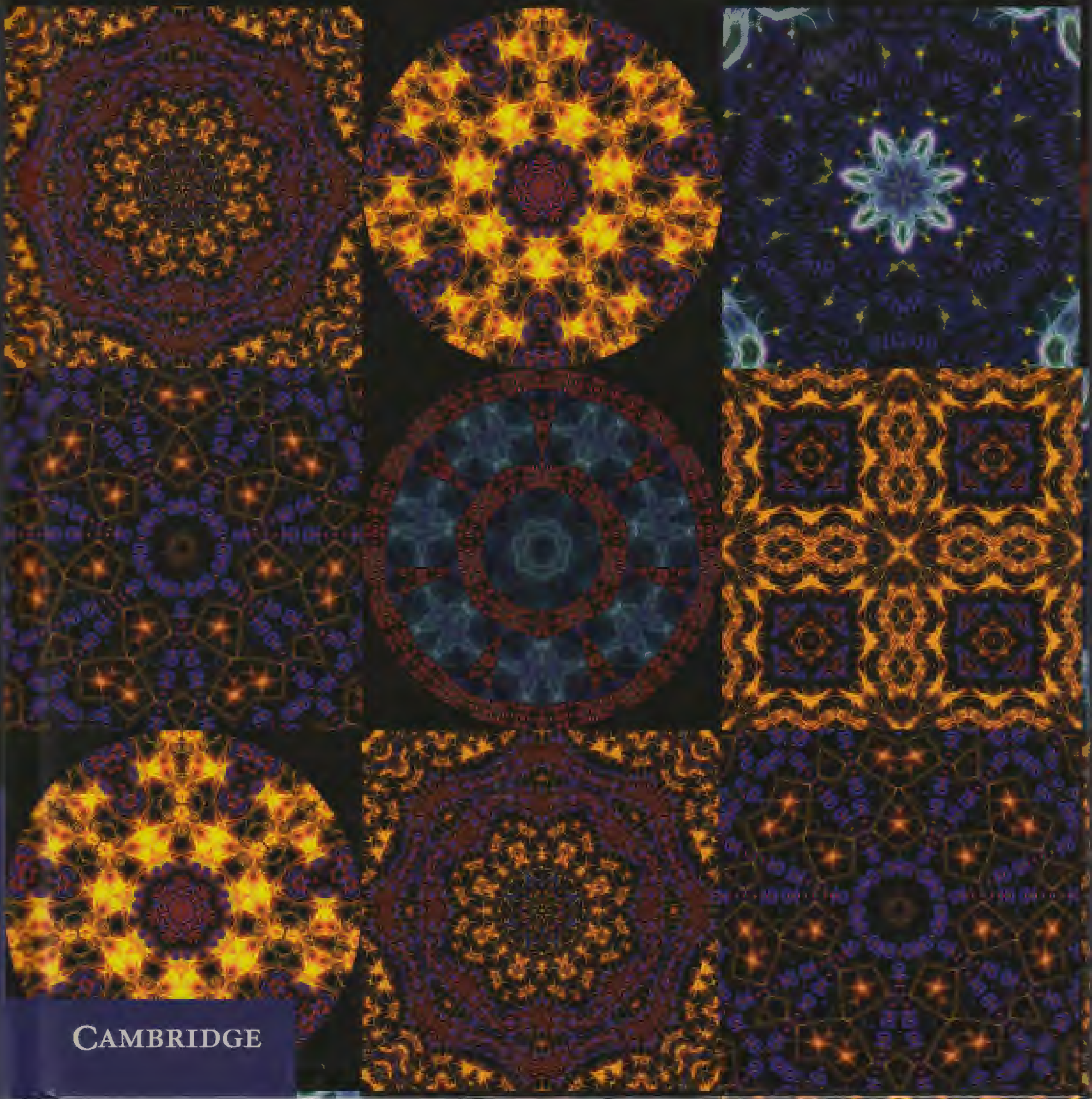
South and Southeast Asian Psycholinguistics

Edited by Heather Winskel and Prakash Padakannaya



CAMBRIDGE

32 Neural bases of lexical tones

Jackson T. Gandour and Ananthanarayan Krishnan

Introduction

This chapter reviews recent findings on the neurolinguistics of speech prosody in tone languages of Southeast Asia. A few caveats are in order from the start. To investigate how language is processed in the human brain, one cannot rely on abstract theoretical models absent of time. The brain is driven by neurophysiology and, as such, our goal must be to focus on how linguistic elements, rules, and representations are *implemented* in the brain. Katz (1964) argued that "it is immaterial whether the mechanism inside the speaker's head is in reality a . . . mechanical system of cardboard flip-flops and rubber bands, or . . . a group of homunculi industriously at work in a tiny office" (p. 129). An opposing view emphasizes that the grammar consists of representations and computations that are executed in the brain in *real time* (Poeppel & Embick, 2006). The primary question for neurolinguistics is how computations and representations are implemented at the appropriate level of biological abstraction. This qualification is necessary because elements and operations are incommensurable between linguistics (e.g., tone, co-articulation) and neuroscience (e.g., neuron, synchronization) at any given level of analysis. With respect to lexical tones, this chapter shows that a complete understanding of the processing of *linguistically relevant features* of the auditory signal can only be achieved within a framework involving a series of computations that apply to *representations* at *different stages of processing* (Hickok & Poeppel, 2004).

