



## Review

# The role of the auditory brainstem in processing linguistically-relevant pitch patterns

Ananthanarayan Krishnan, Jackson T. Gandour \*

Department of Speech Language Hearing Sciences, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038, USA

## ARTICLE INFO

## Article history:

Accepted 15 March 2009

Available online 14 April 2009

## Keywords:

Auditory

Human

Brainstem

Pitch

Language

Frequency following response (FFR)

Iterated rippled noise (IRN)

Mandarin Chinese

Experience-dependent plasticity

Speech perception

## ABSTRACT

Historically, the brainstem has been neglected as a part of the brain involved in language processing. We review recent evidence of language-dependent effects in pitch processing based on comparisons of native vs. nonnative speakers of a tonal language from electrophysiological recordings in the auditory brainstem. We argue that there is enhancing of linguistically-relevant pitch dimensions or features well before the auditory signal reaches the cerebral cortex. We propose that long-term experience with a tone language sharpens the tuning characteristics of neurons along the pitch axis with enhanced sensitivity to linguistically-relevant, rapidly changing sections of pitch contours. Though not specific to a speech context, experience-dependent brainstem mechanisms for pitch representation are clearly sensitive to particular aspects of pitch contours that native speakers of a tone language have been exposed to. Such experience-dependent effects on lower-level sensory processing are compatible with more integrated, hierarchically organized pathways to language and the brain.

© 2009 Elsevier Inc. All rights reserved.

## 1. Introduction

Historically, the brainstem has not been considered to be a part of the brain worthy of interest when it comes to language processing. The conventional wisdom is that “processing operations conducted in the relay nuclei of the brainstem and thalamus is general to all sounds, and speech-specific operations probably do not begin until the signal reaches the cerebral cortex” (Scott & Johnsrude, 2003, p. 100). It is agreed that speech-specific operations are likely circumscribed to the cortex. Nevertheless, we see that the auditory signal is subject to language-dependent effects at subcortical stages of processing. This is not contradictory. By focusing on specific properties of the auditory signal, irrespective of a speech or nonspeech context, we argue that the emergence of acoustic–phonetic *features* relevant to speech perception begins no later than 5–7 ms from the time the auditory signal enters the ear. These general-purpose auditory processes are *tuned* differentially to such features depending upon their linguistic relevance. Such effects of language experience on lower-level sensory processing is compatible with a more integrated approach to language and the brain (Hickok & Poeppel, 2004; Zatorre & Gandour, 2008). This review focuses on how language experience modulates processing of linguistically-relevant pitch in *tone languages* (e.g., Thai) at the level of the brainstem. It is seen that early sensory processing involves more than a simple transmission of

pitch information from the ear to the cerebral cortex. In fact, we see that pitch encoding in the auditory brainstem is an example of experience-dependent neural plasticity.

### 1.1. Linguistic functions of pitch

Tone languages exploit phonologically contrastive pitch at the word or syllable level (Gandour, 1994; Yip, 2003). Such languages are common in the Far East and Southeast Asia. Mandarin Chinese, for example, in addition to consonants and vowels, has four lexical tones (Howie, 1976):  $yi^1$  ‘clothing’ high level [T1];  $yi^2$  ‘aunt’ high rising [T2];  $yi^3$  ‘chair’ low falling-rising [T3];  $yi^4$  ‘easy’ high falling [T4]. Such languages are to be distinguished from those in which pitch variations are usually not contrastive at the syllable or word level (e.g., English). In nontone languages, however, variations in pitch may be used to signal stress and intonation patterns at post-lexical levels of representation. The crucial feature that differentiates between these two types of languages is whether or not pitch variations are contrastive in the lexicon. All languages use pitch variations for intonation, but fewer possibilities are available in tone languages because of co-occurring demands for pitch variation at the lexical level.

### 1.2. Perceptual dimensions of pitch in tone languages

Although there may be concomitant changes in duration, intensity, and phonation, the most important acoustic correlate of tones

\* Corresponding author. Fax: +1 765 494 0771.

E-mail addresses: [rkrish@purdue.edu](mailto:rkrish@purdue.edu) (A. Krishnan), [gandour@purdue.edu](mailto:gandour@purdue.edu) (J.T. Gandour).

is voice fundamental frequency ( $f_0$ ). Its primacy as a cue to tonal identification of citation forms has been confirmed in perception tests using both natural and synthetic speech stimuli. In Mandarin,  $f_0$  contours provide the dominant cue for tone recognition (Howie, 1976; Xu, 1997). Both  $f_0$  height and movement provide sufficient information for high intelligibility of tones in Thai (Abramson, 1962), Mandarin (Howie, 1976), and Cantonese (Vance, 1977). Rapid  $f_0$  movements are required for high intelligibility of contour tones (Abramson, 1978). The overall shape of the  $f_0$  contour is crucially important, as exemplified in high identification of the four Mandarin tones in a frequency range reduced to 4 Hz (Klatt, 1973). The relative importance of  $f_0$  height and movement in cuing tonal distinctions varies depending on the specific pair of tones being contrasted (Lin & Repp, 1989). In Mandarin, the primacy of  $f_0$  cues notwithstanding, concomitant acoustic cues may include the shape of the amplitude envelope, syllable duration, and phonation quality (Fu, Zeng, Shannon, & Soli, 1998; Garding, 1986; Liu & Samuel, 2004; Whalen & Xu, 1992).

Identification and discrimination tasks reveal classical patterns of categorical perception (CP) in Mandarin Chinese listeners for both speech and nonspeech stimuli varying along a linear  $f_0$  continuum from (level [T1] to rising [T2] (Xu, Gandour, & Francis, 2006). CP for pitch direction depends on a listener's experience with a tone language, as shown by the lack of similar CP effects in nontone language listeners. With respect to pitch movement, it appears that a continuum ranging from one or level tone to another is not perceived categorically (Abramson, 1979; Francis, Ciocca, & Ng, 2003), whereas a continuum ranging from a high level tone to a high rising contour tone is perceived categorically (Francis et al., 2003; Wang, 1976).

Indeed, it has been shown that such  $f_0$  features or dimensions, as opposed to categories, are critical to a fuller understanding of tone perception. As a multidimensional perceptual attribute, pitch relies on several acoustic dimensions (e.g., height, slope, direction). Psychophysical evidence for the multidimensional nature of pitch perception in tone languages comes primarily from cross-language multidimensional scaling (MDS) studies of dissimilarity ratings. Based on a sample of tone languages from the Far East (Thai, Cantonese, Mandarin, and Taiwanese) and West Africa (Yoruba) and a nontone language (English), three dimensions are reported to underlie a common perceptual space: average  $f_0$  height, direction of  $f_0$  movement, magnitude of  $f_0$  slope (Gandour, 1978, 1983; Gandour & Harshman, 1978). Their relative importance varies depending on a listener's familiarity with specific types of pitch patterns that occur in their native language. For example, the perceptual saliency of the contour dimension is greater for native speakers of tone languages than for speakers of English, while English listeners give greater weight to the height dimension than do tone language speakers. Such differences in perceptual saliency suggest that *long-term* experience enhances listeners' attention to pitch dimensions that are phonetically relevant in a particular language. Recent MDS analyses of *short-term* training in the recognition of lexical tones (Francis, Ciocca, Ma, & Fenn, 2008; Guion & Pederson, 2007) show that training-related changes in listeners' perceptual space can be attributed to differential weighting of  $f_0$  height and direction dimensions comparable to those observed by Gandour and colleagues. These perceptual data raise questions about the nature of time-varying dimensions of lexical tones that apply to pitch processing at the level of the brainstem.

### 1.3. Physiological mechanisms of pitch

The physiological bases of pitch perception are still a matter of debate as reflected in the extant literature. One view is that the auditory system extracts pitch from complex sounds by deriving a spectral profile from frequency-specific auditory input, followed

by pattern-matching mechanisms (Cohen, Grossberg, & Wyse, 1995). A contrasting view is that the auditory system extracts pitch from the timing of auditory nerve fiber activity irrespective of frequency organization. These temporal models are based solely on the timing information available in the inter-spike intervals represented in the simulated (Krumbholz, Patterson, Seither-Preisler, Lammertmann, & Lutkenhoner, 2003; Meddis & O'Mard, 1997; Patterson, Allerhand, & Giguere, 1995) or actual (Cariani & Delgutte, 1996a, 1996b) auditory nerve activity. They derive a pitch estimate by pooling timing information across auditory nerve fibers without regard to the frequency-to-place mapping. Thus, neural phase-locking related to voice fundamental frequency ( $f_0$ ) plays a dominant role in the encoding of low pitch associated with complex sounds. Temporal encoding schemes provide a unified and parsimonious way of explaining a diverse range of pitch phenomena (Meddis & O'Mard, 1997).

The human frequency following response (FFR), reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem (see Krishnan, 2006, for review of FFR response characteristics and cochlear initiation site). The response is characterized by a periodic waveform which follows the individual cycles of the stimulus waveform. Experimental evidence points to the inferior colliculus (IC) as the source of the FFR generator, as reflected in a latency correspondence between FFR and the potential recorded directly from the IC (Smith, Marsh, & Brown, 1975); selective amplitude reduction of the FFR following cryogenic cooling of the IC (Smith et al., 1975); absence of FFR in individuals with lesions confined to the IC (Sohmer & Pratt, 1977). The shorter latency of the FFR (around 6–9 ms) correlates well with activity from the IC region and is too early to reflect activity from cortical generators (Galbraith, 2008; Galbraith et al., 2000). Furthermore, the nature of the auditory system makes it unlikely that the low-pass filtered phase-locked activity reflected in the FFR is of cortical origin (Alkhoun et al., 2008). However, there is compelling evidence to suggest that this brainstem component is indeed subject to corticofugal modulation (Banai, Abrams, & Kraus, 2007; Suga, Gao, Zhang, Ma, & Olsen, 2000; Suga & Ma, 2003).

### 1.4. Speech and language processing in the brainstem

There has been increasing interest in role of the auditory brainstem in speech processing within the past decade. In terms of speech intelligibility, FFRs show increased amplitude in response to forward speech, as compared to reversed speech, indicating that familiar phonetic and prosodic properties of forward speech selectively activate brainstem neurons (Galbraith et al., 2004). Using the /da/ syllable to elicit the brainstem response, Kraus and colleagues in a series of studies demonstrate how FFRs separately encode source and filter characteristics of the speech signal (for reviews, see Johnson, Nicol, & Kraus, 2005; Kraus & Nicol, 2005), and how brainstem timing predicts cerebral asymmetry for speech (Abrams, Nicol, Zecker, & Kraus, 2006).

In terms of segmental features of speech, FFRs preserve spectral peaks corresponding to the first two formants of both steady-state vowels (Krishnan, 1999, 2002) and time-variant consonants (Krishnan & Parkinson, 2000; Plyler & Ananthanarayan, 2001). Though FFRs are known to preserve pitch-relevant information about complex sounds that produce *time-invariant* pitch (Greenberg, Marsh, Brown, & Smith, 1987), the question arises as to how the brainstem handles suprasegmental features of speech that is characterized by *time-variant* pitch.

