

POSITRON emission tomography (PET) was used in a cross-linguistic study to compare pitch processing in native speakers of English, a nontone language, with those of Thai, a tone language. When discriminating pitch patterns in Thai words, only the Thai subjects showed activation in the left frontal operculum. Activation of this region near the classically defined Broca's area suggests that the brain recognizes functional properties, rather than simply acoustic properties, of complex auditory cues in accessing language-specific mechanisms in pitch perception. *NeuroReport* 9: 2115–2119 © 1998 Rapid Science Ltd.

## Pitch processing in the human brain is influenced by language experience

Jack Gandour,<sup>CA</sup> Donald Wong<sup>1</sup>  
and Gary Hutchins<sup>2</sup>

Department of Audiology and Speech Sciences, Purdue University, West Lafayette, IN 47907-1353; Departments of <sup>1</sup>Anatomy and <sup>2</sup>Radiology, Indiana University School of Medicine, Indianapolis, IN, USA

**Key words:** Language; PET; Phonology; Pitch; Speech perception

<sup>CA</sup>Corresponding Author

### Introduction

Speech is decoded in the brain by mechanisms engaging cortical areas within the left hemisphere, yet language experience may influence which brain circuitry is employed in processing auditory cues (e.g., pitch).<sup>1</sup> Right hemisphere regions have been implicated in pitch perception of non-linguistic stimuli.<sup>2</sup> Whether specialized neural mechanisms at higher cortical levels underlie pitch perception of linguistic stimuli is not known. Positron emission tomography (PET) was used in a cross-linguistic study to compare pitch processing in native speakers of English, a non-tone language, with those of Thai, a tone language (i.e., one in which pitch patterns are phonologically significant in monosyllabic words). When discriminating pitch patterns in Thai words, only the Thai subjects showed activation in the left frontal operculum. Activation of this region near the classically defined Broca's area suggests that the brain recognizes functional properties, rather than simply acoustic properties, of complex auditory cues in selectively accessing language-specific mechanisms in pitch perception. Thus, pitch processing in a linguistic context will preferentially activate specialized speech centers in the left hemisphere.

Prosody is associated with the melodic features of spoken language (e.g., pitch, duration, loudness). In studies of pitch processing, the major focus has been directed at complex perceptual analysis of music<sup>2</sup> and non-linguistic aspects of speech<sup>3,4</sup> mediated by right hemisphere regions. A fundamental question is

whether similar neural mechanisms at higher cortical levels are engaged for pitch perception of linguistic aspects of speech. Tone languages afford a unique opportunity for distinguishing pitch processing in nonlinguistic and linguistic contexts. The current study uses functional brain imaging techniques to compare the patterns of activated foci between Thai and English listeners when performing same-different judgements of pitch patterns in Thai word-pairs (speech stimuli) or their low-pass, speech-filtered versions (non-speech stimuli). This experimental paradigm permits us to tease apart how Thai listeners process pitch patterns in speech and non-speech stimuli. By virtue of their language experience, they are expected to associate pitch patterns with Thai tones, and thereby engage phonological mechanisms in processing speech stimuli.

### Subjects and Methods

**Subjects:** Five (two male, three female) adult, native speakers of Thai were closely matched to five adult speakers of American English (three male, two female) in age (mean  $\pm$  s.d.: Thai  $25.2 \pm 3.1$ , English  $24.6 \pm 2.4$ ) and years of formal education (Thai  $15.8 \pm 2.3$ , English  $16.6 \pm 2.1$ ). All subjects were strongly right-handed ( $LQ \geq 90$ )<sup>5</sup> with no family history of left handedness, musically untrained, with normal hearing sensitivity. Subjects were paid \$75 for their participation, and gave written informed consent. The experimental protocol was approved by the

Institutional Review Board of Indiana University Purdue University Indianapolis.

**Language:** Thai has five contrastive lexical tones, traditionally labelled mid (ˊ), low (ˋ), falling (ˋˊ), high (ˊˊ), and rising (ˊˊˊ): e.g., \kʰaa\ 'stuck' \kʰaa\ 'galangal' \kʰaa\ 'kill' \kʰaa\ 'trade' \kʰaa\ 'leg'. The primary acoustic correlate of Thai tones is voice fundamental frequency. The mid tone can be described phonetically as mid level with a final drop, low tone as low falling, falling tone as high falling, high tone as high rising, and rising tone as low rising.

