.0029$), whereas no correlation was observed between accuracy and PT activation or between subjective ratings and PT activation in either hemisphere.

DISCUSSION

These findings demonstrate that cortical processing of pitch information can be shaped by language experience. The left PT activation in both Chinese and Thai groups reflects enhanced computation elicited by phonological representations of their native tonal categories. However, we are not claiming that the left PT is the primary storage site for these long-term representations. Rather, we argue that the left PT has "access" to discrete phoneme categories. This interpretation extends the model of the PT as a computational hub for the segregation and matching of incoming auditory signals with learned spectrotemporal representations to accommodate top-down as well as bottom-up processing [Griffiths and Warren, 2002]. Functional asymmetry for auditory processing can be driven by higher-level abstract information overriding lower-level physical properties of stimulus inputs. Both top-down and bottom-up processing are essential features of the segregation process. This reciprocity allows for plasticity, i.e., modification of neural mechanisms involved in pitch processing based on language experience. Thus, it appears that computations at a relatively early stage of acoustic-phonetic processing in auditory cor-

tex can be modulated by stimulus features that are phonologically relevant in particular languages.

Converging evidence from a variety of approaches including aphasia [Blumstein, 1995], direct cortical stimulation [Boatman, 2004], and brain imaging [Jacquemot et al., 2003; Jancke et al., 2002; Naatanen et al., 1997; Phillips et al., 2000; Scott and Wise, 2004] similarly indicates that the left PT is sensitive to language experience and plays a critical role in phonological decoding. The presence of spectrally complex speech sounds, however, is not a prerequisite for PT activation. For example, in Morse code an acoustic message is transmitted using a combination of tone patterns that otherwise follows the grammatical rules of a particular language. After intensive training, lateralization of the mismatch negativity response shifts from the right to the left auditory cortex [Kujala et al., 2003]. This finding demonstrates experience-dependent effects of the close associations between the neural representations of the tone patterns and phonemes of the spoken language. Moreover, activation of the left PT is neither domain-specific nor modality-specific. In the music domain, it has been reported that musicians with long-term training and absolute pitch show stronger activation in the left PT than non-musician controls [Ohnishi et al., 2001]. In the visual modality, the perception of natural sign languages requires phonological decoding of complex dynamic hand configurations and hand movements in relation to the body. Both deaf and hearing signers show greater activation of the left PT when processing their native sign language as compared to nonlinguistic gestural controls [MacSweeney et al., 2004], suggesting that the left PT may be specialized for the analysis of phonologically structured material whatever its input modality. It has also been implicated in processing visual to auditory cross-modal associations that result from long-term learning of either language (reading written language [Nakada et al., 2001], lip reading [Calvert et al., 1997], or music (piano "key-touch reading") [Hasegawa et al., 2004]). In sum, activity of the left PT appears to be sensitive to learning experience that involves the development of abstract cortical representations irrespective of cognitive domain or sensory modality.

Although no significant hemispheric laterality effects were found in either the Chinese or the Thai group when examining the native or non-native tone tasks separately, it does not follow that there are no functional differences in the PT of the LH and RH. Bilateral activations can result from functional specialization of the two hemispheres subserving different processes that may be dependent or independent of language experience. In the current study, we infer that pitch processing associated with either Chinese or Thai tones activates auditory cortex in the RH regardless of language experience [Zatorre et al., 2002, and references therein], but that it shifts in laterality to the left PT when a language-sensitive task requires that listeners access higher-order categorical representations. This result is compatible with our view that speech prosody perception involves a dynamic interplay between the

two hemispheres via the corpus callosum [Gandour et al., 2004].

It is unlikely that differential activation in the left PT between native and non-native tones can be attributed to variation in either discrimination sensitivity [Binder et al., 2004] or task difficulty. Left PT activation is not correlated with either accuracy or subjective difficulty ratings even though we observe similar group and task interactions on these two behavioral measures. Instead, left PT activation is correlated with RT irrespective of group or task. Yet RT, presumed to reflect decision-making processes, appears to be positively correlated with activity in inferior frontal regions, not superior temporal regions [Binder et al., 2004]. Since there is no group, task, or group \times task interaction effect in RT, the observed correlation likely reflects interindividual differences in processing general auditory features, which suggests that other factors besides language experience can also modulate activity in the left PT. This finding is consistent with the view that the left PT is not a dedicated language processor, but is in fact engaged in the analysis of complex auditory signals [Griffiths and Warren, 2002].

