

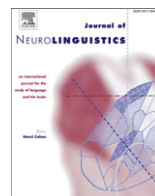


ELSEVIER

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: [www.elsevier.com/locate/jneuroling](http://www.elsevier.com/locate/jneuroling)



## The effects of tone language experience on pitch processing in the brainstem

Ananthanarayan Krishnan, Jackson T. Gandour\*, Gavin M. Bidelman

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

### ARTICLE INFO

#### Article history:

Received 4 June 2009

Received in revised form 4 September 2009

Accepted 5 September 2009

#### Keywords:

Frequency-following response (FFR)

Pitch

Experience-dependent plasticity

Language

Mandarin

Thai

### ABSTRACT

Neural encoding of pitch in the auditory brainstem is shaped by long-term experience with language. The aim herein was to determine to what extent this experience-dependent effect is specific to a particular language. Analysis of variance of brainstem responses to Mandarin and Thai tones revealed that regardless of language identity, pitch-tracking accuracy of whole tones was higher in the two tone language groups (Chinese, Thai) compared to the non-tone language group (English), and that pitch strength of 40-ms tonal sections was generally more robust in tone relative to non-tone languages. Discriminant analysis of tonal sections, as defined by variation in direction and degree of slope, showed that moderate rising pitch was the most important variable for classifying English, Chinese, and Thai participants into their respective groups. We conclude that language-dependent enhancement of pitch representation transfers to other languages with similar phonological systems. From a neurobiological perspective, these findings suggest that neural mechanisms local to the brainstem are tuned for processing pitch dimensions that are perceptually salient depending upon the melodic patterns of a language.

© 2009 Elsevier Ltd. All rights reserved.

### 1. Introduction

Electrophysiological recordings are not only crucial for investigating questions about the hierarchy of pitch processing in the cerebral cortex, but also in subcortical structures (Griffiths, Warren, Scott, Nelken, & King, 2004). Our interest in pitch processing at the level of the brainstem arises from the

\* Corresponding author. Tel.: +1 765 494 3821; 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), [gbidelma@purdue.edu](mailto:gbidelma@purdue.edu) (G.M. Bidelman).

view that a complete understanding of the processing of linguistically relevant dimensions of pitch can only be achieved within a framework involving a series of computations that apply to representations at different stages of processing (Hickok & Poeppel, 2004). At early subcortical stages of processing, the auditory system is found to be malleable as a result of interactions between sensory and cognitive processes (see Kraus & Banai, 2007, for review), revealing that auditory processing may be affected by experience, environmental influences, and active training (p. 105). Indeed, it has been shown that the effects of language experience on pitch processing in the auditory brainstem may vary depending on their perceptual relevance to the prosodic needs of a particular language (see Krishnan & Gandour, 2009, for review).

In the case of pitch, functional neuroimaging 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 interest is the observation that the inferior colliculus (IC) in the midbrain is more sensitive to changes in temporal regularity than the cochlear nucleus.

As a window into the early stages of subcortical pitch processing at the level of the midbrain, we use the human frequency-following response (FFR), a measure of electrophysiological activity in the auditory brainstem. The 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 generator source). The response is characterized by a periodic waveform which follows the individual cycles of the stimulus waveform. Experimental evidence points overwhelmingly to the IC as the source of the FFR generator. The shorter latency of the FFR (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). The very nature of the auditory system itself makes it unlikely that the low-pass filtered phase-locked activity reflected in the FFR is of cortical origin (Alkhoun et al., 2008).

When it comes to linguistically relevant pitch processing in the brain, *tone languages* in the Far East (Mandarin) and Southeast Asia (Thai) are of special interest because they use variations in pitch contrastively at the syllable level (Gandour, 1994; Yip, 2003). Such languages are to be distinguished from those in which variations in pitch are not contrastive at the syllable level (e.g., English). The crucial difference between a tone language and a non-tone language hinges on the phonemic status of pitch variations in the lexicon.