In view of this theoretical perspective, the scope of this review is delimited to empirical data on the perception of lexical tones obtained from functional brain recording. The importance of behavioral data notwithstanding, earlier reviews are available on dichotic listening of lexical tones (Gandour, 2007) and tonal breakdown in production and perception subsequent to damage of the cerebral cortex (Gandour, 1987; 1994, 1998a, 1998b, 2006b; Liang, 2008; Wong, 2002; Wong *et al.*, 2009). Both sources of evidence point to the left hemisphere (LH) as dominant for language functions in native speakers of tone languages. Subcortical structures also play a significant role in the production and comprehension of speech prosody (Van Lancker Sidtis *et al.*, 2006). Case studies of subcortical lesion deficits in Thai are available (Gandour & Dechongkit, 1992; Gandour & Ponglorpisit, 1990) as well as a neurobehavioral framework for assessing dysprosody in this clinical population (Sidtis & Van Lancker Sidtis, 2003).

From a theoretical perspective, we present evidence to show that simple dichotomies, either cognitive domains (e.g., language, music) or acoustic cues (e.g., temporal, spectral), are inadequate as the sole explanatory model for patterns of neural specialization at different levels of the brain (Zatorre & Gandour, 2008). Though the extant literature on the neurobiology of tonal processing in Southeast Asian languages is in its infancy, the evidence reviewed herein demonstrates that neural specializations are driven by both low-level acoustic features and higher-order, linguistic knowledge.

Speech perception is important because it provides multiple windows along the auditory pathway into the cerebral cortex on how continuous, acoustic signals are transformed into representations upon which computations are based at different levels of the brain. Pitch is one of the most important information-bearing components of speech. *Tone languages* offer advantages for investigating neural mechanisms underlying pitch at different levels of processing because of its phonemic status at the word level (Yip, 2002). We will focus on *lexical tones* in Thai (Tingsabadh & Abramson, 1999), Cantonese (Zee, 1999), and Mandarin (Howie, 1976). Voice fundamental frequency (f_0) contours provide the dominant cue for tone recognition (Gandour, 1994).

Functional brain recording

Hemodynamic

Recent advances in functional neuroimaging have now made it possible to observe *in vivo* how changes in behavioral and cognitive task demands lead to changes in neural activity in the human brain (Rugg, 1999). Two methods, positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), respectively, measure changes in brain activity as reflected by properties of blood flow. Though poor in temporal resolution, both methods provide high spatial resolution that enables us to map areas at the level of gyri and sulci in the brain.

Though perception of non-linguistic pitch is associated with activation in the right hemisphere (RH) (Zatorre *et al.*, 1992), imaging studies of lexical tones in Thai and Mandarin have revealed dominant activity in the LH. In judging tones of Thai words (Gandour *et al.*, 2000; cf. Gandour, Wong, & Hutchins, 1998), native Thais, but neither Chinese nor English, showed LH

---

activation in the inferi
pitch and timing pattern
(hums) conditions (Gand
the left IFG for both to
'monetary unit'). Regar
signaled by spectral or
native Thais, but not Ch
native Chinese when pre
homologous regions in th
Interestingly, when ask
English words (not ling
that are activated regar
(Figure 32.1) (Wong, Par
ing engages the LH only
cant to the listener; othe
conclude that LH mecha
irrespective of acoustic c
suprasegmental.

To address the issue
hybrid stimuli consisting
(*tonal chimeras*) and M
words) were presented to
comparison of native vs
language groups was ide
double dissociation betwe
pitch occurred such that s
compared to non-native t
prelexical phonological
tively early stages of ac
stimulus features that ar
(Griffiths & Warren, 200