By virtue of time-variant pitch, tone languages are especially advantageous for isolating effects of encoding of voice pitch at the level of the auditory brainstem. In Mandarin, for example, four words may comprise a minimal quadruplet, minimally

distinguished by variations in pitch; but otherwise identical in terms of consonant and vowel segments (see Section 1.1). In addition to their linguistic significance, all four tones exhibit voice  $f_0$  trajectories and harmonics that lie within the range of easily recordable FFRs (below 2 Hz). The relatively longer duration of citation forms (200–350 ms) necessitates use of slower stimulus repetition rates which in turn enables recording of robust FFRs with little or no neural adaptation.

### 1.5. Hierarchical processing along the auditory pathway and beyond

We have made considerable progress over the past decade in our understanding of the complex series of processing stages that are required to translate speech sounds into meaning at the level of the cerebral cortex. Functional imaging evidence points to multiple, parallel, hierarchically organized processing pathways that are related to speech processing in the cerebral cortex (Hickok & Poeppel, 2000, 2004, 2007; Obleser, Zimmermann, Van Meter, & Rauschecker, 2007; Poeppel, Idsardi, & van Wassenhove, 2008; Scott, 2003; Scott & Wise, 2004; Zatorre, Belin, & Penhune, 2002; Zatorre & Gandour, 2008). Speech processing in the cortex also emerges from differential demands on distributed brain regions shared by both verbal and nonverbal auditory processing (Price, Thierry, & Griffiths, 2005).

In the case of pitch, functional imaging reveals hierarchical processing in subcortical regions along the auditory pathway. Encoding of temporal regularities of pitch begins as early as the cochlear nucleus but is not completed until the auditory cortex (Griffiths, Uppenkamp, Johnsrude, Josephs, & Patterson, 2001). Of relevance to FFRs, they found that the IC is more sensitive to changes in temporal regularity than the cochlear nucleus. Further evidence of a hierarchy of pitch processing is found in the cerebral cortex (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002). When the pitch is varied to produce a melody, activation moves beyond primary auditory cortex with relatively more activity in the right hemisphere.

Electrophysiological recordings are crucial for investigating questions about the hierarchy of pitch processing not only cortically but subcortically as well (Griffiths, Warren, Scott, Nelken, & King, 2004). In addition, our focus on language and pitch in the brainstem stems from the view that a complete understanding of the processing of linguistically-relevant dimensions 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, p. 69). They argue that early processing stages (e.g., brainstem) may perform transformations on the acoustic data that are relevant to linguistic as well as non-linguistic auditory perception. Scott (2003) similarly argues for hierarchical processing at the cortical level, allowing for the possibility of differences in the *degree* of processing of speech and non-speech stimuli. Indeed, we shall suggest that sensory representations of speech parameters at the brainstem are based on graded scales that are modulated by language experience.

### 1.6. Scope of this review of language-dependent processing in the brainstem

It has also been shown that the malleability of the auditory system can be exploited to study the interaction between sensory and cognitive processes at the level of the brainstem (see Kraus & Banai, 2007, for review). They emphasize that “auditory processing is not a rigid, encapsulated process; rather, it interacts intimately with other neural systems . . . and is affected by experience [music, language], environmental influences . . . , and active training . . .” (p. 105).

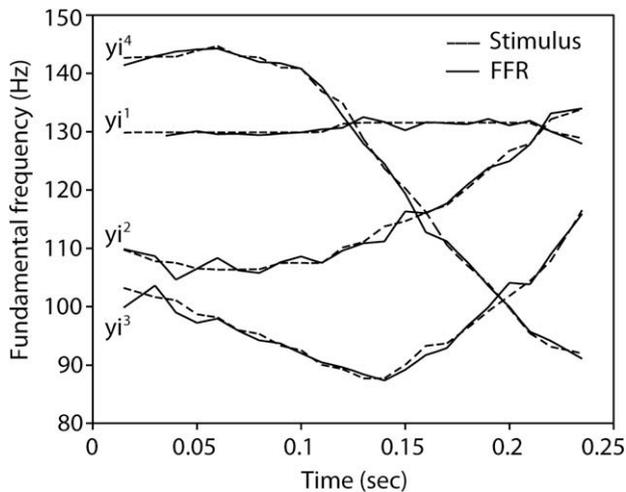
This review will focus on the malleability of *language-dependent* effects on pitch processing in the brainstem. We will de-

scribe major findings from recent studies of experience-dependent neural plasticity in the brainstem by A. Krishnan and his colleagues at Purdue University over the past 5 years. An evaluation of this evidence will lead us to infer that although FFRs preserve pitch information of lexical tones in both native and nonnative listeners, they are more robust in the former because of their long-term exposure to native pitch contours in natural speech. Early sensory encoding of pitch information below the cerebral cortex is sensitive to language experience. Instead of tonal categories, pitch extraction in the brainstem is crucially dependent on specific dimensions or features of pitch contours. The interaction between auditory and cognitive processes is reflected by the well-known differential sensitivity to rising and falling pitch movements in the auditory system which is seen to be enhanced in the brainstem depending upon the prosodic needs of a particular language. We propose that long-term experience with a tone language sharpens the tuning characteristics of neurons along the pitch axis with enhanced sensitivity to linguistically-relevant, rapidly changing sections of pitch contours. Though not specific to a speech context, experience-dependent brainstem mechanisms for pitch representation are clearly sensitive to particular aspects of pitch contours that native speakers of a tone language have been exposed to. All of this evidence will be brought to bear on major theoretical issues concerning the role of the brainstem in language processing. What is the nature of acoustic phonetic features at subcortical levels of processing? What is the role of pitch mechanisms in the brainstem in the encoding of linguistically-relevant dimensions of sounds? Looking at the auditory pathway from the cochlea to auditory cortex and the connections to the brainstem, are language-dependent effects to be ascribed to corticofugal mechanisms or local reorganization in the brainstem itself in the hierarchical processing of the temporal structure of sound?

## 2. Language-dependent plasticity in the encoding of pitch in the human brainstem

### 2.1. Is it feasible to use the scalp-recorded FFR to evaluate the neural representation of pitch in the brainstem?

Auditory nerve single-unit population studies have demonstrated that phase-locking plays a dominant role in the neural encoding of both the spectrum and voice pitch of speech sounds (Cariani & Delgutte, 1996a, 1996b; Meddis & O'Mard, 1997; Patterson et al., 1995). Phase-locked neural activity underlying the scalp-recorded human FFR has also been shown to encode certain spectral features of steady-state and time-variant speech sounds as well as pitch of several complex sounds that produce time-invariant pitch percepts (Aiken & Picton, 2008; Alkhoun et al., 2008; Johnson et al., 2005; Krishnan, 2006). By extension, we hypothesized that the human FFR may preserve pitch-relevant information for speech sounds that elicit time-variant as well as steady-state pitch percepts (Krishnan, Xu, Gandour, & Cariani, 2004). FFRs were recorded differentially between EEG type scalp electrodes placed on the high forehead at midline and at the ipsilateral mastoid in response to the four lexical tones of Mandarin Chinese. Short-term autocorrelation functions and running autocorrelograms (ACGs) were computed from the FFRs to index variations in FFR periodicities over the duration of the response (Krishnan, Swaminathan, & Gandour, *in press*; Krishnan, Xu, Gandour, & Cariani, 2005). ACG represents the running distribution of all-order phase-locked intervals present in the population response and quantifies periodicity and pitch strength variations over time. These measures revealed that the phase-locked neural activity underlying the FFR faithfully follows the pitch changes presented in each stimulus (Fig. 1).



**Fig. 1.** FFR voice pitch contours (solid lines) superimposed on stimulus  $f_0$  contours (broken lines) for the four Mandarin Chinese lexical tones ( $yi^1$  'clothing' high level;  $yi^2$  'aunt' high rising;  $yi^3$  'chair' low falling-rising;  $yi^4$  'easy' high falling). Pitch was extracted using a short-term autocorrelation algorithm (Boersma, 1993) on multiple frames of the signal utilizing a Hanning window of effective length equal to 0.04 s. (Krishnan et al., 2004, with permission of Elsevier.)

These findings for time-variant  $f_0$  suggest that a robust neural temporal representation for pitch is preserved in the phase-locked neural activity of an ensemble of neural elements in the rostral brainstem (Hall, 1979). They extend previous findings of FFR encoding of pitch-relevant information in complex stimuli with time-invariant  $f_0$  (Greenberg et al., 1987). Another interesting finding on voice pitch is that FFR pitch strengths of  $yi^2$  and  $yi^3$  were greater than  $yi^1$  and  $yi^4$ . It has previously been demonstrated that the FFR amplitude for falling tonal sweeps is smaller than that for corresponding rising tonal sweeps (Krishnan & Parkinson, 2000). Similar selectivity to rising tonal sweeps has been observed for cochlear microphonics (Shore & Cullen, 1984), eighth nerve compound action potentials (Shore & Nuttall, 1985), and responses of the ventral cochlear nucleus units (Shore, Clopton, & Au, 1987). It is plausible that differences in FFR pitch strength between Chinese tones with rising (T2, T3) vs. nonrising (T1, T4)  $f_0$  contours, may reflect a differential temporal response pattern for rising and falling tones among the neural elements generating the FFR. Thus, we conclude that the scalp-recorded FFR provides a noninvasive window to view neural processing of voice pitch in human speech sounds at the level of the auditory brainstem.

## 2.2. Does language experience influence the neural representation of pitch in the brainstem?

In the cerebral cortex, the neural substrates of pitch perception in the processing of lexical tones are shaped by language experience (Gandour, 2006a, 2006b, 2007; Zatorre & Gandour, 2008, for reviews). Based on evidence from positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies of pitch processing in Mandarin and Thai, it appears that pitch processing engages the left hemisphere only when the pitch patterns are of linguistic relevance (cf. Wong, 2002; Wong, Parsons, Martinez, & Diehl, 2004). These experiments all employed discrimination tasks, and thus likely reflect temporally aggregated neural events at relatively late attention-modulated stages of auditory processing. The mismatch negativity (MMN), a cortical event-related response associated with auditory discrimination, provides a window on early stages of pitch processing in the cortex. It is elicited using a passive oddball paradigm, eliminating task-related

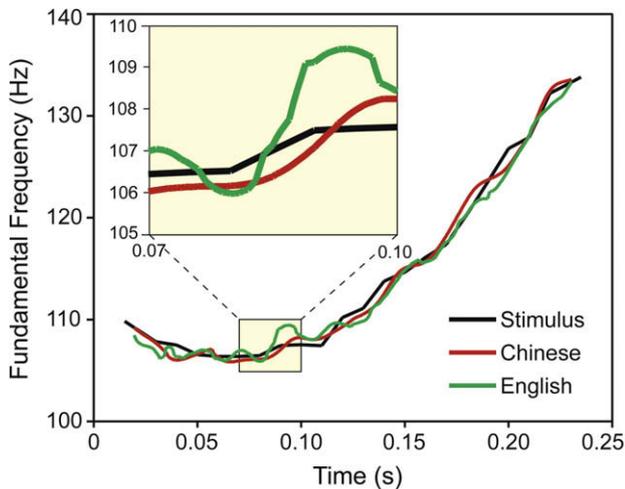
attention or memory confounds, and thereby represents an automatic, preattentive level of processing. As reflected by the MMN, it has been shown that language experience similarly influences the early cortical automatic processing of linguistically-relevant pitch contours (Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Krishnan, & Gandour, 2007; Kaan, Wayland, Bao, & Barkley, 2007), and moreover, that lexical tones are lateralized to the right hemisphere, in contrast to consonants which are lateralized to the left hemisphere (Luo et al., 2006). The Luo et al. data are not irreconcilable with previous PET and fMRI studies. Rather they suggest that hemispheric laterality effects are a result of specialized neural computations that apply to representations at different stages of auditory processing. The leftward asymmetry observed in a discrimination task likely reflects neural computations that occur downstream from preattentive auditory processing.

