

A Cross-Linguistic fMRI Study of Spectral and Temporal Cues Underlying Phonological Processing

Jack Gandour¹, Donald Wong², Mark Lowe², Mario Dzemidzic²,
Nakarin Satthamnuwong¹, Yunxia Tong¹, and Xiaojian Li¹

Abstract

■ It remains a matter of controversy precisely what kind of neural mechanisms underlie functional asymmetries in speech processing. Whereas some studies support speech-specific circuits, others suggest that lateralization is dictated by relative computational demands of complex auditory signals in the spectral or time domains. To examine how the brain processes linguistically relevant spectral and temporal information, a functional magnetic resonance imaging study was conducted using Thai speech, in which spectral processing associated with lexical tones and temporal processing associated with vowel length can be differentiated. Ten Thai and 10 Chinese subjects were asked to perform discrimination judgments of pitch and timing patterns presented in the same auditory stimuli under two different conditions: speech (Thai) and nonspeech (hums). In the speech condition, tasks required judging Thai tones

(T) and vowel length (VL); in the nonspeech condition, homologous pitch contours (P) and duration patterns (D). A remaining task required listening passively to nonspeech hums (L). Only the Thai group showed activation in the left inferior prefrontal cortex in speech minus nonspeech contrasts for spectral (T vs. P) and temporal (VL vs. D) cues. Thai and Chinese groups, however, exhibited similar fronto-parietal activation patterns in nonspeech hums minus passive listening contrasts for spectral (P vs. L) and temporal (D vs. L) cues. It appears that lower level specialization for acoustic cues in the spectral and temporal domains cannot be generalized to abstract higher order levels of phonological processing. Regardless of the neural mechanisms underlying low-level auditory processing, our findings clearly indicate that hemispheric specialization is sensitive to language-specific factors. ■

INTRODUCTION

Hemispheric asymmetry underlying higher cortical functions such as speech is well established in the human brain. Yet, elucidating the neural mechanisms underlying such functional asymmetry remains a central unresolved question. Divergent views have emerged based on premises derived in part from studies of complex auditory signals ranging from nonspeech to speech. Speech perception studies addressing high-level processing have provided evidence in favor of speech-specific circuits in the left hemisphere (LH) of the human brain comparable to species-specific vocalization as posited in animals (e.g., Benson et al., 2001). Other studies, however, argue that LH lateralization for speech is a secondary consequence of LH specialization for the processing of rapidly changing acoustic information that happens to be crucially involved in speech perception (Tallal, Miller, & Fitch, 1993; Schwartz & Tallal, 1980).

More generally, it has been proposed auditory cortical areas in the two hemispheres differ in their relative sensitivity to temporal and spectral features of sounds

(Poeppel, 2001; Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002; Ivry & Robertson, 1998). Studies show that the LH preferentially extracts rapid temporal information using relatively short temporal integration windows (20–50 msec), whereas the right hemisphere (RH) extracts temporal information with longer integration windows (50–250 msec). Signal-processing principles would dictate a temporal-to-spectral trade-off, with shorter integration times providing higher temporal resolution at the expense of lower spectral resolution, while longer temporal windows providing for lower temporal but higher spectral resolution. Thus, hemispheric differences reflect the greater specialization of the LH for rapid temporal processing and the RH for fine spectral processing.

Yet, empirical data on low-level auditory processing remain conflicting and sometimes contradictory. For example, sound sequences made up of pure-tone elements with randomized pitch and duration, instead of being lateralized to the RH and LH, respectively, appear to be analyzed by a common bilateral network (Griffiths, Johnsrude, Dean, & Green, 1999). Moreover, it has been demonstrated that auditory analysis in the human brain may be based on a hierarchy of temporal levels

¹Purdue University, ²Indiana University

(Griffiths, Buchel, Frackowiak, & Patterson, 1998). Fine temporal structure associated with pitch perception implicates the auditory pathway up to and including primary auditory cortex. Longer term temporal structure associated with pitch sequences, however, appears to recruit anterior and posterior temporal regions bilaterally outside primary auditory cortex. It is suggested that these anterior and posterior areas of activation in the temporal lobe are not specific to speech or music (Griffiths et al., 1998, p. 425).

The evidence in support of low-level auditory specializations comes primarily from auditory processing asymmetries for nonspeech sounds in humans and other species and anatomical asymmetries in human auditory cortex (Zatorre et al., 2002). Data are more equivocal for the perception of human speech sounds. Hemispheric differences clearly emerge from domain-specific functions (e.g., speech, music), which are independent of low-level auditory specializations. Especially interesting are hemispheric asymmetries in the perception of speech prosody, that is, variations in the pitch, duration, and loudness of spoken utterances. Pitch variations at the syllable level that change the meaning of individual words are called lexical tones. The perception of both Thai (Gandour, Wong, et al., 2000) and Chinese (Hsieh, Gandour, Wong, & Hutchins, 2001; Klein, Zatorre, Milner, & Zhao, 2001) lexical tones is lateralized to the LH instead of the RH. Indeed, Chinese consonants, vowels, and tones are all lateralized to the LH (Hsieh et al., 2001). Pitch variations that are linguistically significant at the sentence level, on the other hand, are referred to as intonation. Both Chinese tone and intonation involve spectral processing, yet pitch contours associated with lexical tones are processed in the LH by Chinese listeners, whereas pitch contours associated with intonation are processed in the RH (Gandour, Dziedzic, et al., in press). It has also been shown that Thai vowel length, as well as syllable-final consonants, is processed in the LH by Thai listeners (Gandour et al., in press).

An optimal window for exploring how the human brain processes linguistically relevant temporal and spectral information is one in which both vowel length and tone are contrastive. That is, we need a language that exploits variations in duration and voice fundamental frequency, respectively, to distinguish vowels and lexical tones. The co-occurrence of tone and vowel length distinctions in the same language is not uncommon (I. Maddieson, personal communication, January 19, 2001). It is not at all rare in African and American tone languages (e.g., Hausa, Navaho) to find vowel length contrasts, although their tonal inventories are quite limited. Conversely, numerous Asian languages exhibit complex tonal inventories (e.g., Vietnamese), but their vowel length contrasts are severely restricted to one or a few vowels. Thai, a major language of Southeast Asia, provides an optimal window inasmuch as it exhibits five

lexical tones (e.g., $k^b aa^{mid}$ “stuck” $k^b aa^{low}$ “galangal” $k^b aa^{falling}$ “kill” $k^b aa^{high}$ “trade” $k^b aa^{rising}$ “leg”) in addition to vowel length contrasts (e.g., $baat^{low}$ “card” vs. $baat^{high}$ “unit of money”) throughout its nine-vowel system (Tingsabhadh & Abramson, 1993).

Previous studies of the perception of Thai tones (Gandour, Wong, et al., 2000) and vowel length (Gandour et al., in press) implicate neural mechanisms in left posterior inferior frontal regions. That left lateralization occurs for native Thai but not in nonnative listeners is believed to reflect the effects of abstract, higher level aspects of language on brain functioning. Yet, these two earlier studies involved two different stimulus sets and subject groups. To address the issue of temporal versus spectral processing directly, it is imperative that we ask the same group of subjects to make perceptual judgments of tone and vowel length on the same stimuli. Having subjects selectively focus their attention on different aspects of identical auditory stimuli, it seems reasonable to assume that whatever differences emerge in brain activation patterns may be attributed to cognitive processes other than attention, working memory, or size of the temporal integration window.