Differential patterns o
language affiliation alone
features associated with
suprasegmental). Product
RH than vowels (Liu e
Mandarin tones, relative
activation in frontal-parie
asymmetry of tone cor
speech prosody (Friede
Wildgruber *et al.*, 2006)
sentence-level prosody i

activation in the inferior frontal gyrus (IFG). When asked to discriminate pitch and timing patterns in linguistic (Thai *pseudowords*) and non-linguistic (hums) conditions (Gandour *et al.*, 2002), Thais, but not Chinese, activated the left IFG for both tone and vowel length (/*bat*$^{\mathit{low}}$/ 'card' vs. /*baat*$^{\mathit{low}}$/ 'monetary unit'). Regardless of whether the phonological contrast in Thai is signaled by spectral or temporal cues, left frontal areas are dominant for native Thais, but not Chinese. Left frontal areas are similarly activated by native Chinese when presented with Mandarin tones in real words, whereas homologous regions in the RH are activated by English (Hsieh *et al.*, 2001). Interestingly, when asked to discriminate Mandarin tones embedded in English words (not linguistically relevant), it is the *right* insula and IFG that are activated regardless of language background (Chinese, English) (Figure 32.1) (Wong, Parsons, Martinez, & Diehl, 2004). Thus, pitch processing engages the LH only when the pitch patterns are phonologically significant to the listener; otherwise, they are lateralized to the RH. We further conclude that LH mechanisms mediate processing of linguistic information irrespective of acoustic cues or type of phonological unit, i.e., segmental or suprasegmental.

To address the issue of a possible lexical–semantic confound directly, hybrid stimuli consisting of Thai tones superimposed onto Mandarin syllables (*tonal chimeras*) and Mandarin tones onto the same syllables (Mandarin words) were presented to Thai and Mandarin natives (Xu *et al.*, 2006). In a comparison of native vs. non-native tones, overlapping activity between language groups was identified in the left planum temporale. 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. We argue that this neural activity is related to *prelexical* phonological processing of tones, supporting the view that relatively early stages of acoustic–phonetic processing can be modulated by stimulus features that are phonologically relevant in particular languages (Griffiths & Warren, 2002; Hickok & Poeppel, 2007).

Differential patterns of cortical activation, however, are not driven by language affiliation alone. They may also be driven by differences in acoustic features associated with specific types of phonological units (segmental, suprasegmental). Production of Mandarin tones elicit more activity in the RH than vowels (Liu *et al.*, 2006). In a phonological recognition task, Mandarin tones, relative to either consonants or rhymes, show increased activation in frontal-parietal areas of the RH (Li *et al.*, 2010). This rightward asymmetry of tone converges with the role of the RH in mediating speech prosody (Friederici & Alter, 2004; Glasser & Rilling, 2008; Wildgruber *et al.*, 2006). See further discussion of bilateral involvement in sentence-level prosody in tone language speakers (Gandour, 2006a, 2007;

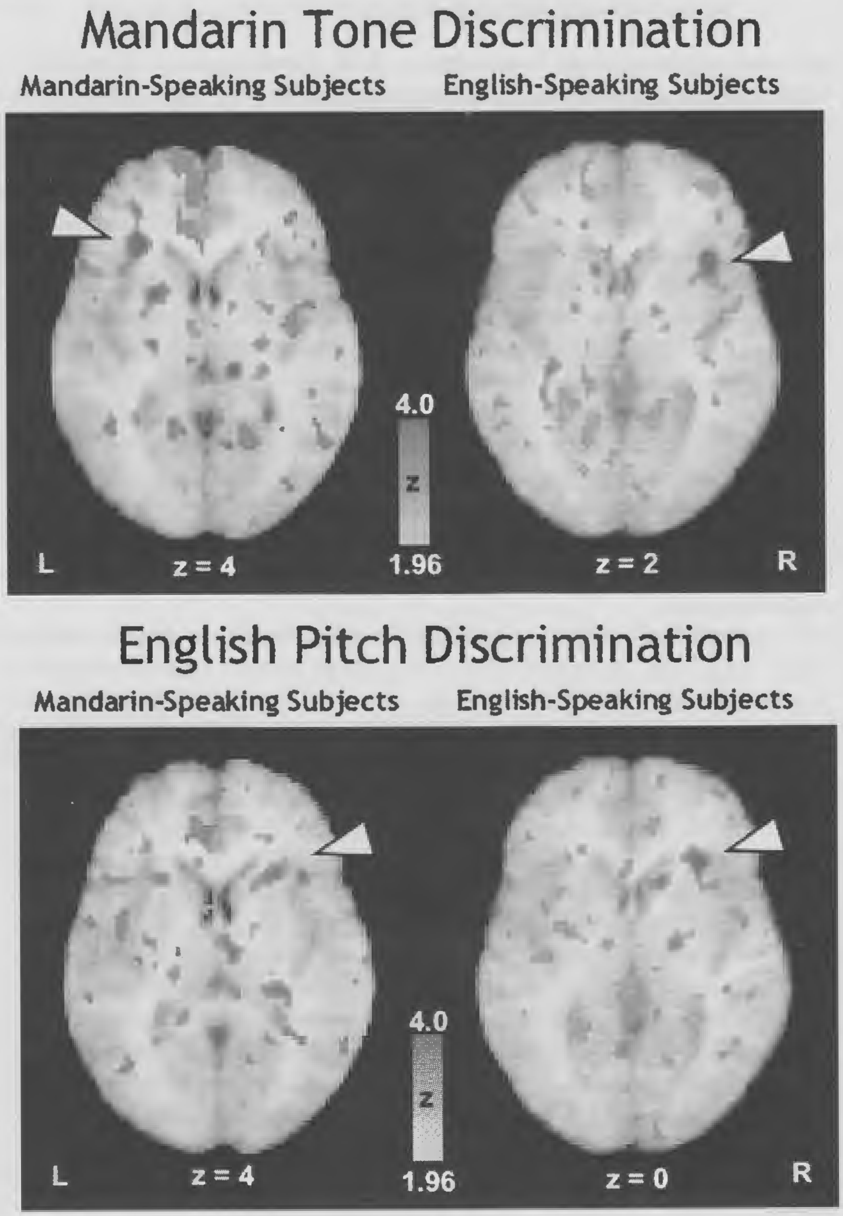
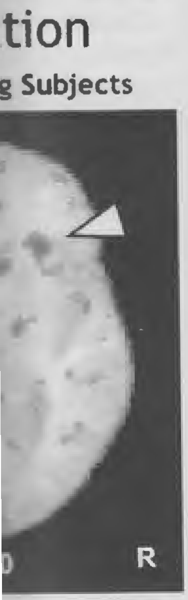