In animals, it is already well-established that experience-dependent neural plasticity is not limited to the cerebral cortex. Response properties and frequency maps in the IC of bats undergo change after auditory conditioning or focal electrical stimulation of the auditory cortex (see Suga, 1990, 1994, for reviews). Auditory experience of altered interaural cues for localization in young owls lead to frequency-dependent changes in interaural time difference tuning and frequency tuning of IC neurons (Gold & Knudsen, 2000).

Other recent data also support experience-dependent neural plasticity at the level of the IC in humans. The latency of wave V in hearing-impaired listeners who use amplification is shorter than in those who do not (Philibert, Collet, Vesson, & Vuillet, 2005). Using the FFR, we observe that brainstem responses improve after auditory training in children with learning impairments (Russo, Nicol, Zecker, Hayes, & Kraus, 2005); pitch tracking accuracy of Mandarin tones is more accurate in nonnative musicians than nonmusicians (Wong, Skoe, Russo, Dees, & Kraus, 2007); experience with sounds composed of acoustic elements relevant to speech leads to developmental changes in brainstem responses (Johnson, Nicol, Zecker, & Kraus, 2008); and pitch track accuracy improves in native English-speaking adults after undergoing short-term training on using Mandarin tones in word identification (Song, Skoe, Wong, & Kraus, 2008). Also relevant is the consequence of a disruption in the normal interaction between local processes and the corticofugal modulation of subcortical function which contributes to plasticity. The deficits in brainstem encoding in children with a variety of language-based learning problems could very well reflect such a disruption in the ability of the corticofugal system to fine tune subcortical processes (Banai, Nicol, Zecker, & Kraus, 2005; Russo et al., 2008; Song, Banai, & Kraus, 2008).

The question then arises whether early, preattentive stages of pitch processing in the brainstem may also be influenced by language experience. We conducted a cross-language study to determine whether native speakers' long-term exposure and experience using pitch patterns in a tonal language has an influence on FFR response properties (Krishnan et al., 2005). FFRs were elicited by the four Mandarin tones presented to native speakers of Mandarin and nontone language speakers of English. If driven by acoustic properties regardless of language experience, FFRs would be expected to be homogeneous across listeners.

Results showed that pitch strength was significantly greater for the Chinese group than for the English across all four Mandarin tones. Grand-average  $f_0$  contours extracted from FFRs elicited by the Tone 2 stimulus ( $yi^2$ ) indicate that both language groups were able to follow the rising  $f_0$  contour of the original speech stimulus (Fig. 2). However, pitch tracking is more variable for the English group compared to the Chinese (see enlarged inset). Indeed, FFR pitch tracking, as measured by rank-transformed cross-correlation coefficients, was significantly greater in the Chinese group as compared to the English across all four Mandarin tones.



**Fig. 2.** Grand-average  $f_0$  contours of Mandarin Tone 2 derived from the FFR waveforms of all subjects across both ears in the Chinese and English groups. The  $f_0$  contour of the original speech stimulus is displayed in black. The enlarged inset shows that the  $f_0$  contour derived from the FFR waveforms of the Chinese group (red) more closely approximates that of the original stimulus ( $yi^2$  'aunt') when compared to the English group (green). (Krishnan et al., 2005, with permission of Elsevier.)

Current temporal encoding schemes of pitch extraction based on the dominant interval in the distribution of inter-spike intervals rely purely on the acoustic properties of the stimulus (Cariani & Delgutte, 1996a, 1996b). In those schemes, no significant differences in the characteristics of pitch encoding are predicted as a function of language experience. Our findings demonstrate otherwise, i.e., that the encoding scheme is not static nor is it dedicated to faithfully extract only the physical properties of the stimulus. Instead, they point to a temporal encoding scheme which is sensitive to language experience. This language-dependent plasticity enhances or primes temporal intervals that carry linguistically-relevant features of pitch contours. The relatively greater pitch strength and smoother pitch tracking in native Mandarin listeners may reflect the operation of this language-dependent encoding scheme.

We conclude that experience-driven adaptive neural mechanisms are involved subcortically that sharpen response properties of neurons tuned for processing pitch contours that are sensitive to the prosodic needs of a particular language. From the perspective of auditory neuroethology, this adjustment in processing pitch contours of Mandarin tones is comparable to neural mechanisms that are developed for processing behaviorally relevant sounds in other nonprimate and nonhuman primate animals (Suga, Ma, Gao, Sakai, & Chowdhury, 2003). Auditory processing is not limited to a simple representation of acoustic features of speech stimuli. Indeed, language-dependent operations may begin before the signal reaches the cerebral cortex.

### 2.3. What are the limits of sensitivity of brainstem neurons to linguistically-relevant features of the speech signal?

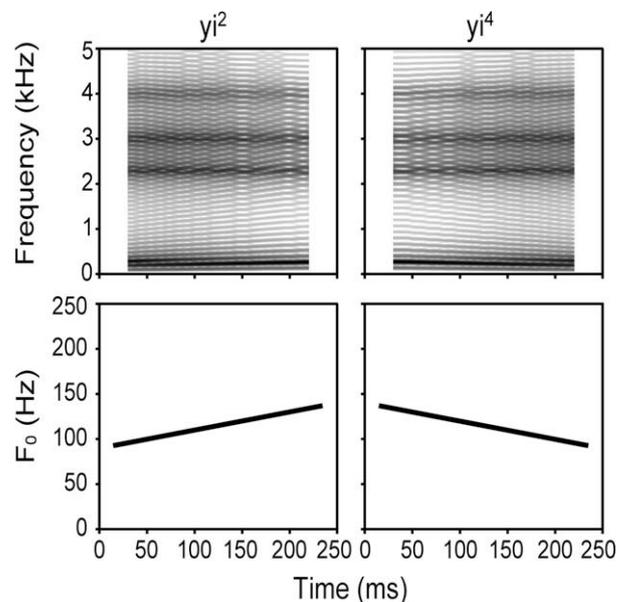
In Krishnan et al. (2005), stimuli exhibited prototypical, curvilinear  $f_0$  contours that were modeled after Mandarin tones that occur in natural speech. If brainstem reorganization is induced by language experience, the question arises as to what the tolerance limits are for linguistic sensitivity at this subcortical level. What specific  $f_0$  properties or features of the pitch stimuli, static or dynamic, are relevant? To what extent can a stimulus deviate from natural speech exemplars before exceeding the upper or the lower limit of linguistic sensitivity of brainstem neurons?

As an initial attempt to address this question (Xu, Krishnan, & Gandour, 2006), we conducted an experiment to determine whether linear  $f_0$  ramps (90–140 Hz, rising; 140–90 Hz, falling), similar to Mandarin T2 and T4 in direction, but dissimilar in trajectory (Fig. 3), elicit brainstem FFRs differentially as a function of language experience (Chinese, English). It is noted that linear ramps do not occur in natural speech because of physiological constraints of the speech production apparatus. In speech perception, however, linearity may be overridden at later cortical stages of processing which recruit attention and memory components. For example, behavioral data from multidimensional scaling (Gandour, 1983) and categorical perception (Xu, Gandour, et al., 2006) of lexical tones reveal language-dependent effects elicited by linear  $f_0$  ramps. But linear ramps are not ecologically representative of what speakers of languages of the world, tonal or otherwise, are exposed to. By examining FFRs elicited by linear approximations of Mandarin T2 and T4, we were able to assess the tolerance limits for priming linguistically-relevant features of the auditory signal useful for pitch extraction at the level of the brainstem.

ANOVAs on FFR pitch strength and pitch tracking accuracy revealed no significant main effects for either language group (Chinese, English) or pitch direction (rising, falling). Neither was there any interaction between language group and pitch direction.

We infer that no language-dependent effects are observed in response to linear rising or falling  $f_0$  ramps because they are not part of native Chinese listeners' experience. Even though the  $f_0$  ramps are dynamic, linear approximations of T2 and T4, they are constant in acceleration and deceleration, respectively. The fact that Mandarin and English FFRs are homogeneous in response to linear trajectories suggests that pitch representations in the brainstem are acutely sensitive to dynamic, curvilinear changes in trajectory throughout the duration of a pitch contour. In the auditory brainstem, neural mechanisms respond to specific dimensions of pitch contours that native speakers have been exposed to. Language-dependent neuroplasticity occurs only when salient dimensions of pitch-relevant to speech perception are present in the auditory signal.

Further support for dimensional specificity comes from a recent study in which FFRs were recorded from Chinese and English par-



**Fig. 3.** Spectrograms and  $f_0$  contours of Mandarin Chinese synthetic speech stimuli ( $yi^2$  'aunt', rising linear ramp;  $yi^4$  'easy', falling linear ramp). (Xu, Krishnan, et al., 2006, adapted with permission of Lippincott Williams & Wilkins.)

ticipants in response to an iterated rippled noise (IRN; see Section 2.4 for detailed description) homolog of a prototypical T2 in contrast to three  $f_0$  variants (two linear, one curvilinear) that do not occur in the Mandarin tonal space (Krishnan, Gandour, Bidelman, & Swaminathan, 2009). Of the two linear variants, one represented a linear ascending ramp (cf. Xu, Krishnan, et al., 2006), the other, a tri-linear approximation of T2, preserving the major points of inflection besides onset and offset. The curvilinear variant was a polynomial flip of T2. No group differences in pitch strength were observed for any of these variants. The absence of language group effects in response to curvilinear, in addition to linear variants of T2, emphasizes that language-dependent neuroplasticity at the level of the brainstem extends only to those pitch patterns that actually occur in the Mandarin tonal space.