In this functional magnetic resonance imaging (fMRI) study, we asked Thai and Chinese subjects to perform discrimination judgments of pitch and timing patterns presented in the same auditory stimuli in two different contexts: linguistic (Thai speech) and nonlinguistic (nonspeech hums). By presenting Thai tones and vowel length to Thai and Chinese listeners, that is, listeners who do not know Thai, we were able to test whether perceptual effects on brain processing are acoustic or linguistic in nature. If acoustic in nature, effects due to this level of processing should be maintained across listeners regardless of language experience. If linguistic in nature, then we predicted perceptual effects that are unique to Thai listeners. Although Chinese listeners have no experience with Thai tones or vowel length, they do have comparable experience with Chinese tones. The question arises whether language experience with the same type of phonological unit is sufficient to lead to similar brain activation patterns as those of native listeners (cf. Gandour, Wong, et al., 2000). If the perceptual effects are driven by higher order phonological considerations, then it was predicted that brain activity of Thai and Chinese listeners will vary depending on language-specific functions of temporal and spectral cues in their respective languages.

RESULTS

Behavioral Data

The behavioral measures of task performance by Thai and Chinese groups are given in Table 1. A two-way analysis of variance on response accuracy revealed a significant two-way interaction between task and group, $F(3,72) = 12.92, p < .0001$. Post hoc Tukey multiple

Table 1. Behavioral Data

Group	Response Accuracy ^a				Response Time ^b				Self-Rating ^c			
	Thai Speech		Nonspeech Hums		Thai Speech		Nonspeech Hums		Thai Speech		Nonspeech Hums	
	T	VL	P	D	T	VL	P	D	T	VL	P	D
Thai	91.6 (2.8)	95.1 (2.8)	92.2 (1.5)	84.4 (7.0)	570 (170)	646 (162)	541 (50)	595 (154)	1.5 (0.7)	2.2 (1.1)	2.4 (0.8)	3.5 (1.1)
Chinese	92.3 (7.2)	77.8 (7.9)	88.0 (4.4)	87.3 (6.7)	503 (122)	495 (147)	512 (108)	513 (129)	1.8 (1.0)	3.9 (1.0)	2.3 (0.8)	3.2 (1.3)

T = tone (Thai speech); VL = vowel length (Thai speech); P = pitch (nonspeech hums); D = duration (nonspeech hums).

^aMean and standard deviation values (in parentheses) are expressed in percentages.

^bMean and standard deviation values (in parentheses) are expressed in milliseconds.

^cMean and standard deviation values (in parentheses) are expressed in scalar units from 1 to 5 (1 = easy, 3 = medium, 5 = hard).

comparisons ($\alpha = .01$) showed that Thai listeners were more accurate at judging Thai vowel length (95%) than Chinese listeners (78%). Each task was self-rated by

listeners on a 1- to 5-point graded scale of difficulty (1 = easy, 3 = medium, 5 = hard). Subjective ratings of task difficulty similarly showed a two-way interaction

Figure 1. Averaged fMRI activation maps for the groups of native Thai and Chinese speakers obtained by comparing speeded response, discrimination judgments of pitch, and duration patterns in a speech context relative to a nonspeech context. The top panel shows cross-linguistic differences in brain activation for the spectral contrast between Thai tones and pitch (T vs. P); the bottom panel for the temporal contrast between Thai vowel length and duration (VL vs. D). Each panel shows a left sagittal section through stereotaxic space of activation maps superimposed onto a representative brain anatomy. Stereotaxic coordinates (mm) are derived from the human brain atlas of Talairach and Tournoux (1988). Left-sided activation foci in frontal and temporo-occipital regions occur in the Thai group only. The activation map threshold is at $t(\alpha) = 5.85, p < 2.5 \times 10^{-9}$ (one-tailed, uncorrected). Student's *t* statistic values are represented by a color scheme where the significance levels increase from red to yellow.