Figure 32.1. Position emission tomography (PET) images show increased activity in the left anterior insular cortex when Chinese natives discriminate pitch patterns embedded in Mandarin words (top panel), but in the homologous area of the RH for those embedded in English words (bottom panel). In contrast, English speakers' activity is circumscribed to the RH regardless of lexical function. (Adapted from *Journal of Neuroscience*, *24*(41), 2004, 9157, with permission from Society of Neuroscience.)

Zatorre & Gandour,
(Wang, Jongman, &
Superior regions
initial stages of
(STS) and middle
logical processing
between core auditory
created an acoustic
from the Mandarin
paradigm (frequent–
variation *within* tonal
lateral STG bilaterally,
(Figure 32.2). In contra
across tonal categories,
were located in the
been implicated in
2004; Meyer *et al.*,
intermediate stage of
bilateral dorsal STG
phonological processing

A

Figure 32.2.
deviant contrasts
Mandarin high
within-category
and across-categ
left MTG). STG
(Adapted from



show increased
tives discriminate
but in the
h words (bottom
ibed to the RH
euroscience,
oscience.)

Zatorre & Gandour, 2008) and in second-language learning of Mandarin tones (Wang, Jongman, & Sereno, 2006).

Superior regions of the temporal lobe are known to be responsible for initial stages of auditory analysis, whereas the superior temporal sulcus (STS) and middle temporal gyrus (MTG) have been implicated in phonological processing (Hickok & Poeppel, 2004, 2007). To assess interaction between core auditory and ventral regions concurrently, Zhang *et al.* (2011) created an acoustic continuum with 11 equal physical intervals ranging from the Mandarin high rising to falling tone. Using a passive oddball paradigm (frequent–standard/rare–deviant), brain areas activated by acoustic variation *within* tonal categories were located in the dorsal and posterior-lateral STG bilaterally, with the strongest activation in the *right* middle STG (Figure 32.2). In contrast, brain areas activated by phonological variation *across* tonal categories, as compared to within-category acoustic variation, were located in the *left* middle MTG. These areas in the temporal lobe have been implicated in the processing of tone and intonation (Gandour *et al.*, 2004; Meyer *et al.*, 2004). The left middle/anterior STS may represent an intermediate stage of processing in a functional pathway linking areas in the bilateral dorsal STG to areas in the left MTG that are engaged in higher-level phonological processing (Liebenthal *et al.*, 2005, 2010).

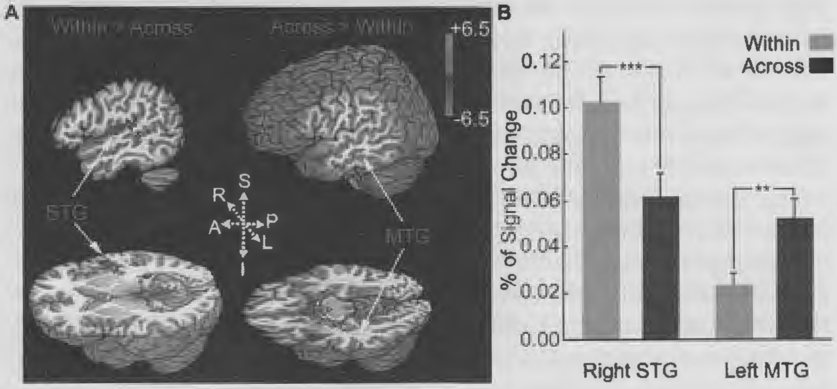


Figure 32.2. Activation in within-category deviant vs. across-category deviant contrasts elicited from a tonal continuum ranging from the Mandarin high rising to falling tone. Regions of activity are shown for within-category > across-category (panel A, STG; panel B, right STG) and across-category > within-category deviants (panel A, MTG; panel B, left MTG). STG, superior temporal gyrus; MTG, middle temporal gyrus. (Adapted from *PLoS One*, ***6***(6), 2011, e20963.)