#### 2.4. Can we extract FFRs from nonspeech homologs of ecologically representative Mandarin tones?

Early stages of processing on the input side may perform computations on the acoustic data that are relevant to linguistic dimensions even when embedded in a nonspeech context. Indeed, perceptual studies of tone perception have demonstrated that the effects of linguistic experience may extend to nonspeech processing under certain stimulus and task (discrimination, identification) conditions (Bent, Bradlow, & Wright, 2006; Luo, Boemio, Gordon, & Poeppel, 2007). In the brainstem, FFRs permit us to index pitch processing in a nonspeech context minus any volitional memory or attention demands. Auditory brainstem responses (ABRs) have been shown to reflect largely separate neural processes in response to nonspeech (broad spectrum click of short duration) and speech (temporally complex syllable of longer duration) stimuli (Song, Bannai, Russo, & Kraus, 2006). Whereas a click vs. /da/ are not equivalent in acoustic complexity, the task before us is to create nonspeech stimuli that are equivalent to speech in terms of pitch attributes.

Our previous FFR findings based on speech stimuli lead us to the question of whether language-specificity in the encoding of pitch at the level of the brainstem applies to linguistically-relevant pitch changes that are devoid of concurrent segmental information, as when pitch patterns associated with lexical tones are presented in a nonspeech context, thus blocking access to the mental lexicon. To answer this question, we must eliminate any potential lexical bias for native listeners. That is, we need to be able to generate auditory stimuli in a nonspeech context that preserve the perception of pitch, but do not have waveform periodicity or highly modulated stimulus envelopes that are characteristic of speech stimuli. If we are to simulate human speech perception, it is imperative that we be able to generate homologous pitch contours that are ecologically representative of what occurs in natural speech. From the perspective of auditory neuroethology (Suga, 1994; Suga et al., 2003), this would enable us to investigate neural mechanisms underlying the processing of pitch contours that are of linguistic relevance comparable to those underlying the processing of behaviorally relevant sounds in other nonprimate and nonhuman primate animals.

Such auditory stimuli exist in the form of iterated rippled noise (IRN), which preserves the temporal regularity of the stimulus without having to repeat the waveform in a periodic manner. An IRN stimulus is generated using a broadband noise which is delayed and added to it repeatedly, and therefore does not have a prominent modulated envelope (Patterson, Handel, Yost, & Datta, 1996; Yost, 1996). The perceived pitch corresponds to the reciprocal of the delay, and the pitch salience increases with the number of iterations of the delay-and-add process. In its simplest form, steady-state stimuli are constructed by adding broadband noise iteratively to itself with a constant delay (Yost, Patterson, & Sheft,

1996). The delay-and-add process introduces temporal regularity into the fine structure of the noise without appreciable changes in the spectrum except for “ripples” at frequency intervals which are the reciprocal of the delay (Patterson et al., 1996). Increases in temporal regularity of steady-state IRN stimuli lead to better temporally-locked neural activity in auditory structures from the cochlear nucleus to cortex (Bendor & Wang, 2005; Griffiths, Buchel, Frackowiak, & Patterson, 1998; Griffiths et al., 2001; Patterson et al., 2002; Shofner, 1999). The IRN algorithm was recently generalized to allow multiple time dependent delays over a range of iteration steps (Denham, 2005), making it possible to detect pitch in *dynamic* IRN by humans. We applied this dynamic IRN algorithm to generate time-variant, curvilinear pitch contours representative of those that occur in natural speech, specifically, polynomial modeling of Mandarin Tone 2 (Swaminathan et al., 2008a, 2008b).

Temporal and spectral properties of the time-variant curvilinear rising tone stimulus (left panels) and electrophysiological response at the brainstem (right panels) at a high iteration step are shown in Fig. 4. Waveform periodicity (1st row) and resolution of curvilinear spectral bands (2nd row) are markedly improved at the high iteration step, as compared to a low iteration step (cf. Swaminathan et al., 2008a, 2008b), in both stimulus and FFR response. The autocorrelograms show clearer bands (black) of temporal regularity in the stimulus and phase-locked activity in the FFR response at the high iteration step (3rd row).

Such dynamic IRN stimuli enable us to evaluate the sensitivity of FFR responses to speech-like pitch contours in a parametrically controllable way without lexical-semantic confounds. Our view is that although cross-language differences in FFR responses may emerge from language experience, the effects of such experience are *not* necessarily specific to speech perception. FFR responses to IRN stimuli modeled after homologous pitch features in the speech context are expected to show that neural mechanisms at the brainstem are targeting particular features of speech rather than speech per se.

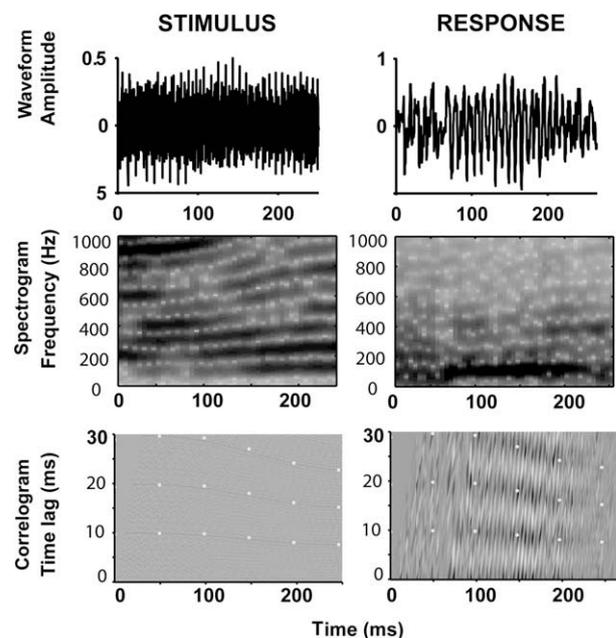


Fig. 4. Waveform, spectrogram, and autocorrelogram, in order from top to bottom, of the curvilinear IRN homolog of Mandarin Tone 2 (left panels) and electrophysiological response at the brainstem (right panels) for a high iteration step. (Swaminathan et al., 2008a, 2008b, adapted with permission of IEEE.)

2.5. Do FFRs to IRN homologs of Mandarin tones vary as a function of language experience?

We hypothesized that pitch representation in the brainstem in response to IRN homologs of the four Mandarin tones, as reflected by FFRs, would be more robust in native speakers of Mandarin Chinese as compared to monolingual English speakers who had no prior knowledge of Mandarin or any other tone language (Krishnan, Swaminathan, et al., in press). Pitch strength (whole contour, 250 ms; 40 ms segments) and pitch tracking accuracy (whole contour) were extracted from the FFRs using autocorrelation algorithms. Narrow band spectrograms were used to extract spectral information.

Fig. 5 (left panels) shows that overall pitch tracking, as measured by the time lag associated with the autocorrelation maximum per language group, is less variable for the Chinese (dashed line) compared to the English group (dotted line). That is, on the

whole,  $f_0$  contours derived from the FFR waveforms of the Chinese group more closely approximate those of the original IRN stimuli. Cross-language comparisons showed that cross-correlation coefficients were significantly larger in the Chinese than the English group for T2, T3, and T4, but not T1. FFR spectral variations are strongly represented up to the 5th harmonic in the Chinese group, but only up to the 3rd or 4th harmonic in the English group. The better representation of multiple pitch-relevant harmonics in the Chinese group points to more accurate pitch tracking as well as stronger pitch representations.

FFR pitch strength, as measured by the average magnitude of the normalized autocorrelation peak per language group, is shown for six tonal sections within each of the four IRN homologs of Mandarin tones (Fig. 5, right panels). Pitch strength values for the Chinese and English groups, respectively, appear above and below the  $f_0$  contour. In T2 and T3, pitch strength of the Chinese group was greater than the English group in three out of six tonal sections

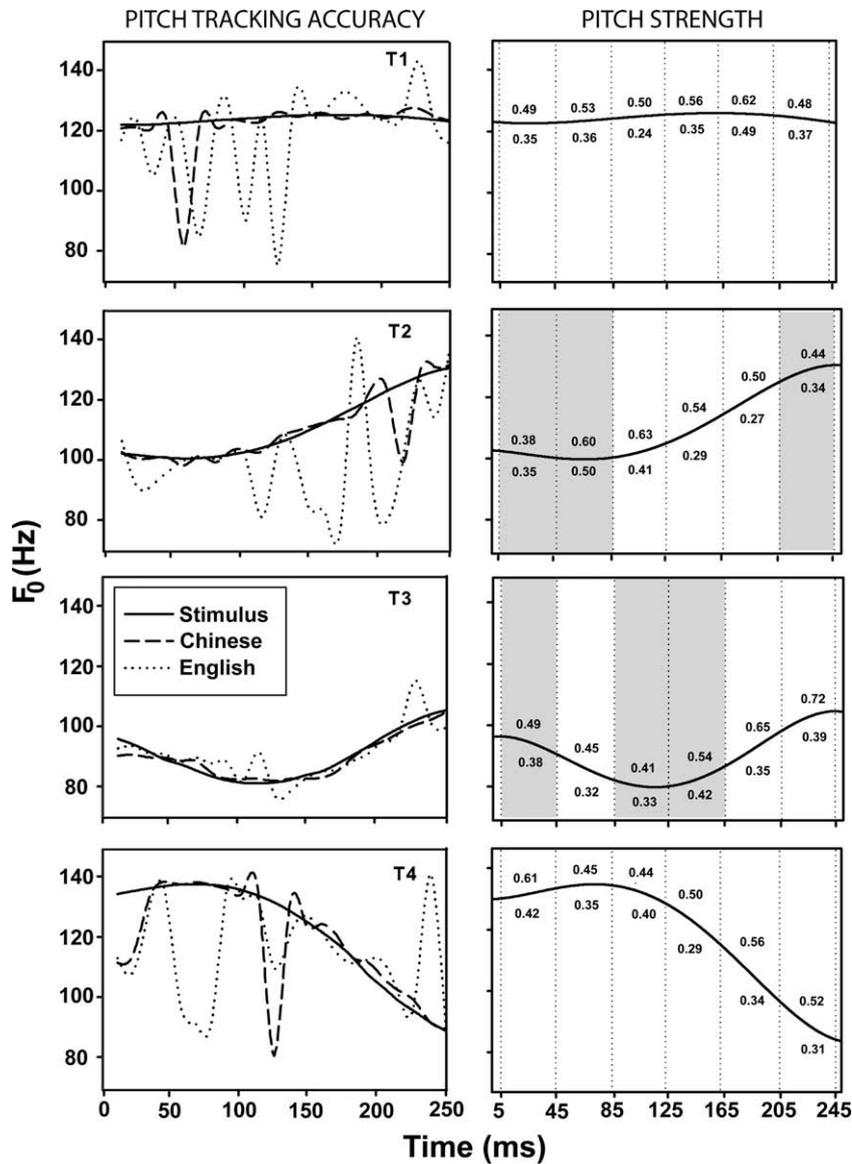


Fig. 5. Pitch tracking accuracy of IRN homologs of Mandarin tones (left) and pitch strength of tonal sections (right) derived from the grand averaged FFR waveforms of Chinese and English subjects. The four Mandarin tonal categories are represented as T1, T2, T3, and T4. Left panels show that the FFR-derived  $f_0$  contours of the Chinese group (dashed line) more closely approximate those of the original IRN stimuli (solid line) when compared to the English group (dotted line). Right panels show that the pitch strength of the Chinese group (value above the solid line) is greater than that of the English group (value below the solid line). Vertical dotted lines demarcate six 40-ms sections within each  $f_0$  contour: 5–45, 45–85, 85–125, 125–165, 165–205, and 205–245. Sections that yielded significantly larger pitch strength for the Chinese group relative to English are unshaded; those that did not are shaded in gray. (Krishnan, Swaminathan, et al., in press, with permission of MIT Press.)