The conjoint activation of Chinese and Thai listeners in the left PrCG (CM: $-43, -6, 29$) mostly overlaps with areas that control muscles of the mouth and larynx of the primary motor cortex [Fox et al., 2001]. Bilateral activation of this area has been observed for passive listening to meaningless monosyllables [Wilson et al., 2004], leading them to hypothesize that this motor area may be recruited in mapping acoustic inputs to articulatory representations during speech perception. In the present study, we observe a left-lateralized task effect for both groups (Chinese: $C^C > C^T$; Thai: $C^T > C^C$) in this same motor area that we ascribe to language experience. Its coactivation with the left PT is consistent with the view that the motor system is involved in speech perception [Liberman and Mattingly, 1985; Liberman and Whalen, 2000].

Finally, we have demonstrated that the use of hybrid speech stimuli coupled with cross-language comparisons is advantageous for identifying language-sensitive regions within a widely distributed cortical network that

subserves the processing of phonologically relevant features of speech. This methodology removes the potential confound between phonological and lexical-semantic processing that is inherent to real word or even pseudoword stimuli. Whereas the focus in this study is on phonologically relevant suprasegmental information (tone), it can easily be extended to focus on segmental information (consonant, vowel) as well.

APPENDIX I

Polynomial equations for generating Chinese and Thai F_0 contours.⁵

APPENDIX I.

Chinese tones

$$\begin{aligned} F_1 &= 123.16 - (11.61/d)x + (67.94/d^2)x^2 - (85.18/d^3)x^3 + (28.43/d^4)x^4 \\ F_2 &= 103.85 - (8.45/d)x - (76.32/d^2)x^2 + (297.91/d^3)x^3 - (185.34/d^4)x^4 \\ F_3 &= 97.45 + (15.05/d)x - (382.13/d^2)x^2 + (772.58/d^3)x^3 - (397.27/d^4)x^4 \\ F_4 &= 133.26 + (4.36/d)x + (165.15/d^2)x^2 - (504.84/d^3)x^3 + (288.71/d^4)x^4 \end{aligned}$$

Thai tones

$$\begin{aligned} F_M &= 120.57 - (26.54/d)x + (49.72/d^2)x^2 - (15.24/d^3)x^3 - (23.08/d^4)x^4 \\ F_L &= 111.52 - (62.35/d)x + (115.69/d^2)x^2 - (117.80/d^3)x^3 + (46.65/d^4)x^4 \\ F_F &= 134.22 + (39.38/d)x - (55.39/d^2)x^2 - (126.85/d^3)x^3 + (97.55/d^4)x^4 \\ F_H &= 116.47 + (23.21/d)x - (56.32/d^2)x^2 + (153.42/d^3)x^3 - (111.85/d^4)x^4 \\ F_R &= 104.00 - (25.55/d)x - (8.84/d^2)x^2 + (108.90/d^3)x^3 - (37.96/d^4)x^4 \end{aligned}$$

Constant d = fixed duration for each subset of nine stimuli (4 C^C + 5 C^T).

APPENDIX II

Forty pairs of speech syllables used in the Chinese tone discrimination task.

APPENDIX II.