**Stimuli:** Stimuli consisted of 80 pairs of natural speech, monosyllabic Thai words and 80 pairs of low pass speech-filtered (200 Hz with 50 dB per octave attenuation rate) versions of the same Thai words. The latter constituted the non-speech pitch condition (Table 1). Initial consonants within each pair differed in one half of the pairs (e.g., same tone-different consonant: \paa\ 'aunt' vs \lao\ 'shield'); lexical tones differed in the other half (e.g., same consonant-different tone: \puu\ 'grandfather' vs \poo\ 'hemp'). Syllable rhymes in any given pair were always different. In the filtered stimuli, the aim was to eliminate semantic and segmental phonetic (i.e., consonant, vowel) information while at the same time preserving suprasegmental (duration, loudness level, pitch contour) information. Stimulus duration was unaffected by filtering; digital editing was used to equalize loudness levels between filtered and unfiltered stimuli. As judged by three Thai listeners (not used in imaging study), none of the filtered stimuli were recognizable as Thai words.

**Tasks:** A total of three stimulus conditions were presented to subjects: baseline, pitch, and tone (Table 1). The baseline condition was silence; subjects were instructed to relax, and no overt motor response was required. In the pitch and tone conditions, subjects were required to make discrimination judgments of pitch patterns and Thai lexical tones, respectively, by clicking a mouse button (left button, same; right button, different). The pitch condition was a nonlinguistic task; the tone condition was a linguistic task for Thai subjects only. Auditory stimuli were presented in pairs on each trial.

Interstimulus interval was 250 ms; intertrial interval 2 s. Stimuli were delivered at a rate of one trial every 3 s. All stimuli were presented binaurally via foam insert earphones (E-A-RTONE 3A) at a comfortable listening level of ~75 dBA.

**PET image acquisition and processing:** During the experimental session, the subject lay supine on the imaging table with his/her eyes blindfolded. A total of eight scans was performed on a Siemens 951/31R system using the bolus  $H_2^{15}O$  method and 90 s scanning intervals. The baseline condition was presented twice, always the first and last scans. Consonant (not reported here), tone, and pitch conditions were presented twice in pseudo-random order to each subject. A baseline subtraction approach requires fewer *a priori* assumptions about the mechanisms underlying task performance, and can be used to verify hierarchical subtractions. Repetition of stimulus conditions permitted us to average data within subjects. For each condition, instructions were given immediately prior to scanning, and several practice trials were given prior to bolus injection.

The following three paired-image subtractions were performed on averaged group data: tone-pitch, tone-baseline, and pitch-baseline. In tone-pitch, the processing component common to both tasks is auditory processing for both Thai and English listeners. The difference between the tone and pitch tasks for Thai listeners is that phonological processing is required in the tone task by virtue of the fact that pitch patterns are associated with lexical tones in Thai.

The presence of significant focal changes in regional cerebral blood flow (rCBF) was tested by the Hammersmith method.<sup>6,7</sup> Values of  $t \geq 0.05$  (one-tailed, corrected) were deemed statistically significant.

## Results

The two language groups showed no significant ( $\alpha = 0.01$ ) differences in behavioral measures of performance on either the pitch or tone task. Reaction times did not differ significantly on either the pitch (English, 0.78 s; Thai, 0.67 s) or tone task (English, 0.95 s; Thai, 0.94 s). Neither did percentage correct differ between groups (pitch:

**Table 1.** Summary of PET experimental paradigm.

Condition	Stimulus	Response	Example
Baseline	Silence	No response	
Pitch (non-speech)	Lowpass filtered Thai words	Click button	
Tone (speech)	Thai words	Click button	/tʰoo/ 'discouraged' /tʰuu/ 'to rub' (No) /laj/ 'to chase' /law/ 'coop' (No) /choo/ 'cluster' /moo/ 'to grind' (Yes)



English, 69%; Thai, 76%; tone: English, 65%; Thai, 66%). Similarly, response sensitivity ( $d'$ )<sup>8</sup> was not significantly different between groups (pitch: English, 1.06; Thai, 1.56; tone: English, 0.82; Thai, 1.01). Nor were there any significant differences in response bias between groups as measured by beta and criterion values.

The two tasks showed no significant differences in behavioral measures of performance. Pooled across groups, reaction time on the pitch task (0.72 s) was not significantly faster than on the tone (0.95 s). Percentage correct on the tone task (72%) was not significantly higher than on the pitch (65%). Subjects' sensitivity did not differ significantly between the pitch ( $d' = 1.31$ ) and tone ( $d' = 0.92$ ) tasks. Nor were there any significant differences in response bias between tasks as measured by beta and criterion values. Therefore, any observed differences in brain

activation patterns between the two tasks are likely to reflect differences in cognitive processing between the two groups.