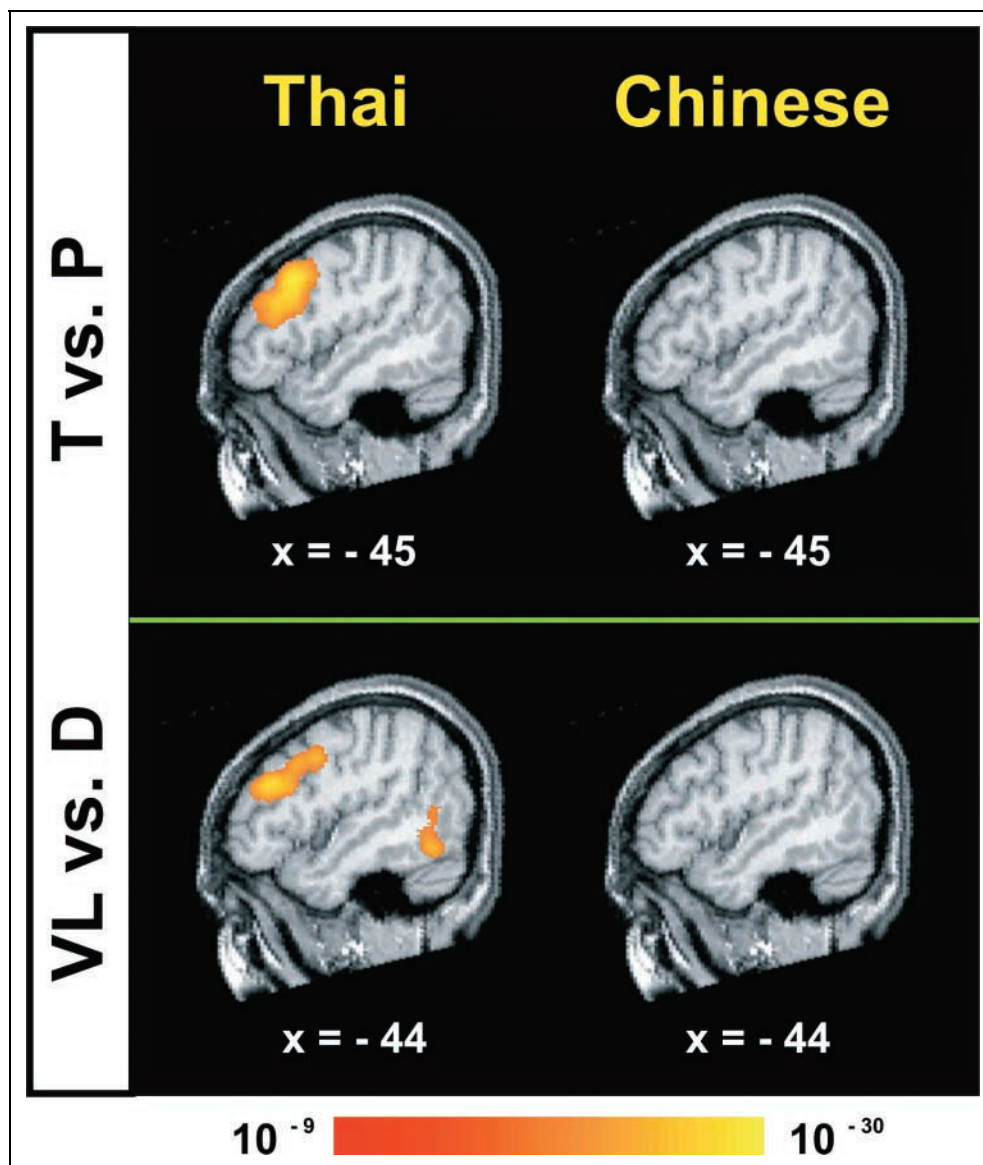


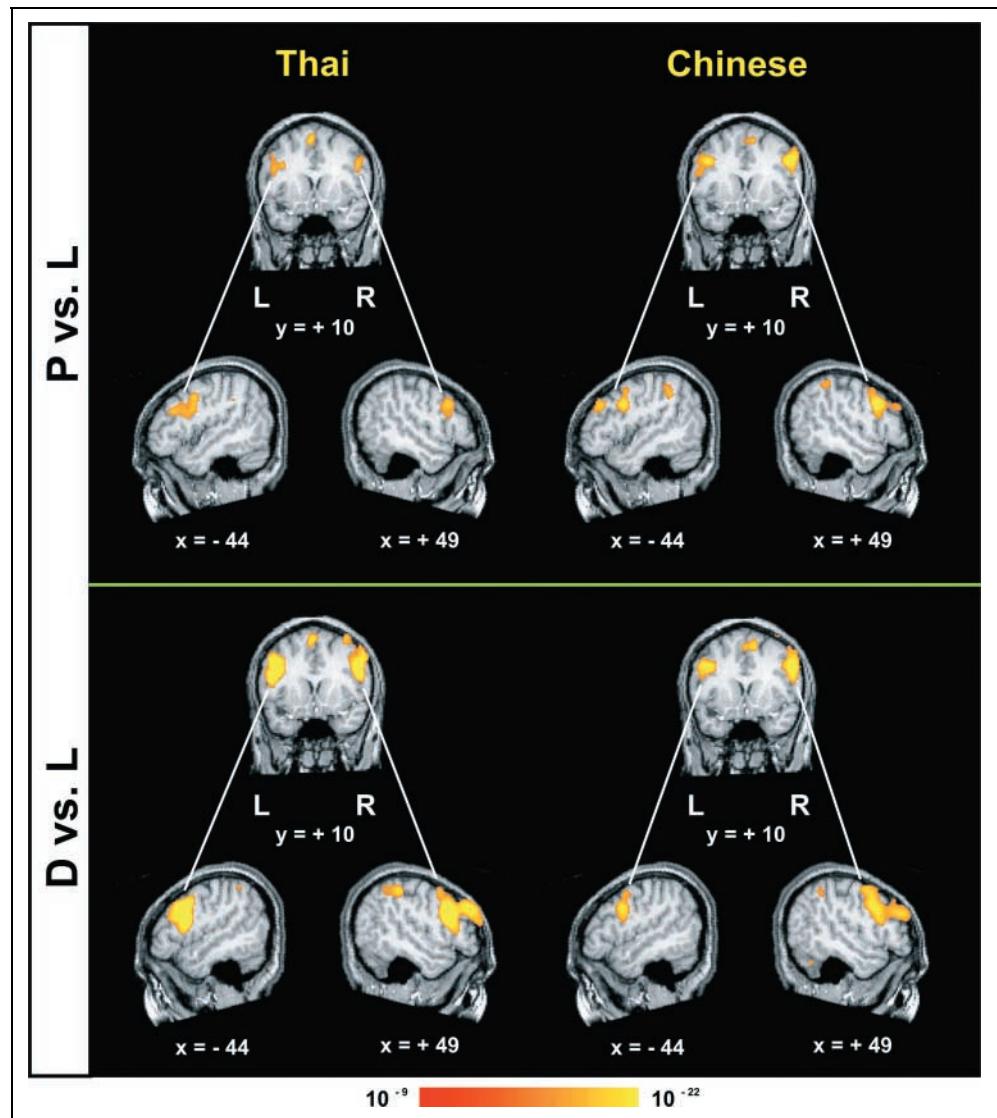
Table 2. Significant Activation Foci in Frontal, Temporal, and Parietal Regions for Task Comparisons within Groups

Region	Thai					Chinese				
	<i>x</i>	<i>y</i>	<i>z</i>	Peak <i>t</i> Value	Extent	<i>x</i>	<i>y</i>	<i>z</i>	Peak <i>t</i> Value	Extent
<i>T versus P: speech versus nonspeech</i>										
Frontal										
L inferior prefrontal	-46	16	37	11.29	10.90					
Temporal										
L anterior STG/STS						-59	-4	3	8.99	3.26
<i>VL versus D: speech versus nonspeech</i>										
Frontal										
L middle frontal gyrus	-43	28	24	10.47	6.45					
Parietal										
L intraparietal sulcus	-22	-70	44	12.31	7.90					
Temporal										
L anterior STG ^a	-49	-7	0	9.02	3.01					
L posterior STG/STS	-56	-25	3	8.67	0.95	-59	-21	10	8.04	3.29
R anterior STG/STS						65	-8	-2	7.68	0.97
L fusiform gyrus	-41	-52	-1	8.92	2.95					
<i>P versus L: nonspeech versus passive listening</i>										
Frontal										
M medial frontal gyrus	-4	11	53	9.22	1.57	-6	4	62	9.43	2.69
L inferior prefrontal	-39	3	38	9.53	10.05	-40	6	31	10.97	8.05
R inferior prefrontal	48	13	31	8.47	2.12	47	10	31	11.69	7.56
Parietal										
L intraparietal sulcus						-44	-41	44	7.72	1.30
R intraparietal sulcus						47	-44	51	7.23	1.00
<i>D versus L: nonspeech versus passive listening</i>										
Frontal										
M medial frontal gyrus	-3	12	57	10.01	3.93	8	14	50	10.22	5.98
L inferior prefrontal	-44	14	25	15.13	18.73	-41	9	31	11.83	7.08
R inferior prefrontal	49	17	23	15.03	26.79	52	13	31	11.44	19.05
Parietal										
L intraparietal sulcus	-27	-58	41	8.78	3.37	-37	-42	39	6.56	0.35
R intraparietal sulcus	48	-35	49	9.35	5.23	36	-58	43	7.42	1.83

Stereotaxic coordinates (mm) are derived from the human brain atlas of Talairach and Tournoux (1988) and refer to the peak *t* value for each region. Extent (ml) refers to the size of activation (number of voxels > threshold). *x* = distance (mm) to right (+) or left (-) of the midsagittal plane; *y* = distance anterior (+) or posterior (-) to vertical plane through the anterior commissure; *z* = distance above (+) or below (-) the intercommissural (AC-PC) line; L = left; R = right; M = at or near midline (*x* coordinates within ± 8 mm); STG = superior temporal gyrus; STS = superior temporal sulcus; T = tone; VL = vowel length; P = pitch; D = duration; L = passive listening.

^aAlthough the peak focus of activation for the Chinese group was observed in the posterior STG, this focus extended in a continuous swath rostrally into the anterior STG.

Figure 2. Averaged fMRI activation maps for the groups of native Thai and Chinese speakers obtained by comparing pitch and duration judgments in a nonspeech context relative to passive listening to nonspeech hums. The top panel shows brain activations for the contrast between pitch and passive listening (P vs. L); the bottom panel for the contrast between duration and passive listening (D vs. L). Each panel shows a coronal section and left/right sagittal sections for each language group. White lines connect inferior prefrontal cortex in coronal sections with corresponding sites in left and right sagittal sections. Across nonspeech tasks, bilateral foci in the inferior frontal region are evident in both language groups, in addition to activation of the SMA. The activation map threshold is at $t(\alpha) = 5.69$, $p < 6.5 \times 10^{-9}$ (one-tailed, uncorrected). See also caption to Figure 1.



between task and group, $F(3,72) = 4.00$, $p < .0108$. Tukey comparisons indicated that the vowel length task was easier for Thai listeners (2.2) than for Chinese listeners (3.9). Response times yielded a main effect of group only, $F(1,72) = 6.84$, $p < .0108$. Pooling across tasks, response times for Thai subjects (588 msec) were longer than those for Chinese subjects (506 msec).