Block 1	Block 2	Block 3	Block 4
捋 – 噎 lu ¹ – ye ¹	掰 – 腰 bai ¹ – yao ¹	为 – 造 wei ⁴ – zao ⁴	漏 – 这 lou ⁴ – zhe ⁴
蛇 – 节 she ² – jie ²	仇 – 雷 chou ² – lei ²	芽 – 醜 ya ² – bu ²	迂 – 家 yū ¹ – jia ¹
炉 – 朝 lu ² – chao ²	猫 – 嗨 mao ¹ – hai ¹	六 – 术 liu ⁴ – shu ⁴	些 – 一 xie ¹ – yi ¹
火 – 簸 huo ³ – bo ⁴	角 – 妈 jiao ³ – ma ¹	替 – 国 ti ⁴ – guo ²	止 – 料 zhi ³ – liao ⁴
窄 – 优 zhai ³ – you ¹	主 – 淆 zhu ³ – xiao ²	讨 – 妞 tao ³ – niu ¹	憋 – 臞 bie ¹ – chuai ²
渠 – 袜 qu ² – wa ⁴	窈 – 尬 tiao ³ – ga ⁴	倔 – 此 jue ⁴ – ci ³	兔 – 多 tu ⁴ – duo ¹
退 – 柴 tui ⁴ – chai ²	拓 – 叵 tuo ⁴ – po ³	池 – 科 chi ² – ke ¹	毁 – 熟 hui ³ – shou ²
粟 – 雪 piao ⁴ – xue ³	据 – 哩 ju ⁴ – li ¹	达 – 迫 da ² – pai ³	切 – 沫 qie ¹ – mo ⁴
时 – 傀 shi ² – kui ³	找 – 矧 zhao ³ – cuo ²	猜 – 毫 cai ¹ – hao ²	肥 – 闹 fei ² – nao ⁴
抛 – 昨 pao ¹ – zuo ²	吼 – 悄 hou ³ – qiao ¹	随 – 皱 sui ² – zhou ⁴	叉 – 序 cha ¹ – xu ⁴

For the Thai tone discrimination task, the same 40 Chinese base syllables (onset + rhyme) were used in combination with the five Thai tones.

ACKNOWLEDGMENTS

This study was supported by the NIH (to J.G. and X.L.) and by the Purdue Research Foundation (to J.G.).

REFERENCES

- Abramson AS (1962): The vowels and tones of standard Thai: acoustic measurements and experiments. Bloomington: Indiana University. Research Center in Anthropology, Folklore, and Linguistics, Pub. 20.
- Abramson AS (1975): The tones of Central Thai: Some perceptual experiments. In: Harris JG, Chamberlain JR, editors. *Studies in Tai linguistics in honor of William J. Gedney*. Bangkok: Central Institute of English Language. p 1–16.
- Abramson AS (1997): The Thai tonal space. In: Abramson AS, editor. *Southeast Asian linguistic studies in honour of Vichin Pan-pong*. Bangkok: Chulalongkorn University Press. p 1–10.
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004): Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7:295–301.
- Blumstein SE (1995): The neurobiology of the sound structure of language. In: Gazzaniga M, editor. *The cognitive neurosciences*. Cambridge, MA: MIT Press. p 915–929.
- Boatman D (2004): Cortical bases of speech perception: evidence from functional lesion studies. *Cognition* 92:47–65.
- Brett M, Johnsrude IS, Owen AM (2002): The problem of functional localization in the human brain. *Nat Rev Neurosci* 3:243–249.
- Burton MW, Small SL, Blumstein SE (2000): The role of segmentation in phonological processing: an fMRI investigation. *J Cogn Neurosci* 12:679–690.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997): Activation of auditory cortex during silent lipreading. *Science* 276:593–596.
- Chao YR (1968): *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Cowan N (1984): On short and long auditory stores. *Psychol Bull* 96:341–370.
- Cox RW (1996): AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Fox PT, Huang A, Parsons LM, Xiong JH, Zamarippa F, Rainey L, Lancaster JL (2001): Location-probability profiles for the mouth region of human primary motor-sensory cortex: model and validation. *Neuroimage* 13:196–209.
- Gandour J (1983): Tone perception in Far Eastern languages. *J Phonet* 11:149–175.
- Gandour J (1994): Phonetics of tone. In: Asher R, Simpson J, editors. *The encyclopedia of language & linguistics*. New York: Pergamon Press. p 3116–3123.
- Gandour J (1998): Aphasia in tone languages. In: Coppens P, Basso A, Lebrun Y, editors. *Aphasia in atypical populations*. Hillsdale, NJ: Lawrence Erlbaum. p 117–141.
- Gandour J (2005): Neurophonetics of tone. In: Brown K, editor. *Encyclopedia of language and linguistics*, 2nd ed. Oxford, UK: Elsevier.