Focal regions of significant rCBF changes were examined for paired-image subtractions between the two active (i.e., requiring an overt discrimination judgement) tasks (tone minus pitch) and each active task from the silent baseline (tone minus baseline, pitch minus baseline). We hypothesized that additional processing regions for the Thai group would emerge in the tone minus pitch subtraction since perception of lexical tone requires phonological processing beyond the earlier cortical stages of pitch processing. An increased rCBF was observed in the left frontal operculum for the Thai group only (Table 2, Fig. 1).

The tone minus baseline subtraction also showed rCBF increases in the left inferior frontal gyrus with

**Table 2.** Stereotaxic coordinates of activation foci.

		Coordinates (mm)			
	Brodmann area	x	y	z	t value
Tone – pitch					
Thai					
L frontal operculum	44/45	-39	14	18	5.5
L anterior cingulate gyrus	32	-3	19	43	4.2
Tone-baseline					
Thai					
L anterior cingulate gyrus	32	-1	17	43	5.5
R superior temporal gyrus	22	55	-22	4	5.2
L superior temporal gyrus	22	-57	-19	2	4.9
L inferior frontal gyrus	45	-42	17	20	4.3
R cerebellum		37	-51	-32	4.1
English					
L anterior cingulate gyrus	32	-1	12	45	6.9
L superior temporal gyrus	22	-53	-35	9	6.5
R superior temporal gyrus	22	60	-24	0	5.6
R cerebellum		21	-53	-20	5.0
L insula		-28	21	2	4.9
L frontal operculum	44	-37	10	22	4.9
Pitch-baseline					
Thai					
L superior temporal gyrus	22	-60	-28	4	5.7
R superior temporal gyrus	22	55	-24	4	5.3
R cerebellum		10	-62	-14	5.3
L superior temporal gyrus	22	-53	-6	0	5.2
L transverse temporal gyrus	41	-37	-26	7	5.1
R superior temporal gyrus	22	57	-10	0	5.0
R cerebellum		35	-53	-29	4.3
English					
L anterior cingulate gyrus	32	-1	5	47	5.8
L superior temporal gyrus	22	-53	-35	11	5.3
R cerebellum		17	-53	-20	5.1
R superior temporal gyrus	22	55	-26	2	5.0
L superior temporal gyrus	22	-51	-17	2	4.3

Significant peak activation foci (blood flow increases only) that exceeded the Hammersmith statistical criterion of significance (adjusted  $p$  threshold = 0.05) in normalized CBF. Stereotaxic coordinates are derived from the human brain atlas of Talairach and Tournoux.<sup>9</sup> 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 CA-CP (anterior commissure-posterior commissure) line (positive = superior). Designation of Brodmann areas, based on Talairach and Tournoux, do not have absolute validity. L = left; R = right.

## Tone - Pitch

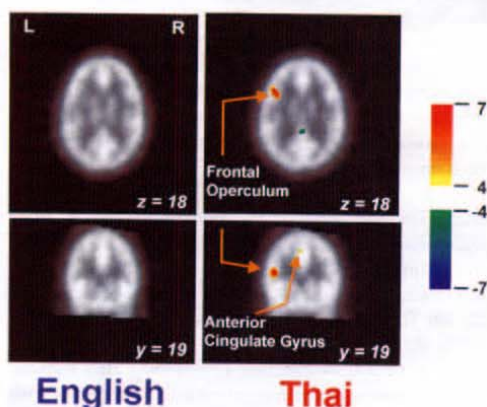


FIG. 1. Averaged PET images obtained from the tone-pitch subtraction for the English and Thai groups. Horizontal (top panels;  $z = 18$ ) and coronal (bottom panels;  $y = 19$ ) sections show foci of significant ( $t = 4.15$ ;  $p_{\text{adjusted}} < 0.05$ ) rCBF increases (red and yellow areas) in the left frontal operculum and the anterior cingulate gyrus for the Thai (right panels), but not for the English group. An area shown in green, the posterior cingulate gyrus (top right panel), is a focus of decreased rCBF.

stereotaxic coordinates (x,y,z) similar to those for the tone minus pitch subtraction (Table 2). As predicted, the pitch minus baseline subtraction did not reveal rCBF increases in the left inferior frontal gyrus because the pitch task involved processing of non-speech stimuli.