Speech versus Nonspeech

In a comparison of pitch judgments of speech relative to nonspeech (T vs. P), frontal lobe activation was observed in the Thai group only. This left-sided inferior prefrontal activation exhibited a peak focus in the middle frontal gyrus near the junction of the precentral sulcus and inferior frontal sulcus, extending ventrally into dorsal and posterior parts of the inferior frontal gyrus (Brodmann's area, BA, 6/44–45; Figure 1; Table 2). This focus extended posteriorly into the precentral sulcus and anteriorly into dorsolateral prefrontal cortex (BA 9). In contrast, the Chinese group showed only temporal lobe

activation in the left anterior superior temporal gyrus (STG, BA 22).

In a comparison of duration judgments of speech relative to nonspeech (vowel length vs. duration, VL vs. D), inferior prefrontal activation in the LH (BA 6/9/44) was also observed only in the Thai group similar to that for T versus P. Additional LH activation unique to the Thai group included a parietal focus in the intraparietal sulcus (BA 7/19) and a temporal focus in the fusiform gyrus (BA 37/19) that spread anteriorly into the middle temporal gyrus and posteriorly into the medial occipital gyrus. Both groups showed temporal lobe activation in left-sided anterior and posterior regions of the STG/superior temporal sulcus (STS, BA 22/42; Table 2), whereas a right-sided area of activation in the posterior STG/STS (BA 22) was observed in the Chinese group only.

Nonspeech versus Passive Listening

A comparison of pitch judgments of hums relative to passive listening (P vs. L) yielded frontal activation

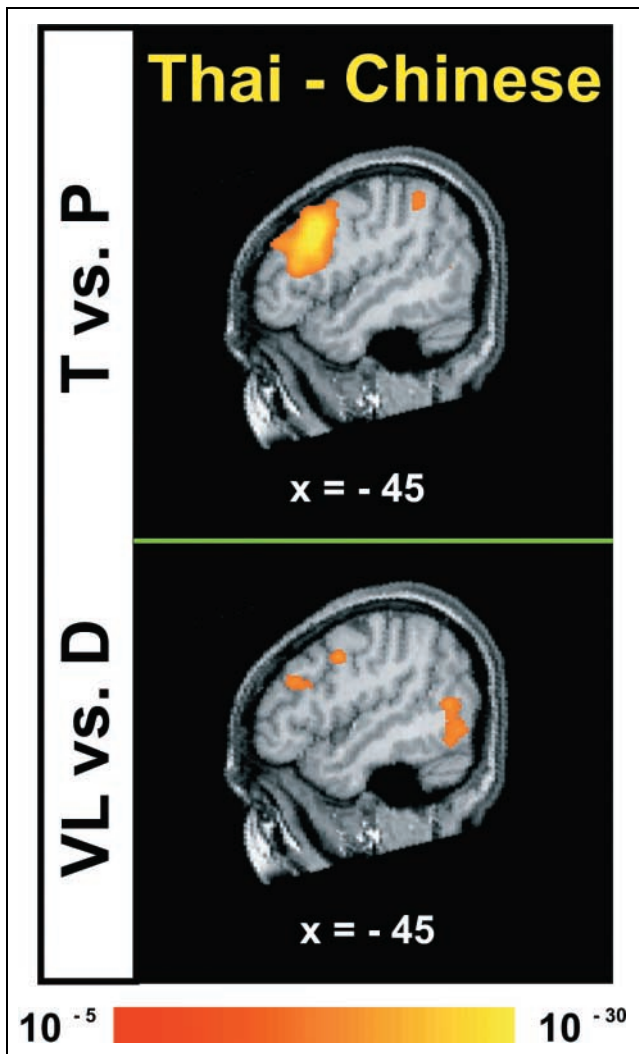


Figure 3. Averaged fMRI between-group activation maps resulting from a direct comparison between the Thai and Chinese groups of pitch and duration patterns in a speech context relative to a nonspeech context. The top panel shows between-group differences in brain activation for the spectral contrast between Thai tones and pitch (T vs. P); the bottom panel for the temporal contrast between Thai vowel length and duration (VL vs. D). Left-sided activation foci are seen in left frontal, parietal, and temporo-occipital regions. The activation map threshold is at $t(\alpha) = 3.80, p < 7.3 \times 10^{-5}$ (one-tailed, uncorrected). See also caption to Figure 1.

bilaterally in the inferior prefrontal cortex (BA 6/9/44) for both Thai and Chinese groups (Figure 2; Table 2). The RH focus extended ventromedially into the anterior insula. Both groups showed additional activation in the medial frontal cortex involving the supplementary motor area (SMA) and pre-SMA regions (BA 6). Only the Chinese group showed parietal activation bilaterally in the supramarginal gyrus (BA 40) near the intraparietal sulcus. Subcortical activations were observed only in the Chinese group in the thalamus bilaterally (Talairach peak coordinates x, y, z : $-9, -12, 14$; $10, -12, 14$). Cerebellar activations were also observed in the Chinese group only involving vermal

(x, y, z : $3, -72, -11$) and left hemispheric (x, y, z : $-24, -66, -21$) regions.

In a comparison of duration judgments of hums relative to passive listening (D vs. L), both groups exhibited a similar frontal activation pattern comprised of foci in the inferior prefrontal cortex (BA 6/9/44) and medial frontal cortex (BA 6) (Figure 2; Table 2). This frontal activation pattern was similar to that observed for P versus L. Moreover, both groups showed parietal activation bilaterally in the supramarginal gyrus near the intraparietal sulcus (BA 40/7). Subcortical activation was observed bilaterally in the thalamus in the Chinese group (x, y, z : $-9, -13, 17$; $9, -15, 15$). Cerebellar activations in the Chinese group were extensive, involving the vermis/paravermis (x, y, z : $4, -74, 18$) and both hemispheric regions (x, y, z : $-26, -66, -22$; $36, -60, -22$). The Thai group showed a cerebellar activation circumscribed to the paravermis (x, y, z : $8, -72, -11$).

Direct Comparison of Tone versus Vowel Length within Groups

A direct comparison of tone and vowel length (T vs. VL) revealed no significantly activated areas for either the Thai or the Chinese group. The reverse comparison (VL vs. T) revealed activity for the Chinese group in the right inferior prefrontal cortex ($x = 50, y = 19, z = 21$; BA 6/9/44).

Table 3. Significant Activation Foci for Direct Comparison between Thai and Chinese Groups

Region	Thai-Chinese				
	<i>x</i>	<i>y</i>	<i>z</i>	Peak <i>t</i> Value	Extent
<i>T versus P: speech versus nonspeech</i>					
Frontal					
L inferior prefrontal	-46	16	37	9.36	15.59
R inferior prefrontal	45	15	29	5.21	1.93
Parietal					
L supramarginal gyrus	-40	-47	52	5.09	1.81
<i>VL versus D: speech versus nonspeech</i>					
Frontal					
L middle frontal gyrus	-44	26	25	5.14	0.44
L precentral gyrus	-45	3	38	5.42	0.78
Temporal					
L fusiform gyrus	-41	-52	-1	5.64	3.81

See note to Table 2.