Electrophysiological

In contrast to hemodynamic methods, electrophysiological recordings yield high temporal (poor spatial) resolution that give us a window on the time-course of events or processes in the brain (Rugg, 1999). The mismatch negativity (MMN) response, elicited using an auditory passive oddball paradigm, is an event-related potential (ERP) with a frontal-central distribution about 150–300 ms after stimulus onset. Language experience, as reflected by the MMN, influences the automatic, preattentive discrimination of speech sounds (Naatanen, 2001).

There are data to support an acoustic basis for hemispheric specialization of pitch at early cortical stages of processing (Poeppel, 2003; Zatorre & Belin, 2001). Lexical tones elicit a stronger MMN response from native speakers of Mandarin in the RH than in the LH, whereas consonants evoke the opposite asymmetry (Luo *et al.*, 2006). Consonants and tones are characterized by rapidly changing temporal and slowly changing spectral variation, respectively. Such hemispheric asymmetries suggest that early cortical processing of linguistic pitch is driven primarily by acoustic features *before* being mapped onto a semantic representation at later stages. MMNs also yield RH dominance across levels of prosodic representation (tone, intonation) in Mandarin regardless of whether segmental information is preserved in the stimuli (speech, non-speech) (Ren, Yang, & Li, 2009). The influence of phonemic status on the perception of Mandarin tones shows that acoustic and phonological information may be processed *concurrently* within the MMN time window (Xi *et al.*, 2010). Across-category deviants elicited larger MMNs relative to within-category deviants in the LH, probably reflecting more cognitive memory-based comparator processing (cf. Maess *et al.*, 2007). Both within- and across-category deviants elicited marginally larger MMNs in the RH, presumably reflecting its role in sensory processing.

There are no veridical ERP studies of tone perception in Thai to date. Using native (Thai) and non-native (Mandarin) words presented to Thai listeners (Sittiprapaporn *et al.*, 2003), the MMN elicited by a native Thai word was greater than that of a non-native Mandarin word, and lateralized to the LH. But we cannot disentangle effects of word recognition from tone perception. In a study of the perception of Thai tones (Kaan *et al.*, 2007), contrary to expectation, no significant differences in amplitude or latency were observed among native Thai, Mandarin, and English speakers. But their findings actually reveal the influence of preceding consonants on pitch (Gandour, 1974). There is no language-group effect because consonantal perturbations on pitch are found in all languages of the world.

Yet it has been demonstrated that the MMN may serve as an index of acoustic features that are differentially weighted by language experience.

Using Mandarin tonal pairs in a speech context (Chandrasekaran, Krishnan, & Gandour, 2007b), one acoustically similar, the other dissimilar, the MMN amplitude of native Chinese was found to be larger than that of English speakers for the dissimilar pair only. We infer that early cortical pitch processing may be shaped by the relative saliency of underlying dimensions or *features* of pitch. A multidimensional scaling analysis of MMN responses confirmed that Chinese are more sensitive to the *pitch contour* feature, relative to *pitch height*, than English (Chandrasekaran, Gandour, & Krishnan, 2007). Thus, MMN may serve as an index of pitch features that are differentially weighted depending on the number of lexical tones and their distribution within a particular tone space (Gandour, 1983). Using non-speech homologues of Mandarin "high level" and "high rising" tones as well as a linear approximation of the latter (Chandrasekaran, Krishnan, & Gandour, 2007a), native Chinese showed larger MMN responses than English to the natural, curvilinear pitch contours only. Taken together, these findings suggest that experience-dependent neural plasticity in early cortical processing of linguistic pitch is sensitive to naturally occurring pitch features but not specific to speech per se.