(Fig. 5, right panels, unshaded). In T1 and T4, pitch strength of the Chinese group was greater than the English group in all tonal sections.

A discriminant analysis was used to determine the extent to which individual subjects can be classified into their respective language groups based on a weighted linear combination of their pitch strength of three 40-ms temporal intervals that were maximally differentiated in terms of slope (flat, rising, falling). About 83% of subjects were correctly classified into their respective language groups. The group centroid of the Chinese group was larger than that of the English. Univariate tests of pitch strength confirmed that more dynamic changes in pitch (rising, falling) had a greater influence on the FFR responses of the Chinese group relative to the English, whereas less dynamic changes in pitch (flat) did not yield a language group effect.

Pitch strength of the rising  $f_0$  trajectory was the most important in discriminating listeners by language affiliation. Both psychoacoustic (Collins & Cullen, 1978; Klatt, 1973; Nabelek, 1978; Schouten, 1985) and physiologic studies (Krishnan & Parkinson, 2000; Shore & Nuttall, 1985; Shore et al., 1987) indicate better sensitivity for rising vs. falling tones. Multidimensional scaling analyses show that the perceptual dimension related to direction of pitch change is spatially distributed primarily in terms of rising vs. nonrising  $f_0$  movements (Gandour, 1983; Gandour & Harshman, 1978). In a previous FFR study (Krishnan et al., 2004), pitch strength of T2 and T3, averaged across the stimulus, is greater than that of T1 and T4. Both T2 and T3 contain steeper rising pitch sections. This response asymmetry in FFRs presumably reflects greater neural synchrony (Shore & Nuttall, 1985) and more coherent temporal response patterns to rising than to falling tones (Shore et al., 1987). Perhaps the greater weighting towards rising pitch for the Chinese group in this study reflects an experience-dependent enhancement of both neural synchrony and temporal response pattern coherence among the neural elements generating the FFR.

Which time-varying features or dimensions of the lexical tones are arguably more relevant to pitch processing at the level of the brainstem? As reflected by the positive correlation between pitch strength and acceleration *across tones*, we propose that the degree of acceleration or deceleration is a critical variable that influences pitch extraction in the rostral brainstem. As reflected by the number of sections of maximal acceleration/deceleration *within each tone* itself, language experience is observed to have an influence on pitch strength primarily in those tonal sections exhibiting higher degrees of acceleration or deceleration. These findings are consistent with speech production data showing that  $f_0$  patterns in Mandarin have a greater amount of dynamic movement as a function of time and number of syllables than those in English (Eady, 1982). This experience-dependent effect, however, occurs in the brainstem only when stimuli reflect curvilinear dynamic contours representative of Mandarin tones as opposed to *linear* dynamic approximations (cf. Xu, Krishnan, et al., 2006).

Overall we have demonstrated that experience-dependent neural mechanisms for pitch representation at the brainstem level are sensitive to specific dimensions of pitch contours that native speakers of a tone language are frequently exposed to in natural speech. They suggest that brainstem neurons are differentially sensitive to changes in pitch without regard to the context in which they are presented. We infer that the role of the brainstem is to facilitate cortical level processing of pitch-relevant information by optimally capturing those dimensions of the auditory signal that are of linguistic relevance.

By focusing on tonal sections instead of the whole tone, we are able to determine whether language-dependent effects are better conceptualized as applying throughout the tonal contour, or alternatively, as applying to sections that exhibit certain acoustic dimensions irrespective of tonal category. Such experimental out-

comes support the view that at early stages of brain processing, neural mechanisms underlying speech perception are shaped by particular dimensions of pitch patterns regardless of the stimulus context in which they are embedded.

### 2.6. Do speech stimuli induce more robust neural phase-locking, as reflected in the FFR, than nonspeech stimuli (IRN) with degraded periodicity?

At an early preattentive 'subcortical' stage of processing, FFRs elicited in response to Mandarin tones reveal smoother pitch tracking in native vs. nonnative listeners, no matter the context, speech or nonspeech (Krishnan, Swaminathan, et al., in press; Krishnan et al., 2005). By measuring 'pitch strength', peak of autocorrelation function, we are able to focus on individual sections of pitch contours. In our next cross-language study (Swaminathan et al., 2008a, 2008b), FFRs were recorded from Chinese and English participants in response to four Mandarin tonal contours presented in a speech and nonspeech context. This made it possible to carry out a direct comparison between speech and nonspeech (IRN) stimuli to determine whether the brainstem mechanisms responsible for extracting pitch information are susceptible to stimulus degradation. Another objective was to determine whether experience-dependent neural mechanisms for pitch representation in the brainstem are sensitive to specific time-varying features of pitch contours that native speakers of a tone language are familiar with *regardless of context*.

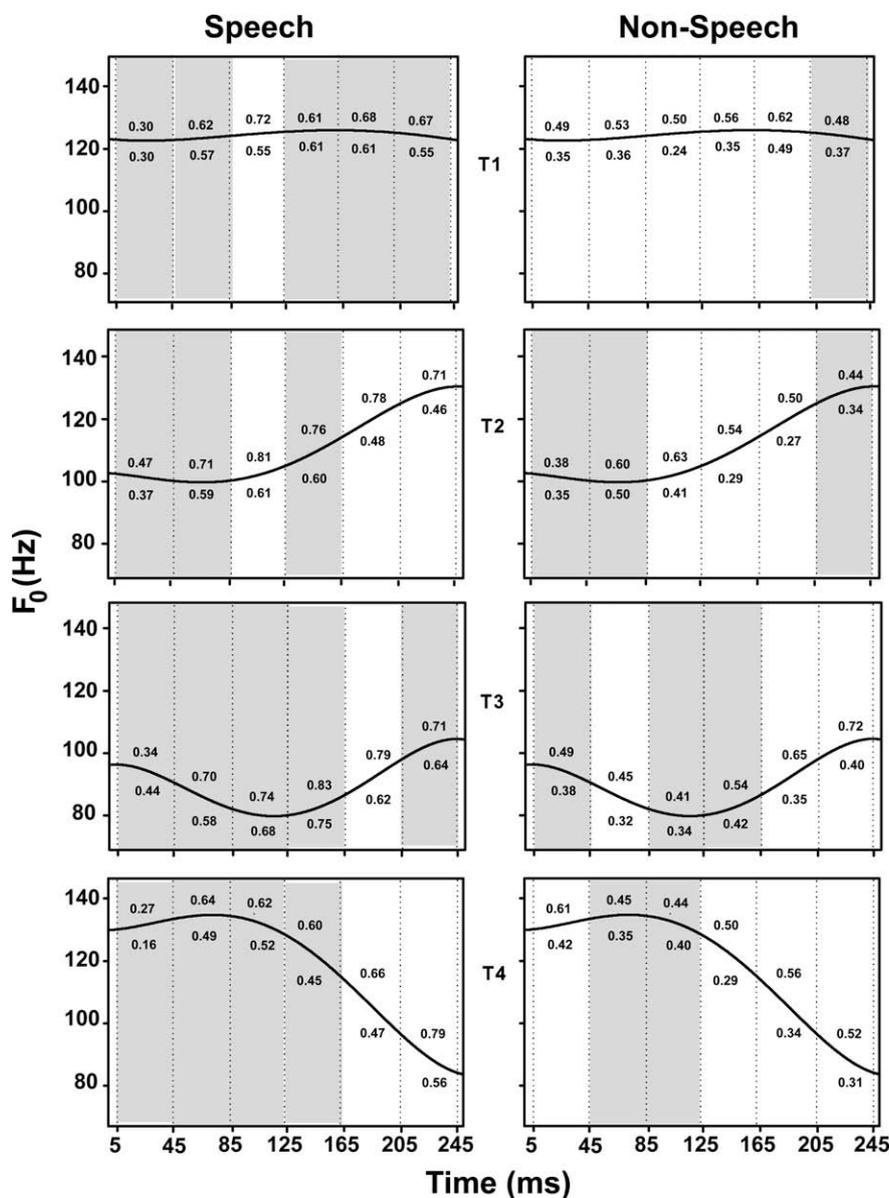
As expected, given the relatively less robust temporal periodicity in the stimulus waveform, pitch strength was observed to be greater for speech than nonspeech stimuli for both English and Chinese listeners. Notwithstanding, dynamic IRN stimuli do preserve fine-grained measures of pitch representation at the level of the brainstem.

Regardless of context, pitch strength of the Chinese group was greater than that of the English (Fig. 6). However, group differences in pitch strength were not uniform throughout the duration of FFR responses to either speech or their IRN homologs. In some tonal sections that have rapid changes (e.g., T4, S4; speech), the two language groups did not differ in pitch strength. In others that are relatively smooth (e.g., T1, S3; speech), they did. Nonetheless, we infer that neural mechanisms in the brainstem are not responding to lexical tones per se, but rather to specific time-varying acoustic properties of the input stimuli. Moreover, the degree of acceleration (e.g., T3, S5) and deceleration (e.g., T4, S5) of the pitch trajectories seems to be a critical factor. We hypothesize that cross-language differences in the sustained phase-locked activity of the brainstem reflect an enhancement of selectivity to pitch-relevant periodicities that correspond with rapidly changing dynamic segments of the pitch contour.

## 3. Theoretical framework for brainstem, pitch, and language

### 3.1. Neural mechanisms underlying experience-dependent pitch enhancement in the brainstem

Long-term experience with language and music (Krishnan et al., 2005; Musacchia, Sams, Skoe, & Kraus, 2007; Wong et al., 2007) and short-term auditory training (Hayes, Warrier, Nicol, Zecker, & Kraus, 2003; Russo et al., 2005; Song, Banai, et al., 2008; Song, Skoe, et al., 2008) can influence neural encoding in the brainstem. We hereby propose an empirically-driven theoretical framework to account for our data showing experience-dependent pitch representation in the brainstem. This framework has three components: (i) a functional connectivity model that considers the IC, MGB (medial geniculate body) and the AC (auditory cortex) as an ipsilateral dominant functional unit with dominant inputs from the contralat-



**Fig. 6.** Pitch strength of tonal sections derived from the FFR waveforms in response to speech (left) and nonspeech (right) stimuli for the two language groups. The four Mandarin tonal categories are represented as T1, T2, T3, and T4. Consistent across speech and nonspeech stimuli, pitch strength of the Chinese group (value above the solid line) is greater than that of the English group (value below the solid line). Sections that yield significantly larger pitch strength for the Chinese group relative to English (unshaded) are those sections which generally exhibit larger acceleration or deceleration values. See also caption of Fig. 5. (Swaminathan et al., 2008a, 2008b, with permission of Lippincott Williams & Wilkins.)