Between-Group Comparison of Speech versus Nonspeech

A direct comparison between the Thai and Chinese groups of pitch judgments of speech relative to nonspeech (T vs. P) and duration judgments of speech relative to nonspeech (VL vs. D) confirms the robustness of activation foci observed in left-sided fronto-parieto-temporal regions within the Thai group (Figure 3; Table 3).

An intrasubject analysis confirmed that the between-group effects in the left inferior prefrontal cortex were not due to a few subjects dominating the group map. For the T versus P comparison, eight Thai subjects showed significant activation ($p < .01$) in a region at or near the junction of the inferior frontal and precentral sulci versus only two Chinese subjects. Similarly, for the VL versus D comparison, 9 Thai subjects showed significant activation in this region versus only five of the Chinese subjects.

DISCUSSION

Hemispheric Specialization for Speech Prosody

A controversial issue is whether hemispheric lateralization of speech processing is speech-specific or whether functional asymmetries can be accounted for by computational demands evoked by complex auditory signals that contain critical information in the spectral domain or time domain. Although these competing approaches are not necessarily mutually exclusive, our findings clearly indicate that hemispheric differences in speech perception are sensitive to higher order domain-specific factors. It may well be the case that the spectral/temporal hemispheric differences are characteristic only of low-level auditory processing, and that speech sounds are transmitted to higher levels of processing, domain-specific knowledge of the language comes into play. These domain-specific effects have consistently implicated frontal lobe structures rather than the auditory cortex (Hsieh et al., 2001; Gandour, Wong, et al., 2000; Gandour et al., in press). Similarly, in this study, left inferior prefrontal cortex is activated in speech minus nonspeech contrasts, regardless of whether the processing is in the spectral or temporal domain (T vs. P, VL vs. D) for the Thai group only. Auditory stimuli for speech and nonspeech conditions were closely matched in terms of prosodic features; task demands were comparable as far as attention and working memory is concerned. No matter whether the phonological contrast is signaled primarily by spectral variation (Thai tones) or temporal variation (Thai vowel length), the LH appears to be dominant in processing phonologically relevant distinctions in a listener's native language.

The role of higher order, language-specific factors in hemispheric specialization is reinforced by comparing brain activation patterns in Thai and Chinese

listeners when judging the same acoustic cues in a nonspeech context. The Thai and Chinese groups indeed exhibit similar fronto-parietal activation patterns when judging nonspeech hums for either spectral or temporal cue (P vs. L, D vs. L). When the stimuli are no longer perceived as speech, the language-specific effects disappear.

Functional specialization for speech processing appears to involve more than simply the phonological significance of auditory cues in a listener's native language. With respect to prosody, differences in hemispheric function may be related to the size of temporal integration windows over which prosodic elements are encoded in the speech signal. For example, Chinese tone and intonation span over relatively shorter (local) and longer (global) temporal domains, respectively. It has been demonstrated that Chinese listeners, in contrast to English listeners, process pitch contours in the LH when they are associated with lexical tones, but in the RH when they are associated with intonation (Gandour, Dziedzic, et al., in press). Cross-linguistic differences between Chinese and English groups suggest that these temporal integration windows reflect higher order mechanisms that are not necessarily tied to absolute ranges along a timing continuum. Instead, these temporal integration windows appear to be defined relative to local and global aspects of speech prosody of a particular language. In this study, both tone and vowel length contrasts, defined at the syllable level, reflect processing of local aspects of speech prosody. Thus, it is not surprising that Thai listeners show prefrontal activation lateralized to the LH when processing both tone and vowel length in their native language.

The type of acoustic cue underlying a phonological contrast does not appear to influence hemispheric specialization. What is important is whether a particular cue serves a phonological function in a listener's native language. Like Thai, the Chinese language exploits variations in F_0 to signal tonal contrasts. Unlike Thai, temporal variations in vowels are not exploited phonologically. Despite experience with lexically significant F_0 contours in Chinese, the Chinese group still fails to show LH lateralization for Thai tone (Gandour, Wong, et al., 2000). However, when judging Chinese tones, Chinese listeners show LH lateralization as predicted by the speech-specific hypothesis (Hsieh et al., 2001). Unique LH lateralization for Thai vowel length in the Thai group is consistent with an earlier fMRI study in which English listeners served as the control group (Gandour et al., in press).

Neural Circuitry Involved in Phonological Processing

Overt segmentation of speech that is required to perform phonetic discrimination tasks results in activation of posterodorsal aspects of the inferior prefrontal cortex

(Burton, Small, & Blumstein, 2000). This process entails articulatory recoding that occurs only when subjects are required to extract segmental information from the continuous acoustic signal and make phonetic judgments about speech sounds; for example, English consonants (Burton et al., 2000; Zatorre, Evans, Meyer, & Gjedde, 1992). Compatible data have been reported for suprasegmental pitch features associated with Thai (Gandour, Wong, et al., 2000) and Chinese tones (Hsieh et al., 2001), and suprasegmental duration features associated with Thai vowel length (Gandour et al., in press). Our current findings on Thai tone and vowel length provide additional support for this notion of speech segmentation. Such findings conflict with the view that speech production is mediated exclusively by anterior cortical structures and speech perception by posterior.

It is hypothesized that segmentation depends on the number of phonetic features distinguishing the paired items (Burton et al., 2000). A direct test of the hypothesis requires that nonminimal pairs (e.g., *dip–doom* pairs) be compared directly to minimal pairs (e.g., *dip–tip* pairs). Burton et al. (2000) showed that tasks that require segmentation of the initial consonant (e.g., *dip–doom* pairs) activate inferior prefrontal cortex, whereas those that do not require segmentation (e.g., *dip–tip* pairs) do not. Their study, however, employed two different groups of subjects for the two different segmentation conditions. In addition, their control condition was not equal in acoustic complexity to that of the experimental condition. In the current study, all stimulus pairs that were designed to elicit a different judgment differed by one feature only (e.g., *luum^{rising}/luum^{high}*, *maan^{high}/man^{high}*). Yet, we find left prefrontal activation in Thai listeners for both tone and vowel length minimal pairs. It is possible that the notion of speech segmentation needs to be reexamined or that the extraction of suprasegmental units differs from that of segmental (cf. Hsieh et al., 2001). Although our experimental design does not permit us to tease apart stages of processing, it appears likely that cross-linguistic differences in inferior prefrontal activation reflects higher level processes implicated in extraction of phonological elements.

Temporal lobe activity in this study appears to reflect acoustic phonetic processing of speech sounds. Comparisons of pitch (P vs. L) or duration (D vs. L) in nonspeech sounds relative to passive listening did not elicit any significant temporal foci for either the Thai or Chinese group. This would suggest that no additional regions in the temporal lobe are activated beyond those already engaged for lower level auditory processing in the passive listening baseline. This putative lower level stage of processing is supported by the absence of a language group effect. Interestingly, the Chinese group shows an anterior focus in the STG of the LH, instead of the RH, for the tone comparison in speech sounds

relative to nonspeech sounds (T vs. P). This latter finding is hard to reconcile with a spectral–temporal bias for hemispheric specialization of auditory processing that would predict RH lateralization of Chinese listeners' processing of Thai tones. However, it is consistent with previous work that has demonstrated speech-specific areas located in the anterior and posterior parts of the left STG (Giraud & Price, 2001).