As reflected by FFRs, previous comparisons between native speakers of tone (Mandarin) and non-tone (English) languages show that native experience with lexical tones enhances pitch encoding at the level of the brainstem irrespective of speech or nonspeech context (Krishnan, Swaminathan, & Gandour, 2009; Krishnan, Xu, Gandour, & Cariani, 2005; Swaminathan, Krishnan, & Gandour, 2008). Moreover, language-dependent pitch encoding mechanisms appear to be especially sensitive to time-varying dimensions (e.g., acceleration) that span subparts rather than the whole of pitch contours (Krishnan, Swaminathan, et al., 2009; Xu, Krishnan, & Gandour, 2006). Using tri-linear and linear approximations to a natural, *curvilinear* pitch contour, no language-dependent effects are observed regardless of how close a *linear* pitch pattern approximates a native lexical tone (Krishnan, Gandour, Bidelman, & Swaminathan, 2009). Curvilinearity itself, though necessary, is insufficient to enhance pitch extraction of the auditory signal at the level of the brainstem. A *nonnative* curvilinear pitch pattern similarly fails to elicit a language-dependent effect. We therefore conclude that language-dependent neuroplasticity occurs only when salient dimensions of pitch in the auditory signal are part of the listener's experience and relevant to speech perception.

As far as we know, no study has yet to compare FFRs between two different tone languages. The aim of this study is to fill this knowledge gap by providing information on the lower and upper bounds of these language-dependent effects. We extend our previous work on language-dependent neural plasticity by including another tone language group (Thai) that will serve as either an experimental or control group depending on which set of lexical tones, Thai or Mandarin, respectively, are being presented. The choice of Mandarin and Thai gives us an opportunity to compare tone languages with phonological systems that are typologically similar in terms of number of tones (Mandarin, 4; Thai, 5) and type of tonal contours (Mandarin: 1 level, 2 rising, 1 falling; Thai: 2 level, 2 rising, 1 falling). Yet the voice fundamental frequency ( $f_0$ ) trajectories of their corresponding lexical tones exhibit varying degrees of similarity between one another. By choosing corresponding pairs of lexical tones

from the two languages that represent three degrees of phonetic similarity (high, moderate, low), we are able to assess the effects of phonetic similarity from long-term experience on the neural representation of pitch in the brainstem. The pair of Mandarin and Thai tones that exhibits highly similar  $f_0$  trajectories is Mandarin T4 (high falling) and Thai TF (high falling); moderately similar  $f_0$  trajectories is Mandarin T2 (high rising) and Thai TR (low falling-rising); dissimilar  $f_0$  trajectories is Mandarin T1 (high level) and Thai TM (mid falling).

If pitch representation in the brainstem is dependent on experience with tonal contours from a particular language, we can assess whether pitch representation varies between the native and nonnative tone language groups as a function of degree of phonetic similarity for corresponding tonal pairs. By including an English group, i.e., participants with no previous exposure to a tone language, we can evaluate whether the two tone language groups together show more robust pitch representation than naïve listeners regardless of the language identity of individual lexical tones. Of significance to the neurobiology of language, these experimental outcomes would suggest that as in the cerebral cortex, early, subcortical pitch encoding may be influenced by long-term exposure to linguistically relevant parameters of the auditory signal.