The notion that pitch features are driving early tonal processing is further supported in Cantonese (Tsang *et al.*, 2011). MMN and P3a (involuntary attention-switching response after MMN in the passive oddball paradigm, and peaking at about 300 ms) were elicited from two tonal pairs, one that differed in pitch height (height-large vs. height-small), and the other in pitch contour (contour-early vs. contour-late). Pitch height exerted a strong effect on the peak amplitude and latency of MMN (see Gandour, 1983), whereas pitch contour did not (Figure 32.3). But the mere presence of a change in pitch contour, early or late, did influence P3a. This may reflect the salience of the turning point in signaling contour tones (Khouw & Ciocca, 2007). Such findings point to language-dependent enhancement of neural mechanisms for differential processing of pitch features depending upon the prosodic needs of a particular language. At later stages of processing, the influence of pitch height and slope features is demonstrated in a categorical perception experiment of rising pitch contours shared in both Cantonese and Mandarin (Zheng *et al.*, 2012). The P300, an index of task demands (Polich, 2007), is a voluntary attention-switching response elicited in an active oddball paradigm with a parietal distribution about 300 ms after stimulus onset. As reflected by P300 amplitude, Cantonese listeners discriminated the tonal stimuli better than Mandarin. Zheng *et al.* hypothesize that Cantonese listeners make finer distinctions in f_0 height and slope because of differences in tonal inventories (Peng, 2006). The Cantonese tonal space (six tones) is *denser* than that of Mandarin (four tones).

While speech-specific operations are likely circumscribed to the cortex (Scott & Johnsrude, 2003), the auditory signal may be subject to language-dependent

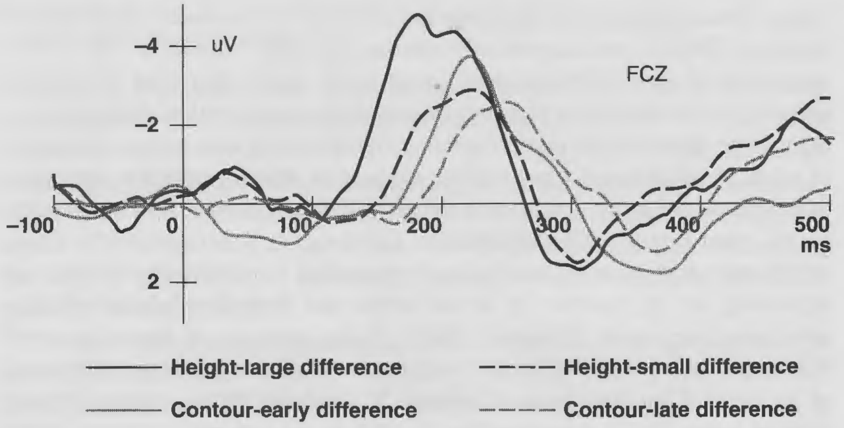


Figure 32.3. Peak amplitude and latency of MMN and P3a show that pitch contour and pitch height are important dimensions used in early processing of Cantonese tones. MMNs were larger in tonal pairs that differ greatly in initial pitch height (height–large, height–small). In contrast, pitch contour influenced the latency of P3a (contour–early, contour–late). FCZ, frontal–central electrode recording site. (Adapted from *Neuroscience Letters*, *487*(3), 2011, 270, with permission from Elsevier Press.)

effects at *subcortical* stages of processing. We argue that the emergence of acoustic–phonetic *features* relevant to tone perception begins no later than 15–18 ms from the time the auditory signal enters the ear (Krishnan & Gandour, 2009). The human frequency-following response (FFR) is used to measure pitch-related electrophysiological activity in the brainstem. This response reflects sustained phase-locked activity in a population of neural elements within the brainstem (Krishnan, 2007), and is characterized by a periodic waveform that follows the individual cycles of the stimulus waveform. As reflected by FFRs, comparisons between native speakers of tone (Mandarin) and non-tone (English) languages show that native experience with lexical tones enhances pitch encoding at the level of the brainstem irrespective of speech or non-speech context (Krishnan *et al.*, 2005, 2009; Swaminathan, Krishnan, & Gandour, 2008). Pitch strength (magnitude of the normalized autocorrelation peak) of 40-ms segments revealed that the Chinese group exhibits more robust pitch representation of those segments containing *rapidly changing* pitch movements across all four tones (Krishnan, Swaminathan, & Gandour, 2009). Long-term experience is believed to sharpen the tuning characteristics of the best modulation frequency neurons along the pitch axis with particular sensitivity to linguistically relevant dynamic segments. Pitch strength of the rising f_0 trajectory (Figure 32.4) is the most important in discriminating listeners by

---

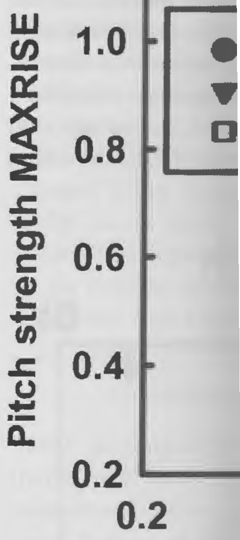


Figure 32.4. D
rising pitch is
language speak
89, with permi

language affiliation (K
pitch features of high
degraded listening co
Bidelman, 2010a). Thi
at increasingly higher a
speech (Krishnan *et a*
processing in the brain
are behaviorally relev