Locating short-term storage of phonological information in verbal working memory remains an empirical question (Becker, MacAndrew, & Fiez, 1999). According to Baddeley's (1992) model of working memory, verbal storage consists of (1) a phonological buffer for short-term maintenance of phonological information and (2) a subvocal rehearsal process that refreshes the contents of the buffer. The first subcomponent is believed to be mediated by parietal sites, the latter by frontal (Smith & Jonides, 1999). Our findings suggest that this storage buffer is sensitive to attention-modulated processing of suprasegmental information and its functional role in a particular language (cf. Shaywitz et al., 2001). Selective attention to durational distinctions in Thai speech results in LH sites in the intraparietal sulcus and adjacent inferior and superior parietal lobules for the Thai group (VL vs. D). In contrast, the Chinese group shows no significant parietal activity to either Thai tones or vowel length. Yet, the Chinese group shows parietal activation bilaterally when comparing pitch (P vs. L) and duration (D vs. L) properties of nonspeech hums. Similarly, parietal lobe activity is bilateral for the Thai group's duration comparisons of nonspeech hums (D vs. L). Collectively, these findings reinforce the idea that neural circuitry for the verbal storage system is influenced by the modality of the information stored (speech vs. nonspeech).

The absence of any language group effect for the direct comparison of tone and vowel length (T vs. VL) suggests that no additional neural mechanisms are recruited to perform the tone judgments regardless of language background. The reverse comparison (VL vs. T), however, shows activation in the right inferior prefrontal cortex for the Chinese group. This right-sided frontal activation may reflect additional processing associated with attention or working memory. Word-final nasals and glides count as part of the coda in Thai, whereas in Chinese they are part of the nucleus. For Chinese listeners, Thai vowel-nasal and vowel-glide sequences are perceived indissolubly as a complex rhyme, thus making it difficult for them to isolate the vowel from nasal.

Functional neuroimaging studies have suggested that the left ventral occipito-temporal cortex (BA 37) may play a role in phonological processing (e.g., Rumsey et al., 1997). In this study, the left fusiform gyrus is activated in the VL versus D comparison for the Thai group only. These occipito-temporal activations are usually interpreted to reflect visualization processes

associated with the performance of language tasks. Perhaps, this task elicited some visualization of Thai script in spite of the fact that the stimuli were pseudowords. Thai script clearly distinguishes between short and long vowels. Thus, the absence of any activation in this region for Chinese listeners is predictable since they are unfamiliar with Thai script. More research obviously needs to be done to elucidate the role of this region in language processing.

Effects of Task Performance on Brain Activation Patterns

It is unlikely that differences in task performance between Thai and Chinese groups can account for their differential patterns of activation in the left prefrontal cortex. As measured by response accuracy and self-ratings of task difficulty, Thai and Chinese groups performed alike on the tone (T) task, yet only the Thai group shows left-sided activation in prefrontal regions. These two groups also performed similarly on the tone (T) and nonspeech hum tasks (P, D). Notwithstanding, the Chinese group exhibits similar brain activation patterns as the Thai group for nonspeech tasks only. The Chinese group did perform at a significantly lower level of proficiency than the Thai group on the VL task. This cross-linguistic difference in performance is presumably due to structural differences in the rhyme constituent of Chinese and Thai syllables.

Conclusions

This study clearly demonstrates the importance of abstract, prosodic features in distinguishing between brain processing of speech and nonspeech auditory signals. In particular, cross-linguistic differences in brain activity associated with Thai tones and vowel length indicate that hemispheric specialization is sensitive to language-specific, prosodic functions in addition to low-level acoustic features. Precisely how these higher order mechanisms interact with the low-level perceptual input stage remains a challenge for future research.

METHODS

Subjects

Ten adult Thai (5 men and 5 women) and 10 adult Chinese (3 men and 7 women) speakers were closely matched in age (Thai: $M = 25.8$, $SD = 4.05$; Chinese: $M = 25.4$, $SD = 2.8$) and years of formal education (Thai: $M = 17.6$, $SD = 2.1$; Chinese: $M = 18.8$, $SD = 2.1$). All subjects were strongly right-handed as measured by the Edinburgh Handedness Inventory (Oldfield, 1971) and exhibited normal hearing sensitivity at frequencies of 0.5, 1, 2, and 4 kHz. All subjects gave informed consent in compliance with a protocol approved by the Institutional Review Board of Indiana

University, Purdue University Indianapolis, and Clarian Health.

Languages

Both tone and vowel length oppositions occur in Thai, whereas only tonal oppositions occur in Chinese. Duration is distinctive for all nine Thai vowels (Tingsabhadh & Abramson, 1993). Acoustically, this duration distinction is preserved in stressed syllables across a variety of speaking conditions (Abramson, 2001; Potisuk, Gandour, & Harper, 1998). Perceptually, duration has been shown to be the primary cue in signaling the contrast between short- and long-vowel phonemes (Abramson & Ren, 1990; Abramson, 1962). Vowel duration, on the other hand, is not distinctive in Chinese.

The primary acoustic correlate of Thai tones is voice fundamental frequency (F_0) (Abramson, 1962). 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. Perceptually, F_0 has been demonstrated to be the primary acoustic correlate of Thai tones (Abramson, 1962, 1978). Chinese is also a tone language. Yet, it differs from Thai in both the number and shape of its lexical tones (Howie, 1976).

Stimuli

Stimuli consisted of 40 pairs of monosyllabic Thai pseudowords and 40 pairs of “hummed” versions of the same words for the speech condition and nonspeech condition, respectively. By using pseudowords, we minimized lexically induced semantic category or word frequency effects, while maximizing subjects’ focused attention on phonetic characteristics of the stimuli. Stimulus pairs were of three subtypes in both speech and nonspeech conditions: same vowel duration, different pitch pattern (18 minimal pairs; e.g., *tooŋ^{mid}/tooŋ^{rising}*); same pitch pattern, different vowel duration (18 minimal pairs; e.g., *pum^{falling}/puum^{falling}*); same vowel duration, same pitch pattern (4 identical pairs; e.g., *doon^{falling}/doon^{falling}*). Syllable structures were either CVC or CVVC, where the final consonant was either a nasal (m n ŋ) or a glide (j w).

The aim of using hummed stimuli was to eliminate segmental information (i.e., consonant, vowel) while at the same time preserving suprasegmental information (i.e., duration, pitch). Hums were derived from resynthesized speech (Praat) by using six formants (bandwidths): 600 (50), 1400 (100), 2400 (200), 3400 (300), 4500 (400), and 5500 (500) Hz. In the Thai speech stimuli, long vowels were, on average, about twice as long as short vowels (VV: $M = 262$ msec, $SD = 38$; V: $M = 129$ msec, $SD = 22$). In the hummed stimuli, the absolute difference in duration between the short and long hums was adjusted to 175 msec for all different

vowel duration pairs. These adjustments resulted in long-short hum ratios ranging from 1.5 to 1.7. Hums corresponding to the same vowel duration pairs were made equal in duration via time expansion or compression. Fundamental frequency and intensity characteristics of the original Thai pseudowords were preserved in their hummed counterparts. As judged by three Thai listeners, none of the hummed stimuli contained any recognizable Thai words or parts thereof. All stimuli were 16-bit digital sound files sampled at 44.1 kHz.