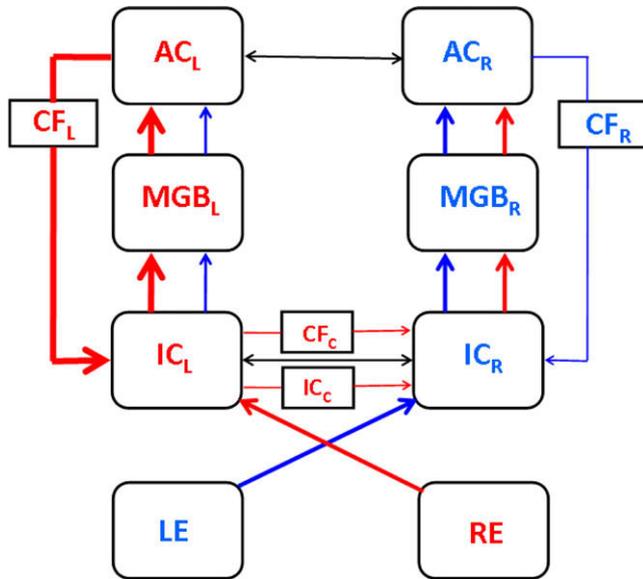
eral ear; (ii) a dominant ipsilateral corticocollicular component that presumably facilitates experience-dependent reorganization of pitch mechanisms in the brainstem during critical periods of speech and language development to optimally encode linguistically-relevant pitch; and (iii) a temporally based pitch mechanism in the IC. This framework can also be expanded to incorporate independent local mechanisms subsequent to reorganization by corticofugal influence.

### 3.2. The role of corticofugal pathways in experience-dependent pitch encoding

The human auditory pathway (Fig. 7) is contralateral dominant such that the lateralized response in the left auditory cortex ( $AC_L$ ) directly reflects that in the  $MGB_L$  and  $IC_L$  and is dominated by stimulation of the contralateral ear (Langers, van Dijk, & Backes, 2005;

Schonwiesner, Krumbholz, Rubsamen, Fink, & von Cramon, 2007). As far as we know, there are as of yet no data on experience-dependent effects on ear dominance at preattentive *subcortical* stages of tonal processing. If electrophysiological activity at the brainstem is driven by acoustics only, the null hypothesis is that effects should be homogeneous across listeners regardless of language experience. An alternative hypothesis is that ear asymmetry in the IC is also influenced by language experience, and that their effects will be greater than those predicted by structural asymmetry alone.

Let us illustrate a case of functional asymmetry in IC responses elicited by RE stimulation, shown by thicker red lines compared to the LE (Fig. 7). For linguistically-relevant sounds, this functional asymmetry is initially reinforced by corticofugal input ( $CF_L$ ) to promote reorganization in the IC for language acquisition. LE stimulation may still yield stronger responses for native relative to nonnative listeners because of the weaker contralateral ( $CF_C$  and



**Fig. 7.** Anatomic model of the auditory pathway illustrating asymmetry in IC responses favoring RE stimulation, shown by the relatively thicker red lines compared to LE stimulation (blue lines). For linguistically-relevant sounds, this functional asymmetry is initially reinforced by the  $CF_L$  input to promote reorganization in the IC for perceptual learning. LE, left ear; RE, right ear; IC, inferior colliculus (brainstem); subscripts L and R, left and right, respectively; MGB, medial geniculate body (thalamus); AC, (primary) auditory cortex; CF, corticofugal; subscript C, contralateral. At the level of the IC and AC, the solid black double arrow indicates that pitch information may be transmitted contralaterally in either direction.

$IC_C$ ) inputs to the  $IC_R$ . Linguistically-relevant sounds may also trigger the relatively smaller  $CF_R$  influence. In contrast, nonnative listeners may exhibit a functional asymmetry favoring LE stimulation, shown by the relatively thicker blue lines compared to the RE. This latter asymmetry effect is less than that observed for native listeners because there is no experience-dependent reorganization in the  $IC_R$  affected by the  $CF_R$  to enhance pitch representation. Thus, we see how corticofugal mechanisms can facilitate experience-dependent reorganization for pitch processing in the brainstem.

Other theoretical schemes have been proposed that invoke corticofugal influence to modulate subcortical sensory processing. The reverse hierarchy theory (RHT) provides a representational hierarchy to describe the interaction between sensory input and top-down processes to guide plasticity in primary sensory areas (Ahissar & Hochstein, 2004; Nahum, Nelken, & Ahissar, 2008; Shamma, 2008). Its basic claim is that neural circuitry mediating a certain percept can be modified starting at the highest representational level and progressing to lower-levels in search of more refined high resolution information to optimize perception. RHT has been invoked as a plausible explanation for top-down influences on subcortical sensory processing (Banai et al., 2007).

Another proposed circuitry for mediating learning-induced plasticity is the cortico-colliculo-thalamo-cortico-collicular (CTCC) loop (Xiong, Zhang, & Yan, in press). This framework contains bottom-up (colliculo-thalamic, thalamo-cortical) and top-down (corticofugal) projections, forming a tonotopic CTCC loop that is assumed to be the only neural substrate carrying accurate auditory information. The CTCC model is similar to our functional connectivity model, both in structure and proposed function. Additionally, the CTCC loop incorporates several neuromodulatory inputs forming a core neural circuit mediating sound-specific plasticity associated with perceptual learning. Auditory stimuli induce activation of the loop; and neuromodulatory inputs are

claimed to induce large scale frequency-specific plasticity in the loop.

Top-down guided plasticity may also be mediated by memory-based theories. It has been proposed that sensory specific memory guides plasticity, instead of higher-level memory functions mediated in prefrontal and parietal regions, to enhance encoding at the sensory level and increase the probability of creating accurate sensory memory traces (Pasternak & Greenlee, 2005). Long-term stored episodes of native pitch patterns (cf. pitch templates) may also guide plasticity by engaging the corticofugal pathways (Golding, 1998). Thus, the corticofugal system is triggered during learning to enhance sensory encoding of specific dimensions that are behaviorally relevant.

While the corticofugal system may be crucially involved in the experience-driven reorganization of subcortical neural mechanisms, it is not necessarily dedicated to maintenance of long-term, permanent on-line subcortical processing. We already know that the corticofugal system can lead to subcortical egocentric selection of behaviorally relevant stimulus parameters in nonprimate and nonhuman primate animals (Suga & Ma, 2003; Suga et al., 2000, 2003). In the case of humans, the corticofugal system likely shapes the reorganization of the brainstem pitch encoding mechanisms for enhanced pitch extraction in earlier stages of language development when plasticity presumably would be most vigorous (Keuroghlian & Knudsen, 2007; Kral & Eggermont, 2007). Once this reorganization is complete over the critical developmental period, local mechanisms in the brainstem are sufficient to extract linguistically-relevant pitch information in a robust manner without further corticofugal interaction. Importantly, however, we would continue to expect corticofugal influence to play an important role under a variety of challenging auditory circumstances (e.g., noise degradation, learning novel stimuli, retraining of hearing-impaired, and adaptation to unfamiliar 'accents').

Several lines of evidence can be adduced in favor of local mechanisms in the brainstem as mediating the experience-dependent plasticity observed in our cross-language FFR data. First, the corticofugal egocentric selection is short-term, and takes time (latency) to be activated, whereas the FFR response latency is only about 6–8 ms. Even considering the sustained portion of our FFR response (260 ms time window for our stimuli), the slower time constants of adaptive plasticity (Dean, Robinson, Harper, & McAlpine, 2008) and the medial olivary cochlear bundle (MOCB) reflex (Backus & Guinan, 2006) are much too sluggish to effectively influence the dynamic pitch pattern over its entire duration. Moreover, the MOCB reflex would require us to invoke bottom-up processes, instead of top-down, that can influence the neural plasticity seen at the IC level. Secondly, plasticity in the IC of mice persists after deactivation of the corticofugal system (Yan, Zhang, & Ehret, 2005). In fact it has been demonstrated in animal models that once tuning is well-established with corticofugal modulation, local mechanisms can maintain the plasticity without permanent corticofugal influence (Gao & Suga, 1998; Ji, Gao, & Suga, 2001). Thirdly, and a more compelling argument in favor of a local mechanism is the absence of a cross-language FFR effect at the brainstem level elicited by linear, or even nonnative curvilinear,  $f_0$  contours (Krishnan, Gandour, et al., 2009; Krishnan, Swaminathan, et al., in press; Krishnan et al., 2005; Xu, Krishnan, et al., 2006). Indeed, multidimensional scaling data show that dimensions underlying the perception of linear  $f_0$  contours, similar to those in Xu, Krishnan et al. (2006), are weighted differentially as a function of language experience (Gandour, 1983; Gandour & Harshman, 1978). Cortical modulation of the brainstem response would have led us to expect, contrary to fact, differential FFR responses to the linear  $f_0$  contours between Chinese and English listeners. These observations together are more consistent with the view that the operation of this language-dependent encoding scheme, presumably induced by na-

tive experience with tones, is local to the generator(s) of the FFR in the auditory brainstem. From the perspective of auditory neuroethology (Suga, 1994; Suga et al., 2003), FFRs allow us to investigate subcortical neural mechanisms underlying the processing of linguistically-relevant pitch contours that are comparable to those neural mechanisms underlying the processing of behaviorally relevant sounds in other nonprimate and nonhuman primate animals.

### 3.3. Local brainstem mechanisms underlying experience-dependent pitch encoding

We infer that the cross-language differences in pitch representation reflect an enhanced tuning to inter-spike intervals that correspond to rapidly changing dynamic segments of the pitch contour. To explain the brainstem mechanism underlying FFR pitch extraction and how language experience may alter this mechanism, we adopt the temporal correlation analysis model (Fig. 8) described by Langner (1992 and 2004). Coincidence detection neurons in the IC perform a correlation analysis on the delayed and undelayed temporal information from the cochlear nucleus to extract pitch-relevant periodicities that are spatially mapped onto a periodicity pitch axis. This encoding scheme is accomplished by neurons with different best modulation frequencies arranged in an orderly fashion orthogonal to the tonotopic frequency map. Its sensitivity can be enhanced by long-term experience, as reflected by the smoother tracking of whole pitch contours and the greater pitch strength of 40-ms sections thereof. Cross-language comparisons further reveal that this encoding scheme is more sensitive to dynamic segments of pitch contours in the native Chinese group relative to the nonnative English group. It is possible that long-term experience sharpens the tuning characteristics of the best modulation frequency neurons along the pitch axis with particular sensitivity to linguistically-relevant dynamic segments. This sharpening is likely mediated by local excitatory and inhibitory interactions that are known to play an important role in signal

selection at the level of the brainstem (Ananthanarayan & Gerken, 1983, 1987). Such interaction may take the form of an active facilitation/disinhibition of the pitch intervals corresponding to the dynamic segments and inhibition of other pitch periods. Neuromodulatory inputs to the CTCC loop could influence the balance between excitation and inhibition (Xiong et al., in press).