The English group showed no frontal lobe activation in the tone minus pitch subtraction. However, an increased rCBF was unexpectedly found in the left frontal operculum for the tone minus baseline subtraction (Table 2). For English subjects, listening to Thai words in the tone task would be tantamount to listening to a foreign language. Although English is a non-tone language, English listeners may have implicitly focused on consonants and vowels, both of which are phonologically significant in English, before attempting to extract the pitch patterns associated with Thai tones. The processing of consonants and vowels in performing the tone task may account for activation of left frontal cortex in the English group.

A comparison of the two language groups in the tone minus baseline subtraction revealed increased rCBF in the left insula for the English group only (Table 2). This cortical area has been implicated in phonetic processing,<sup>10-12</sup> including possibly subvocal rehearsal of vocal pitch.<sup>13</sup> Thus, insular activation in the English subjects might be expected given the increased demand placed on their phonetic processing

within a context devoid of semantic processing. In contrast, the lack of significant insular activation in the Thai subjects may be related to the fact that the speech stimuli were heard as real words presented in a context that unavoidably required both semantic and phonological processing. This group difference in insular activation, however, cannot be accounted for by the extent to which the task is learned or automatic.<sup>14</sup> Both groups received the same instructions and exposure to the auditory stimuli prior to scanning. Thus, the degree in which the insula is engaged in a task involving listening to speech stimuli may reflect the differing cortical strategies that the two groups employed in performing the same phonetic task.

Both groups showed increased rCBF at several common foci across the two baseline subtractions, tone minus baseline and pitch minus baseline: superior temporal gyrus bilaterally, anterior cingulate gyrus, and right cerebellum (Table 2). Such a distributed neural network has also been reported in earlier PET studies involving complex auditory tasks.<sup>2-4</sup> Bilateral activation of the superior temporal gyri is related to the initial cortical stages for processing complex auditory stimuli, regardless of whether the stimuli are speech or nonspeech. Increased rCBF to the anterior cingulate cortex in these active tasks is likely to be attributed to the subjects' receiving of instructions before the task, their preparation, planning, and anticipation of the cognitive task, rather than task-related processing itself.<sup>15</sup> Increased rCBF to the right cerebellum is related to the motor response evoked by the right hand in mouse clicking. In the pitch minus baseline subtraction (Table 2), neither group showed any significant activation in the right frontal cortex. The absence of activation in this region, in contrast to earlier PET studies,<sup>2-4</sup> is probably due to differences in memory load associated with pitch tasks and in the control tasks used for comparison with pitch.

## Discussion

This cross-linguistic comparison provides support for the view that complex auditory processing leading to speech perception undergoes discrete processing stages, each involving separate cortical areas of a distributed neural network. Whether linguistic or nonlinguistic, perceptual analysis of auditory stimuli occurs in the temporal lobe. However, when a phonological decision is to be made, subjects must access articulatory representations involving neural circuits that include Broca's area.<sup>3,4</sup> The robust activation of the left frontal operculum in the Thai group provides cross-linguistic, physiological evidence that lexical tone perception involves both left frontal and



temporal regions. These findings challenge the classical view that speech production is mediated exclusively by anterior brain structures, speech perception by posterior.<sup>16</sup>

The group difference reflected in the insular activation is consistent with the existence of parallel and separate pathways postulated by others for linking the temporo-parietal to the frontal language area in speech processing,<sup>17,18</sup> one via the arcuate fasciculus, the second involving the insula as a relay center. It is possible that these pathways are differentially activated depending on cognitive strategies used by Thai and English subjects when processing Thai words.

A central, yet controversial issue concerning the neural mechanisms underlying the processing of communication sounds is whether speech perception is mediated by unique neural networks in the human brain.<sup>19–21</sup> In the present study, the differential processing of pitch patterns as non-speech or speech provides support for the functional load theory of prosody.<sup>1</sup> This suggests that auditory parameters of the speech signal are not simply encoded in higher cortical areas by their complex acoustic properties,<sup>19,20,22</sup> but instead by their linguistic relevance in a particular language.<sup>21</sup> For the comparison of the tone to pitch task, only the Thai group shows activation in Broca's area because pitch variations are perceived by native Thai listeners as phonologically significant at the lexical level in their language. However, when the same Thai listeners are presented with homologous pitch contours in a non-linguistic context, they do not show a similar left frontal lobe activation. These findings are difficult to reconcile with the view that speech perception simply involves recruiting circuits that already exist for complex auditory analysis. For example, it has been postulated that the cortical processing of rapid temporal acoustic cues (20–40 ms; cf. Ref. 19 and 23) underlies the well-known left-hemispheric specialization for human speech perception. Discrimination of lexical tones in the present study, however, requires much longer temporal intervals (250–350 ms) for integrating voice fundamental frequency cues. Therefore, left hemisphere activation in the processing of lexical tones cannot be attributed to rapid temporal processing.