Language-depend
especially sensitive to
natural speech. We fai
how close a linearly ac
native lexical tone (K
A curvilinear pattern i
extraction of the audit
curvilinear pitch patte
(Krishnan *et al.*, 2009)

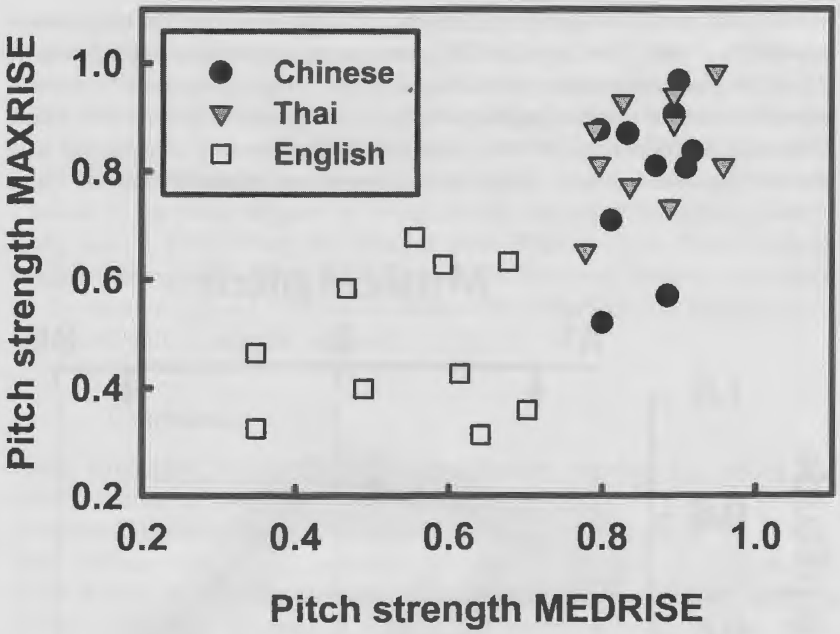


Figure 32.4. Discriminant analysis of pitch strength indicates that moderate rising pitch is important for distinguishing tone language from non-tone language speakers. (Adapted from *Journal of Neurolinguistics*, *23*(1), 2010, 89, with permission from Elsevier Press.)

language affiliation (Krishnan, Gandour, & Bidelman, 2010b). These dynamic pitch features of high perceptual saliency are found to be more resistant to degraded listening conditions in Mandarin listeners (Krishnan, Gandour, & Bidelman, 2010a). This advantage in voice pitch encoding is maintained even at increasingly higher acceleration rates that fall outside the boundary of natural speech (Krishnan *et al.*, 2010), demonstrating that neuroplasticity for pitch processing in the brainstem is not limited to the domain in which pitch contours are behaviorally relevant.

Language-dependent pitch encoding mechanisms in the brainstem are especially sensitive to the curvilinear shape of pitch contours that occur in natural speech. We fail to observe any language-dependent effects no matter how close a linearly accelerating or decelerating pitch pattern approximates a native lexical tone (Krishnan *et al.*, 2009; Xu, Krishnan, & Gandour, 2006). A curvilinear pattern itself, though necessary, is insufficient to enhance pitch extraction of the auditory signal at the level of the brainstem. A *non-native* curvilinear pitch pattern similarly fails to elicit a language-dependent effect (Krishnan *et al.*, 2009).

The question arises whether the neural encoding of pitch in the brainstem is specific to a particular language. In a cross-language comparison (Mandarin, Thai), language-dependent enhancement of pitch representation was demonstrated to transfer to other languages with similar prosodic systems (Krishnan, Gandour, & Bidelman, 2010b). Another related question has to do with transfer between separate domains – language and music. FFRs show that