Stimuli in the vowel length and tone tasks were identical, thus controlling for differences in word length or differences in frequency of occurrence of consonants, vowels, or tones (Prathanee, 1998; Gandour & Gandour, 1982). No pseudoword appears in more than one pair. No pair was judged more than once during a scanning run, thus eliminating confounding effects of prior exposure. Similarly, nonspeech stimuli were identical across tasks.

Tasks

The fMRI paradigm consisted of two tasks that required subjects to make overt comparisons involving Thai speech stimuli (T = Thai tones; VL = Thai vowel length) and two tasks involving nonspeech hums (P = pitch

contours; D = duration) (Table 4). In these four tasks, subjects were instructed to direct their attention to particular features of auditory stimuli, make discrimination judgments, and respond by pressing a mouse button (left = same, right = different). In the remaining task, subjects were instructed to listen passively to nonspeech hums (L = passive listening), make no comparisons of the paired stimuli, and respond by alternately pressing the mouse buttons after each trial.

The speech tasks (T, VL) were designed to elicit phonetic processing with the goal of identifying brain areas involved in perceptual processing of spectral and temporal information associated with linguistically significant elements. The contrasts T versus P and VL versus D allowed us to compare discrimination judgments of pitch and duration patterns, respectively, in a speech context relative to a nonspeech context. The P task was used as a baseline condition for removing the effects of selective attention to pitch patterns in a nonspeech context relative to those in a speech context (T vs. P). The D task was used as a baseline condition for removing the effects of selective attention to duration patterns in a nonspeech context relative to those in a speech context (VL vs. D). The passive listening task (L) was designed to capture cognitive processes inherent to automatic, perceptual analysis including early auditory processing, executive functions mediating sustained attention and arousal, working memory, and motor response formation. Thus, we were able to compare selective attention to pitch (P vs. L) and duration (D vs. L) processing in nonspeech hums relative to automatic, unfocused attention to the same stimuli.

All speech and nonspeech stimuli were digitally edited to have equal maximum energy level in decibel sound pressure level (dB SPL). Auditory stimuli were presented binaurally using a computer playback system (Psychology Software Tools) and a pneumatic-based audio system (Avotec). The plastic sound conduction tubes were threaded through tightly occlusive foam eartips inside the earmuffs that attenuated the average SPL of the continuous scanner noise by ~30 dB. Average intensity of all experimental stimuli was 95 dB SPL and was matched across all conditions. Average intensity of the scanner noise was 80-dB SPL after attenuation by the earmuffs.

A scanning run consisted of two tasks presented in blocked format (32 sec) alternating with 16-sec rest periods. Each block contained 10 trials. The order of scanning runs and trials within blocks were randomized for each subject. Instructions were delivered to subjects in their native language via headphones during rest periods immediately preceding each task: “listen” for passive listening to nonspeech stimuli (L), “duration” for same-different judgments on Thai vowels (VL) and nonspeech hums (D), “pitch” for same-different judgments on Thai tones (T) and hums (P). Average trial duration was about 3.2 sec, including an interstimulus interval of 200 msec and a response interval of 2 sec.

Table 4. fMRI Paradigm

Task	Stimuli	Example	Response	
Pitch (tone)	Thai	p ^h in ^R	p ^h iin ^R	same
		haaj ^M	haaj ^H	different
		kaŋ ^L	kaaŋ ^L	same
		hiŋ ^L	hiŋ ^R	different
Pitch	Hums	□ ^R	□ □ ^R	same
		□ □ ^M	□ □ ^H	different
		□ ^L	□ □ ^L	same
		□ ^L	□ ^R	different
Duration (vowel length)	Thai	p ^h in ^R	p ^h iin ^R	different
		haaj ^M	haaj ^H	same
		kaŋ ^L	kaaŋ ^L	different
		hiŋ ^L	hiŋ ^R	same
Duration	Hums	□ ^R	□ □ ^R	different
		□ □ ^M	□ □ ^H	same
		□ ^L	□ □ ^L	different
		□ ^L	□ ^R	same

Thai tones: M = mid; L = low; F = falling; H = high; R = rising. Thai vowel length: single vowel letter = short; double vowel letter = long. Hum duration: □ = short; □ □ = long.

Total stimulus duration (excluding the response intervals) during a 32-sec block was about 12 sec.

Accuracy, reaction time, and subjective ratings of task difficulty were collected to measure task performance. Before scanning, Thai subjects were trained to a high level of accuracy on all tasks: tone ($M = 96.7$, $SD = 3.1$), vowel length ($M = 97.0$, $SD = 3.4$), pitch ($M = 93.8$, $SD = 3.8$), and duration ($M = 86.4$, $SD = 5.9$). Chinese subjects were similarly trained to a high level of accuracy on the tone ($M = 93.7$, $SD = 6.3$), pitch ($M = 94.6$, $SD = 4.0$), and duration ($M = 88.0$, $SD = 3.1$) tasks, and to a moderately high level of accuracy on the vowel length ($M = 83.5$, $SD = 4.6$).

Image Acquisition

Scanning was done on a 1.5-T Signa GE scanner (Waukesha, WI) equipped with transmit-receive radio-frequency coils optimized for whole-brain echo-planar imaging (EPI). Each 200-volume EPI series was begun with 16 baseline volumes (32 sec) followed by 184 volumes during which the two comparison conditions (32 sec) alternated with intervening 16-sec rest periods. Axially oriented EPI images were acquired using a gradient-echo pulse sequence with the following image parameters: repetition time/echo time (TR/TE) 2 sec/50 msec; matrix 64×64 ; flip angle (FA) 90° ; 24×24 cm field of view (FOV); receiver bandwidth 125 kHz. Fifteen 7-mm thick axial slices with a 2-mm interslice gap were required to image the entire brain. The initial 11 volumes were discarded to account for presaturation and hemodynamic delay effects (Bandettini, Jesmanowicz, Wong, & Hyde, 1993). The final five volumes were also excluded so that the dataset for each volume would consist of exactly 184 time points: four cycles for each of the two conditions and seven rest periods between them.

Subjects were scanned with eyes closed and room lights dimmed. Both head–neck padding and dental bite bar were used to minimize the effects of head motion. Each scan was analyzed for head motion (Woods, Cherry, & Mazziotta, 1992) because of sensitivity to the small magnitude (0.2%) of activation-related BOLD (blood oxygen level dependent) signal changes. All fMRI data were Hamming-filtered spatially, which increased the BOLD contrast-to-noise ratio with only a small loss of spatial resolution (Lowe & Sorenson, 1997). Data with maximal peak-to-peak displacement greater than 0.15 mm were rejected because image registration algorithms do not completely remove motion effects (Jiang et al., 1995). Data for 2 out of 22 subjects were excluded due to excessive head motion. Pooling across scanning runs, average peak-to-peak displacement was 0.01 mm for 3 subjects, 0.02 mm for 13 subjects, 0.03 mm for 3 subjects, and 0.04 mm for 1 subject. Prior to functional imaging scans, high-resolution anatomic images were acquired in 124 contiguous axial slices using a 3-D Spoiled-Grass (3-D SPGR) sequence (slice thickness

1.1–1.2 mm; TR/TE 35/8 msec; number of excitations 1; 30° FA; matrix 256×128 ; FOV 24×24 cm; receiver bandwidth 32 kHz) for anatomic localization and coregistration. Total duration of the imaging session was about 80 min. In all cases, functional scanning commenced within approximately 20 min of the start of the imaging session.