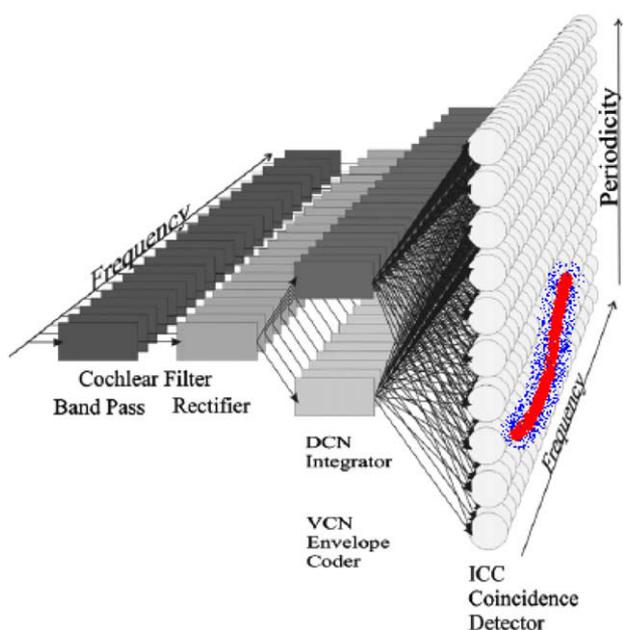
We also observe that the temporal distribution of the phase-locked activity to individual harmonics differs as a function of language experience. Our results not only show a clear dominance of the second harmonic for all stimuli but also better representation of multiple harmonics in the Chinese group relative to English. This finding complements our data on voice pitch representation (Krishnan et al., 2004). In the Chinese group, stronger pitch and more accurate pitch tracking co-occur with relatively stronger representation of pitch-relevant harmonics. Just the opposite is the case for the English group. Moreover, psychoacoustic and physiological data indicate that complex stimuli produce stronger and more accurate pitch percepts when spectral components are prominent in the dominance region for pitch (2nd to about the 5th harmonic) (Cariani & Delgutte, 1996a; Schwartz & Purves, 2004). Consistent with our theoretical framework, we conclude that the role of the brainstem is to facilitate cortical level processing of pitch-relevant information by optimally capturing those dimensions of the auditory signal that are of linguistic relevance.

### 3.4. Emergence of acoustic phonetic feature representations at subcortical stages of processing

The overriding question in speech perception is to figure out how the incoming auditory signal is transformed into an abstract lexical representation (see Poeppel et al., 2008, for review). At the lexical level, we can safely assume that speech sounds are represented as a bundle of distinctive, binary features. Though less is known about intermediate representations, our FFR data lead us to hypothesize that they are based on graded or *n*-ary features. Yet, as we have seen, these intermediate representations are not unrelated to distinctive features. In speech processing, distinctive features are grounded in lower-level sensory dimensions or features that emerge along the auditory pathway. By this parsimonious account, feature-based representations percolate up the processing hierarchy in an analysis-by-synthesis model to the 'phonological primal sketch' whereby the conversion from continuous to discrete representations takes place (Poeppel et al., 2008, Figs. 1 and 4).

There is considerable evidence for the view that basic perceptual attributes (e.g., frequency, duration, intensity) are represented separately in auditory sensory memory (see Caclin et al., 2006, for review, pp. 1960–1961). Of special interest to us are perceptual attributes that rely on several acoustic dimensions. A recent MMN study of timbre, a multidimensional perceptual attribute, shows that its dimensions are represented separately in auditory sensory memory (Caclin et al., 2006). In addition to acoustic changes, it is also noted that the MMN and N100 m may reflect abstract featural representations of phonological segments (Eulitz & Lahiri, 2004; Obleser, Lahiri, & Eulitz, 2004).

Notwithstanding, pitch is a multidimensional perceptual attribute that relies on several acoustic dimensions (e.g., height, slope, direction). In a cross-language MMN study of Mandarin tones (T1, T2, T3), multidimensional scaling analysis revealed that two pitch dimensions, interpretively labeled as 'height' and 'contour', underlie early cortical pitch processing (Chandrasekaran, Gandour, et al., 2007; Chandrasekaran, Krishnan, et al., 2007). Using subject weights on the contour dimension, a discriminant analysis showed that Chinese and English individuals were classified successfully into their respective language group. These data suggest that the MMN can serve as an index of pitch features that are differentially weighted depending on a listener's experience with lexical tones.



**Fig. 8.** Temporal correlation analysis model for extracting pitch periodicities from the cochlea to IC (Langner, 2004) including information reflected in FFRs. The relatively greater pitch strength and smoother pitch tracking in native Mandarin (red) compared to English (blue) listeners suggests that the sensitivity of this pitch encoding scheme can be enhanced by long-term experience. DSN, dorsal cochlear nucleus; VCN, ventral cochlear nucleus; and ICC, inferior colliculus. (Langner, 2004, adapted with permission of Plenum Press.)

The finding that language experience influences preattentive processing of pitch dimensions at early stages of cortical processing is compatible with our cross-language data on pitch processing in the brainstem. FFR data suggest that there may be enhancing or priming of linguistically-relevant pitch dimensions or features well before the auditory signal reaches the cerebral cortex. In the case of Mandarin tones, for example, we see that native listeners' long-term learning experience leads to more robust representations of especially *rapid, nonlinear* changes in pitch movement. No language-dependent effects, on the other hand, are seen in response to *linear*  $f_0$  ramps because native speakers have not been exposed to them in natural speech. It remains to be determined just exactly how fine-grained is linguistic specificity on pitch representations at subcortical stages of processing.

## Acknowledgment

Research supported by the National Institutes of Health R01 DC008549-01A1 (A.K.).

## References

- Abrams, D. A., Nicol, T., Zecker, S. G., & Kraus, N. (2006). Auditory brainstem timing predicts cerebral asymmetry for speech. *Journal of Neuroscience*, 26(43), 11131–11137.
- Abramson, A. S. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments*. Bloomington: Indiana U. Research Center in Anthropology, Folklore, and Linguistics, Pub. 20.
- Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and Speech*, 21, 319–325.
- Abramson, A. S. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 127–134). London: Academic Press.
- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10), 457–464.
- Aiken, S., & Picton, T. (2008). Envelope and spectral frequency following responses to vowel sounds. *Hearing Research*, 245(1–2), 35–47.
- Alkhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., et al. (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology*, 119, 922–933.
- Anathanarayan, A. K., & Gerken, G. M. (1983). Post-stimulation effects on the auditory brain stem response partial-masking and enhancement. *Electroencephalography and Clinical Neurophysiology*, 55(2), 223–226.
- Anathanarayan, A. K., & Gerken, G. M. (1987). Response enhancement and reduction of the auditory brain-stem response in a forward-masking paradigm. *Electroencephalography and Clinical Neurophysiology*, 66(4), 427–439.
- Backus, B. C., & Guinan, J. J. Jr., (2006). Time-course of the human medial olivocochlear reflex. *The Journal of the Acoustical Society of America*, 119(5 Pt 1), 2889–2904.
- Banai, K., Abrams, D., & Kraus, N. (2007). Sensory-based learning disability: Insights from brainstem processing of speech sounds. *International Journal of Audiology*, 46(9), 524–532.
- Banai, K., Nicol, T., Zecker, S. G., & Kraus, N. (2005). Brainstem timing: Implications for cortical processing and literacy. *Journal of Neuroscience*, 25(43), 9850–9857.
- Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054), 1161–1165.
- Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 32(1), 97–103.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, 17, 97–110.
- Caclin, A., Brattico, E., Tervaniemi, M., Naatanen, R., Morlet, D., Giard, M.-H., et al. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, 18(12), 1959–1972.
- Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 76(3), 1698–1716.
- Cariani, P. A., & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology*, 76(3), 1717–1734.
- Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, 25, 195–210.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, 1128(1), 148–156.
- Cohen, M. A., Grossberg, S., & Wyse, L. L. (1995). A spectral network model of pitch perception. *Journal of the Acoustical Society of America*, 98(2), 862–879.
- Collins, M. J., & Cullen, J. K. Jr., (1978). Temporal integration of tone glides. *Journal of the Acoustical Society of America*, 63(2), 469–473.
- Dean, I., Robinson, B. L., Harper, N. S., & McAlpine, D. (2008). Rapid neural adaptation to sound level statistics. *Journal of Neuroscience*, 28(25), 6430–6438.
- Denham, S. (2005). Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model. *Biosystems*, 79(1–3), 199–206.
- Eady, S. J. (1982). Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech*, 25(1), 29–42.
- Eulitz, C., & Lahiri, A. (2004). Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience*, 16(4), 577–583.
- Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36, 268–294.
- Francis, A. L., Ciocca, V., & Ng, B. K. (2003). On the (non)categorical perception of lexical tones. *Perception and Psychophysics*, 65(7), 1029–1044.
- Fu, Q. J., Zeng, F. G., Shannon, R. V., & Soli, S. D. (1998). Importance of tonal envelope cues in Chinese speech recognition. *Journal of the Acoustical Society of America*, 104(1), 505–510.
- Galbraith, G. (2008). Deficient brainstem encoding in autism. *Clinical Neurophysiology*, 119(8), 1697–1700.
- Galbraith, G., Amaya, E. M., de Rivera, J. M., Donan, N. M., Duong, M. T., Hsu, J. N., et al. (2004). Brain stem evoked response to forward and reversed speech in humans. *Neuroreport*, 15(13), 2057–2060.
- Galbraith, G., Threadgill, M., Hemsley, J., Salour, K., Songdej, N., Ton, J., et al. (2000). Putative measure of peripheral and brainstem frequency-following in humans. *Neuroscience Letters*, 292, 123–127.
- Gandour, J. T. (1978). The perception of tone. In V. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 41–76). New York: Academic Press.
- Gandour, J. T. (1983). Tone perception in far Eastern languages. *Journal of Phonetics*, 11, 149–175.
- Gandour, J. T. (1994). Phonetics of tone. In R. Asher & J. Simpson (Eds.), *The encyclopedia of language and linguistics* (Vol. 6, pp. 3116–3123). New York: Pergamon Press.
- Gandour, J. T. (2006a). Brain mapping of Chinese speech prosody. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *Handbook of East Asian psycholinguistics* (Vol. 1, pp. 308–319). Cambridge, UK: Cambridge University Press. Chinese.
- Gandour, J. T. (2006b). Tone: Neurophonetics (2nd ed.). In K. Brown (Ed.), *Encyclopedia of language and linguistics* (Vol. 12, pp. 751–760). Oxford, UK: Elsevier.
- Gandour, J. T. (2007). Neural circuitry underlying the perception of linguistic prosody. In C. Gussenhoven & T. Raid (Eds.), *Tones and tunes: Experimental studies in word and sentence prosody* (Vol. 2, pp. 3–25). Berlin: Mouton de Gruyter.
- Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, 21(1), 1–33.
- Gao, E., & Suga, N. (1998). Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21), 12663–12670.
- Garding, E. (1986). Tone 4 and tone 3 discrimination in modern standard Chinese. *Language and Speech*, 29, 281–293.
- Gold, J. I., & Knudsen, E. I. (2000). A site of auditory experience-dependent plasticity in the neural representation of auditory space in the barn owl's inferior colliculus. *Journal of Neuroscience*, 20(9), 3469–3486.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.
- Greenberg, S., Marsh, J. T., Brown, W. S., & Smith, J. C. (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hearing Research*, 25(2–3), 91–114.
- Griffiths, T. D., Buchel, C., Frackowiak, R. S., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, 1(5), 422–427.
- Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., & Patterson, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nature Neuroscience*, 4(6), 633–637.
- Griffiths, T. D., Warren, J. D., Scott, S. K., Nelken, I., & King, A. J. (2004). Cortical processing of complex sound: A way forward? *Trends in Neurosciences*, 27(4), 181–185.
- Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O. S. Bohn & M. Munro (Eds.), *Second language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Fllege* (pp. 57–77). Amsterdam: John Benjamins.
- Hall, J. W. 3rd, (1979). Auditory brainstem frequency following responses to waveform envelope periodicity. *Science*, 205(4412), 1297–1299.
- Hayes, E. A., Warrior, C. M., Nicol, T. G., Zecker, S. G., & Kraus, N. (2003). Neural plasticity following auditory training in children with learning problems. *Clinical Neurophysiology*, 114(4), 673–684.
- Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4(4), 131–138.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, 92(1–2), 67–99.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.

- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.
- Ji, W., Gao, E., & Suga, N. (2001). Effects of acetylcholine and atropine on plasticity of central auditory neurons caused by conditioning in bats. *Journal of Neurophysiology*, 86(1), 211–225.
- Johnson, K. L., Nicol, T. G., & Kraus, N. (2005). Brain stem response to speech: a biological marker of auditory processing. *Ear and Hearing*, 26(5), 424–434.
- Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental plasticity in the human auditory brainstem. *Journal of Neuroscience*, 28(15), 4000–4007.
- Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research*, 1148, 113–122.
- Keuroghlian, A. S., & Knudsen, E. I. (2007). Adaptive auditory plasticity in developing and adult animals. *Progress in Neurobiology*, 82(3), 109–121.
- Klatt, D. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. *Journal of the Acoustical Society of America*, 53(1), 8–16.
- Kral, A., & Eggermont, J. J. (2007). What's to lose and what's to learn: Development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Research Reviews*, 56(1), 259–269.
- Kraus, N., & Banai, K. (2007). Auditory-processing malleability: Focus on language and music. *Current Directions in Psychological Science*, 16(2), 105–110.
- Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends in Neurosciences*, 28(4), 176–181.
- Krishnan, A. (1999). Human frequency-following responses to two-tone approximations of steady-state vowels. *Audiology and Neuro-Otology*, 4(2), 95–103.
- Krishnan, A. (2002). Human frequency-following responses: Representation of steady-state synthetic vowels. *Hearing Research*, 166(1–2), 192–201.
- Krishnan, A. (2006). Human frequency following response. In R. F. Burkard, M. Don, & J. J. Eggermont (Eds.), *Auditory evoked potentials: Basic principles and clinical application* (pp. 313–335). Baltimore: Lippincott Williams & Wilkins.
- Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience dependent neural representation of dynamic pitch in the brainstem. *Neuroreport*, 20(4), 408–413.
- Krishnan, A., & Parkinson, J. (2000). Human frequency-following response: Representation of tonal sweeps. *Audiology and Neuro-Otology*, 5(6), 312–321.
- Krishnan, A., Swaminathan, J., & Gandour, J. T. (in press). Experience dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*.
- Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: Representation of pitch contours in Chinese tones. *Hearing Research*, 189(1–2), 1–12.
- Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Research Cognitive Brain Research*, 25(1), 161–168.
- Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13(7), 765–772.
- Langers, D. R., van Dijk, P., & Backes, W. H. (2005). Lateralization, connectivity and plasticity in the human central auditory system. *Neuroimage*, 28(2), 490–499.
- Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, 60(2), 115–142.
- Langner, G. (2004). Topographic representation of periodicity information: The 2nd neural axis of the auditory system. In J. Syka & M. Merzenich (Eds.), *Plasticity of the central auditory system and processing of complex acoustic signals* (pp. 21–26). New York: Plenum Press.
- Lin, H. B., & Repp, B. H. (1989). Cues to the perception of Taiwanese tones. *Language and Speech*, 32, 25–44.
- Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, 47, 109–138.
- Luo, H., Boemio, A., Gordon, M., & Poeppel, D. (2007). The perception of FM sweeps by Chinese and English listeners. *Hearing Research*, 224(1–2), 75–83.
- Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., et al. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences of the United States of America*, 103(51), 19558–19563.
- Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, 102(3), 1811–1820.
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, 104(40), 15894–15898.
- Nabelek, I. V. (1978). Temporal summation of constant and gliding tones at masked auditory threshold. *Journal of the Acoustical Society of America*, 64(3), 751–763.
- Nahum, M., Nelken, I., & Ahissar, M. (2008). Low-level information and high-level perception: The case of speech in noise. *PLoS Biology*, 6(5), e126.
- Obleser, J., Lahiri, A., & Eulitz, C. (2004). Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience*, 16(1), 31–39.
- Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related fMRI. *Cerebral Cortex*, 17(10), 2251–2257.
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6(2), 97–107.
- Patterson, R. D., Allerhand, M. H., & Giguere, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, 98(4), 1890–1894.
- Patterson, R. D., Handel, S., Yost, W. A., & Datta, A. J. (1996). The relative strength of the tone and noise components in iterated ripple noise. *Journal of the Acoustical Society of America*, 100(5), 3286–3294.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4), 767–776.
- Philibert, B., Collet, L., Vesson, J. F., & Veuillet, E. (2005). The auditory acclimatization effect in sensorineural hearing-impaired listeners: Evidence for functional plasticity. *Hearing Research*, 205(1–2), 131–142.
- Plyler, P. N., & Ananthanarayan, A. K. (2001). Human frequency-following responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology*, 12(10), 523–533.
- Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363(1493), 1071–1086.
- Price, C., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences*, 9(6), 271–276.
- Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, 156(1), 95–103.
- Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., et al. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, 119(8), 1720–1731.
- Schonwiesner, M., Krumbholz, K., Rubsam, R., Fink, G. R., & von Cramon, D. Y. (2007). Hemispheric asymmetry for auditory processing in the human auditory brain stem, thalamus, and cortex. *Cerebral Cortex*, 17(2), 492–499.
- Schouten, M. E. (1985). Identification and discrimination of sweep tones. *Perception and Psychophysics*, 37(4), 369–376.
- Schwartz, D. A., & Purves, D. (2004). Pitch is determined by naturally occurring periodic sounds. *Hearing Research*, 194(1–2), 31–46.
- Scott, S. K. (2003). How might we conceptualize speech perception? The view from neurobiology. *Journal of Phonetics*, 31, 417–422.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26(2), 100–107.
- Scott, S. K., & Wise, R. J. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition*, 92(1–2), 13–45.
- Shamma, S. (2008). On the emergence and awareness of auditory objects. *PLoS Biology*, 6(6), e155.
- Shofner, W. P. (1999). Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms. *Journal of Neurophysiology*, 81(6), 2662–2674.
- Shore, S. E., Clopton, B. M., & Au, Y. N. (1987). Unit responses in ventral cochlear nucleus reflect cochlear coding of rapid frequency sweeps. *Journal of the Acoustical Society of America*, 82(2), 471–478.
- Shore, S. E., & Cullen, J. K. Jr., (1984). Cochlear microphonic responses of the peripheral auditory system to frequency-varying signals. *American Journal of Otolaryngology*, 5(1), 34–42.
- Shore, S. E., & Nuttall, A. L. (1985). High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts. *Journal of the Acoustical Society of America*, 78(4), 1286–1295.
- Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology*, 39(5), 465–472.
- Sohmer, H., & Pratt, H. (1977). Identification and separation of acoustic frequency following responses (FFRs) in man. *Electroencephalography and Clinical Neurophysiology*, 42(4), 493–500.
- Song, J. H., Banai, K., & Kraus, N. (2008). Brainstem timing deficits in children with learning impairment may result from corticofugal origins. *Audiology and Neuro-Otology*, 13(5), 335–344.
- Song, J. H., Banai, K., Russo, N. M., & Kraus, N. (2006). On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiology and Neuro-Otology*, 11(4), 233–241.
- Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, 20(10), 1892–1902.
- Suga, N. (1990). Biosonar and neural computation in bats. *Scientific American*, 262(6), 60–68.
- Suga, N. (1994). Processing of auditory information carried by complex species-specific sounds. In M. S. Gazzaniga & E. Bizzi (Eds.), *The cognitive neurosciences* (pp. 295–318). Cambridge, MA: MIT Press.
- Suga, N., Gao, E., Zhang, Y., Ma, X., & Olsen, J. F. (2000). The corticofugal system for hearing: Recent progress. *Proceedings of the National Academy of Sciences of the United States of America*, 97(22), 11807–11814.
- Suga, N., & Ma, X. (2003). Multiparametric corticofugal modulation and plasticity in the auditory system. *Nature Reviews Neuroscience*, 4(10), 783–794.
- Suga, N., Ma, X., Gao, E., Sakai, M., & Chowdhury, S. A. (2003). Descending system and plasticity for auditory signal processing: Neuroethological data for speech scientists. *Speech Communication*, 41(1), 189–200.
- Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008a). Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE Transactions on Biomedical Engineering*, 55(1), 281–287.

- Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008b). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport*, *19*(11), 1163–1167.
- Vance, T. (1977). Tonal distinctions in Cantonese. *Phonetica*, *34*, 93–107.
- Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, *208*, 61–72.
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, *49*(1), 25–47.
- Wong, P. C. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*, *59*(2), 83–95.
- Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience*, *24*(41), 9153–9160.
- Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422.
- Xiong, Y., Zhang, Y., & Yan, J. (in press). The neurobiology of sound-specific auditory plasticity: A core neural circuit. *Neurosciences and Biobehavioral Reviews*.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.
- Xu, Y., Gandour, J. T., & Francis, A. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, *120*(2), 1063–1074.
- Xu, Y., Krishnan, A., & Gandour, J. T. (2006). Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport*, *17*(15), 1601–1605.
- Yan, J., Zhang, Y., & Ehret, G. (2005). Corticofugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice. *Journal of Neurophysiology*, *93*(1), 71–83.
- Yip, M. (2003). *Tone*. New York: Cambridge University Press.
- Yost, W. A. (1996). Pitch of iterated rippled noise. *Journal of the Acoustical Society of America*, *100*(1), 511–518.
- Yost, W. A., Patterson, R., & Sheft, S. (1996). A time domain description for the pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America*, *99*(2), 1066–1078.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46.
- Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 1087–1104.