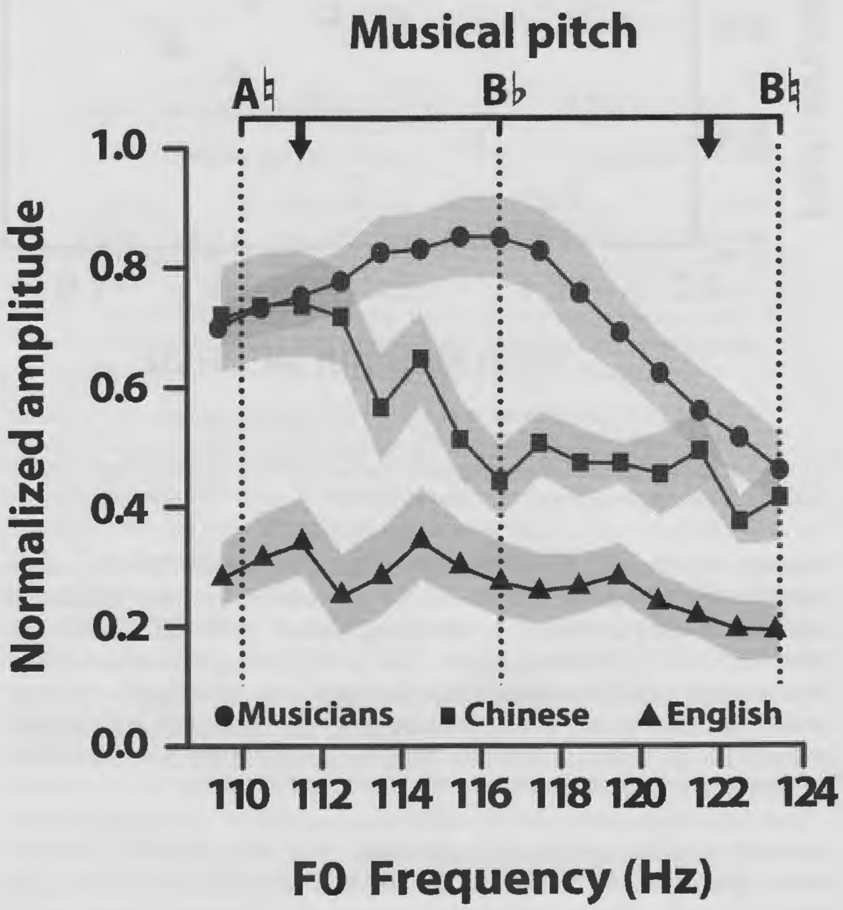


Figure 32.5. Comparisons of spectral f_0 magnitudes reveal that pitch encoding is enhanced in musicians as compared to Chinese or non-musicians in the rapidly changing portion of Mandarin tone 2 (high rising) corresponding to the note B♭ of a discrete musical scale. (Adapted from *Journal of Cognitive Neuroscience*, *23*(2), 2011, 431, with permission from MIT Press.)

---

music training fac
Chandrasekaran, Kr
show that tone lan
FFRs were stronger
to a *musical* pitch i
ingly, English-spea
Chinese in just thos
tually salient notes
suggest that experie
by the relative salie
associated with a pa

Conclusion

Tonal languages p
transformation of t
attention-modulated
brain. Instead of si
to be driven by mu
different *real-time*
A more complete ac
the interplay betwe
addition to those de
effects on lower-le
hierarchically organ
processing, general-
perceptually salient
linguistic status. Fr
ments in brainstem
comparable to neura
iorally relevant soun
(Suga *et al.*, 2003).

Acknowled

Research supported
01A1 (A.K.). Repri
Department of Spee
Lafayette, IN, USA

music training facilitates pitch processing of Mandarin tones (cf. MMN: Chandrasekaran, Krishnan, & Gandour, 2009; Wong *et al.*, 2007). FFRs also show that tone language experience facilitates music processing. Chinese FFRs were stronger than those of English-speaking non-musicians in response to a *musical* pitch interval (Bidelman, Gandour, & Krishnan, 2011). Surprisingly, English-speaking musicians' FFRs were even superior to those of Chinese in just those subparts of a lexical tone that can be related to perceptually salient notes along the musical scale (Figure 32.5). These findings suggest that experience-dependent plasticity of brainstem responses is shaped by the relative saliency of acoustic dimensions underlying the pitch patterns associated with a particular domain.

Conclusions

Tonal languages provide an especially valuable window for tracing the transformation of the pitch signal from early preattentive sensory to later attention-modulated tasks at cognitive stages of processing in the human brain. Instead of simple dichotomies, hemispheric laterality of pitch is seen to be driven by multiple dichotomies or scalar features that may apply at different *real-time* intervals at cortical or subcortical levels of the brain. A more complete account of tonal processing will require that we understand the interplay between general sensory-motor and cognitive processes in addition to those derived from linguistic knowledge. Experience-dependent effects on lower-level sensory processing are compatible with an integrated, hierarchically organized auditory pathway to the brain. In the case of tonal processing, general-purpose auditory processes are tuned differentially to perceptually salient features of the auditory signal depending upon their linguistic status. From the perspective of auditory neuroethology, enhancements in brainstem neural activity relevant to encoding of voice pitch is comparable to neural mechanisms that are developed for processing behaviorally relevant sounds in other non-primate and non-human primate animals (Suga *et al.*, 2003).

Acknowledgments

Research supported by the National Institutes of Health R01 DC008549–01A1 (A.K.). Reprint requests should be addressed to Jackson Gandour, Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, IN, USA 47907 (email: gandour@purdue.edu).