fMRI Data Analysis

In order to analyze the acquired data for task-dependent changes, it was necessary to construct a reference function that allowed a comparison of the tasks of interest. In each fMRI scan, there were blocks of three states: rest, state 1, and state 2. For example, one fMRI scan had a rest condition in 16-sec blocks, VL in 32-sec blocks, and D in 32-sec blocks. For all scans, rest and states 1 and 2 were interleaved in the following manner: 16-sec rest, 32-sec state 1, 16-sec rest, 32-sec state 2, 16-sec rest, 32-sec state 1, 16-sec rest, 32-sec state 2, 16-sec rest, 32-sec state 1, 16-sec rest, 32-sec state 2, 16-sec rest, 32-sec state 1, 16-sec rest, 32-sec state 2, 16-sec rest. This results in 200 images of each slice. A reference function is constructed for each task comparison of interest. To make the comparison (state 1–state 2), the reference function was defined as:

$$\text{reference} = 1i = [12, 28], [61, 76], [109, 124], [157, 172]$$

$$= -1i = [37, 52], [85, 100], [133, 148], [181, 196]$$

where i is the index of the i -th-acquired volume (first volume: $i = 1$). The least squares method described in Lowe and Russell (1999) was used to calculate a Student's t statistic value and percent signal change for each task comparison by comparing the derived reference function to the acquired data. This fit gave the amplitude (A) of activation-related signal change and baseline intensity (BI), with the relative signal change $S = A/BI$.

Individual whole brain statistical maps were interpolated to $256 \times 256 \times 256$ cubic voxels (0.9375 mm/side). Individual anatomic images and single-subject interpolated activation maps were projected into a standardized stereotaxic coordinate system (Talairach & Tournoux, 1988). Single-subject activation maps were then summed pixel-by-pixel, combined into within-group activation maps, and displayed on anatomic images from a representative subject. Stereotaxic location of activation peaks and extent of activation were identified by drawing regions-of-interest around activation foci at Student's t statistic thresholds (one-tailed, uncorrected) of $t(\alpha) = 5.85$, $p < 2.54 \times 10^{-9}$ for all comparisons (T vs. P, VL vs. D, P vs. L, D vs. L, T vs. VL).

Acknowledgments

Funding was provided by research grants from the National Institutes of Health R01DC04584-02 and James S. McDonnell Foundation.

Reprint requests should be sent to Jack Gandour, Department of Audiology and Speech Sciences, Purdue University, Heavilon Hall, West Lafayette, IN 47907-1353, USA, or via e-mail: gandour@purdue.edu.

The data reported in this experiment have been deposited in The fMRI Data Center (<http://www.fmridc.org>). The accession number is 2-2002-1131E.

REFERENCES

- Abramson, A. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments* (vol. 20). Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics.
- Abramson, A. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, 21, 319–325.
- Abramson, A. (2001). The stability of distinctive vowel length in Thai. In K. Tingsabadh, & A. Abramson (Eds.), *Essays in Tai linguistics* (pp. 13–26). Bangkok: Chulalongkorn University Press.
- Abramson, A., & Ren, N. (1990). Distinctive vowel length: Duration vs. spectrum. *Journal of Phonetics*, 18, 79–92.
- Baddeley, A. (1992). Working memory. *Science*, 255, 556–559.
- Bandettini, P. A., Jesmanowicz, A., Wong, E. C., & Hyde, J. S. (1993). Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine*, 30, 161–173.
- Becker, J., MacAndrew, D., & Fiez, J. (1999). A comment on the functional localization of the phonological storage subsystem of working memory. *Brain and Cognition*, 41, 27–38.
- Benson, R., Whalen, D., Richardson, M., Swainson, B., Clark, V., Lai, S., & Liberman, A. M. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language*, 78, 364–396.
- Burton, M., Small, S., & Blumstein, S. (2000). The role of segmentation in phonological processing: An fMRI investigation. *Journal of Cognitive Neuroscience*, 12, 679–690.
- Gandour, J., Dziedzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Sathamnuwong, N., & Lurito, J. (in press). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language*.
- Gandour, J., and Gandour, M. J. (1982). The relative frequency of tones in Thai. In D. Bradley (Ed.), *Papers in Southeast Asian linguistics, tonation* (vol. 8, pp. 155–159). Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University.
- Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, 12, 207–222.
- Gandour, J., Wong, D., Lowe, M., Dziedzic, M., Sathamnuwong, N., Tong, Y., & Lurito, J. (in press). Neural circuitry underlying perception of duration depends on language experience. *Brain and Language*.
- Giraud, A., & Price, C. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience*, 13, 754–765.
- Griffiths, T. D., Buchel, C., Frackowiak, R., & Patterson, R. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, 1, 422–427.
- Griffiths, T. D., Johnsrude, I., Dean, J. L., & Green, G. G. (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *NeuroReport*, 10, 3825–3830.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.
- Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, 76, 227–252.
- Ivry, R., & Robertson, L. (1998). *The two sides of perception*. Cambridge: MIT Press.
- Jiang, A., Kennedy, D., Baker, J., Weiskoff, R., Tootel, R., Woods, R., Benson, R., Kwong, K., Thomas, J., Brady, B., Rosen, B., & Belliveau, J. (1995). Motion detection and correction in functional MRI imaging. *Human Brain Mapping*, 3, 224–235.
- Klein, D., Zatorre, R., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*, 13, 646–653.
- Lowe, M. J., & Russell, D. P. (1999). Treatment of baseline drifts in fMRI time series analysis. *Journal of Computer Assisted Tomography*, 23, 463–473.
- Lowe, M., & Sorenson, J. (1997). Quantitative comparison of functional contrast from BOLD-weighted spin-echo and gradient-echo-planar imaging at 1.5 Tesla and H₂¹⁵O PET in the whole brain. *Magnetic Resonance in Medicine*, 37, 723–729.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 25, 679–693.
- Potsuk, S., Gandour, J., & Harper, M. P. (1998). Vowel length and stress in Thai. *Acta Linguistica Hafniensia*, 30, 39–62.
- Prathane, B. (1998). Frequency of occurrences of sound in general daily life conversation at Khon Kaen University. *Siriraj Hospital Gazette*, 50, 1058–1063.
- Rumsey, J. M., Nace, K., Donohue, B., Wise, D., Maisog, J. M., & Andreason, P. (1997). A positron emission tomographic study of impaired word recognition and phonological processing in dyslexic men. *Archives of Neurology*, 54, 562–573.
- Schwartz, J., & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science*, 207, 1380–1381.
- Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Fullbright, R. K., Skudlarski, P., Mencl, W. E., Constable, R. T., Marchione, K. E., Fletcher, J. M., Klorman, R., Lacadie, C., & Gore, J. C. (2001). The functional neural architecture of components of attention in language-processing tasks. *Neuroimage*, 13, 601–612.
- Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, 283, 1657–1661.
- Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging*. New York: Thieme.
- Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiology of speech perception. *Annals of New York Academy of Science*, 682, 27–47.
- Tingsabadh, K., & Abramson, A. (1993). Thai. *Journal of the International Phonetic Association*, 23, 24–28.
- Woods, R., Cherry, S., & Mazziotta, J. (1992). Rapid automated algorithm for aligning and reslicing PET images. *Journal of Computer Assisted Tomography*, 16, 620–633.
- Zatorre, R., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11, 946–953.
- Zatorre, R., Belin, P., & Penhune, V. B. (2002). Structure and function of the auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6, 37–46.
- Zatorre, R., Evans, A., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846–849.