- Gandour J, Xu Y, Wong D, Dziedzic M, Lowe M, Li X, Tong Y (2003): Neural correlates of segmental and tonal information in speech perception. *Hum Brain Mapp* 20:185–200.
- Gandour J, Tong Y, Wong D, Talavage T, Dziedzic M, Xu Y, Li X, Lowe M (2004): Hemispheric roles in the perception of speech prosody. *Neuroimage* 23:344–357.
- Griffiths TD, Warren JD (2002): The planum temporale as a computational hub. *Trends Neurosci* 25:348–353.
- Hasegawa T, Matsuki K, Ueno T, Maeda Y, Matsue Y, Konishi Y, Sadato N (2004): Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Brain Res Cogn Brain Res* 20:510–518.
- Howie JM (1976): *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.
- Ivry R, Robertson L (1998): *The two sides of perception*. Cambridge, MA: MIT Press.
- 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.
- Jancke L, Wustenberg T, Scheich H, Heinze HJ (2002): Phonetic perception and the temporal cortex. *Neuroimage* 15:733–746.
- Kujala A, Huotilainen M, Uther M, Shtyrov Y, Monto S, Ilmoniemi RJ, Naatanen R (2003): Plastic cortical changes induced by learning to communicate with non-speech sounds. *Neuroreport* 14:1683–1687.
- Ladefoged P (2001): *A course in phonetics*. Fort Worth, TX: Harcourt College Publishers.
- Lieberman AM, Mattingly IG (1985): The motor theory of speech perception revised. *Cognition* 21:1–36.
- Lieberman AM, Whalen DH (2000): On the relation of speech to language. *Trends Cogn Sci* 4:187–196.
- MacSweeney M, Campbell R, Woll B, Giampietro V, David AS, McGuire PK, Calvert GA, Brammer MJ (2004): Dissociating linguistic and nonlinguistic gestural communication in the brain. *Neuroimage* 22:1605–1618.
- Naatanen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A et al. (1997): Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385:432–434.
- Nakada T, Fujii Y, Yoneoka Y, Kwee IL (2001): Planum temporale: where spoken and written language meet. *Eur Neurol* 46:121–125.
- Ohnishi T, Matsuda H, Asada T, Aruga M, Hirakata M, Nishikawa M, Katoh A, Imabayashi E (2001): Functional anatomy of musical perception in musicians. *Cereb Cortex* 11:754–760.
- Oldfield RC (1971): The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Phillips C, Pellathy T, Marantz A, Yellin E, Wexler K, Poeppel D, McGinnis M, Roberts T (2000): Auditory cortex accesses phonological categories: an MEG mismatch study. *J Cogn Neurosci* 12:1038–1055.
- Pike KL (1948): *Tone languages*. Ann Arbor, MI: University of Michigan Press.
- Poeppel D (2003): The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time.’ *Speech Commun* 41:245–255.
- Price CJ, Friston KJ (1997): Cognitive conjunction: a new approach to brain activation experiments. *Neuroimage* 5:261–270.
- Price CJ, Moore CJ, Friston KJ (1997): Subtractions, conjunctions, and interactions in experimental design of activation studies. *Hum Brain Mapp* 5:264–272.
- Scott SK, Wise RJ (2004): The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92:13–45.
- Steinmetz H, Rademacher J, Huang YX, Hefter H, Zilles K, Thron A, Freund HJ (1989): Cerebral asymmetry: MR planimetry of the human planum temporale. *J Comput Assist Tomogr* 13:996–1005.
- 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.
- Tingsabadh MRK, Abramson AS (1993): Thai. *J Int Phonet Assoc* 23:25–28.
- Wang WS, Li KP (1967): Tone 3 in Pekinese. *J Speech Hear Res* 10:629–636.
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004): Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702.
- Winer BJ, Brown DR, Michels KM (1991): *Statistical principles in experimental design*. New York: McGraw-Hill.
- Wong PC (2002): Hemispheric specialization of linguistic pitch patterns. *Brain Res Bull* 59:83–95.
- Xu Y (1997): Contextual tonal variations in Mandarin. *J Phonet* 25:61–83.
- Zatorre RJ, Belin P (2001): Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953.
- Zatorre RJ, Belin P, Penhune VB (2002): Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* 6:37–46.