Contrary to the 'speech is not special' view, our findings on pitch processing by Thai listeners support the hypothesis that humans possess specialized cortical modules activated only by speech.<sup>21</sup>

## Conclusion

We have added physiological evidence to earlier psychological and clinico-anatomical evidence<sup>24,25</sup> that functional characteristics of auditory stimuli differentially influence the brain circuits used at higher cortical stages for speech processing. We have also demonstrated that the grand enterprise of mapping language and prosody onto the human brain can be vitally enhanced by cross-linguistic studies.

## References

1. Van Lancker D. *Int J Hum Commun* **13**, 201–277 (1980).
2. Zatorre R, Evans A and Meyer E. *J Neurosci* **14**, 1908–1919 (1994).
3. Zatorre R, Evans A, Meyer E and Gjedde A. *Science* **256**, 846–849 (1992).
4. Zatorre R, Meyer E, Gjedde A and Evans A. *Cerebr Cortex* **6**, 21–30 (1996).
5. Oldfield R. *Neuropsychologia* **9**, 97–113 (1971).
6. Friston K, Frith C, Liddle P et al. *J Cerebr Blood Flow Metab* **10**, 458–466 (1990).
7. Friston K, Frith C, Liddle P and Frackowiak R. *J Cerebr Blood Flow Metab* **11**, 81–95 (1991).
8. MacMillan N and Creelman C. *Detection Theory: A User's Guide*. New York: Cambridge University Press, 1991.
9. Talairach J and Tournoux P. *Co-planar Stereotaxic Atlas of the Human Brain, 3-D Proportional System: An Approach to Cerebral Imaging*. New York, Springer Verlag, 1988.
10. Dronkers N. *Nature* **384**, 159–161 (1996).
11. Paulsen E, Frith C and Frackowiak R. *Nature* **362**, 342–345 (1993).
12. Paulsen E, Frith U, Snowling M et al. *Brain* **119**, 143–157 (1996).
13. Perry D, Alivisatos B, Evans A et al. *J Acoustic Soc Am* **93**, 2403 (1993).
14. Raichle M, Fiez J, Videen T et al. *Cerebr Cortex* **4**, 8–26 (1994).
15. Murtha S, Chertkow H, Beauregard H et al. *Hum Brain Mapp* **4**, 103–112 (1996).
16. Paus T, Perry D, Zatorre R et al. *Eur J Neurosci* **8**, 2236–2246 (1996).
17. Lichtheim L. *Brain* **7**, 433–484 (1885).
18. McCarthy R and Warrington E. *Brain* **107**, 463–485 (1984).
19. Schwartz J and Tallal P. *Science* **207**, 1380–1381 (1980).
20. Fitch R, Miller S and Tallal P. *Ann Rev Neurosci* **20**, 331–353 (1997).
21. Liberman A and Mattingly I. *Science* **243**, 489–494 (1989).
22. Van Lancker D and Sidtis J. *J Speech Hear Res* **35**, 953–970 (1992).
23. Studdert-Kennedy M and Mody M. *Psychonomic Bull Rev* **2**, 508–514 (1995).
24. Gandour J. *Phonetics and phonology*. In: Stemmer B and Whitaker H, eds. *Handbook of Neurolinguistics*. New York: Academic Press, 1998: 207–219.
25. Baum S and Pell M. The neural bases of prosody: Insights from lesion studies and neuroimaging. In: Lynch M, ed. *The Cognitive Science of Prosody: Interdisciplinary Perspectives*. Amsterdam: North-Holland, in press.

ACKNOWLEDGEMENTS: This research was supported by the Indiana University School of Medicine (Physiologic Imaging Research Center) and Purdue University (Showalter Trust). Thanks to Rich Fain, Susan Giger, Li Hsieh, Kevin Perry, Nakarin Sathamnuwong, Diana Van Lancker, and Brett Weinzapfel for their contributions to this effort.

Received 7 April 1998;

accepted 16 April 1998