## 2. Methods

### 2.1. Participants

Ten adult native speakers of Mandarin Chinese (5 male, 5 female), hereafter referred to as Chinese (C), 10 adult native speakers of Thai (5 male, 5 female), hereafter referred to as Thai (T), and 10 adult native speakers of American English (4 male, 6 female), hereafter referred to as English (E), participated in the fFR experiment. The three groups were closely matched in age (Chinese:  $M = 27.4$ ,  $SD = 3.1$ ; Thai:  $M = 25.8$ ,  $SD = 4.7$ ; English:  $M = 22.8$ ,  $SD = 3.2$ ), education (Chinese:  $M = 17.6$ ,  $SD = 3.1$ ; Thai:  $M = 17.7$ ,  $SD = 2.6$ ; English:  $M = 16.3$ ,  $SD = 1.8$ ), and were strongly right handed ( $\geq 85\%$ ) as measured by the Edinburgh Handedness inventory (Oldfield, 1971). All participants exhibited normal hearing sensitivity (better than 20 dB HL in both ears) at octave frequencies from 500 to 4000 Hz. In addition, participants reported no previous history of neurological or psychiatric illnesses. Each participant completed a language history questionnaire (Li, Sepanski, & Zhao, 2006). Native speakers of Mandarin and Thai were born and raised in mainland China or Thailand, respectively, and none had received formal instruction in English before the age of 6 ( $M = 10.6$ ,  $SD = 2.2$ ). Chinese participants had no previous exposure to Thai, and conversely, Thai participants had no previous exposure to Mandarin. The American English group had no prior experience with any tonal language whatsoever. All participants completed a music history questionnaire (Wong & Perrachione, 2007). None of the participants had more than three years of formal music training on any combination of instruments nor had studied music within the past five years. All individuals were students, enrolled at Purdue University at the time of their participation. All were paid for their participation and gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

### 2.2. Stimuli

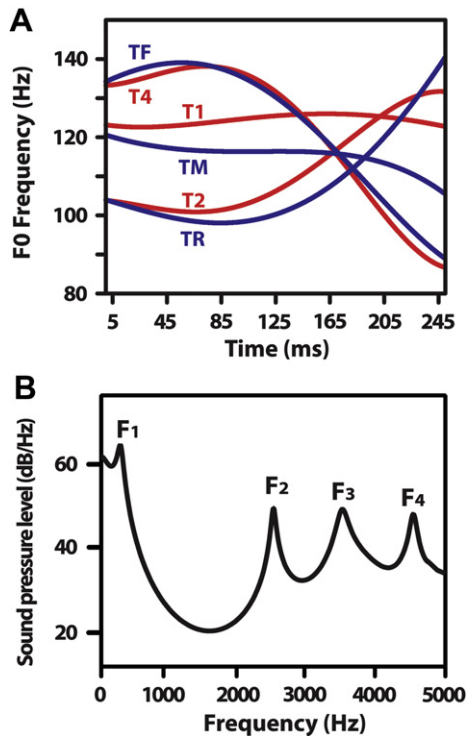
Stimuli consisted of a set of three Mandarin and three Thai words that are distinguished minimally by tonal contour. Using a cascade/parallel formant synthesizer (Klatt, 1980; Klatt & Klatt, 1990), a synthetic version of the syllable/yi/with Mandarin tone 1 was created. Vowel formant frequencies were steady-state and were held constant (in Hz):  $F1 = 300$ ;  $F2 = 2500$ ;  $F3 = 3500$ ; and  $F4 = 4530$ . All stimuli were time-normalized to 250 ms. Tone 1 was chosen as a template for analysis-resynthesis as it exhibits less pitch movement than the other five tones. The various Mandarin and Thai pitch patterns were overlaid onto this synthetic vowel using a pitch-synchronous overlap and add (PSOLA) algorithm implemented in Praat (Boersma & Weenink, 2008). Synthesis parameters for  $f_0$  contours were modeled after Mandarin (Xu, 1997) and Thai (Abramson, 1962)  $f_0$  contours that occur in natural speech citation forms using 4th-order polynomial equations (Xu, Gandour, et al., 2006). The three Mandarin words ( $yi^1$  'clothing' [T1];  $yi^2$  'aunt' [T2];  $yi^4$  'easy' [T4]) and three Thai words ( $yi^M$  'derogatory

title' [TM];  $yi^R$  'nasal sound' [TR];  $yi^F$  'candy' [TF] had identical vowel quality ([i]), duration (250 ms), and formant structure. Yet all three words within their respective language were minimally distinguished by  $f_0$  trajectories (Fig. 1). Phonetically speaking, Mandarin tones T1, T2, and T4 have been described as high level, high falling, and high rising, respectively; Thai tones TM, TR, and TF as mid falling, low falling-rising, and high falling, respectively (Ladefoged, 2006, p. 251).

### 2.3. Data acquisition

FFR recording protocol and data analysis are similar to those reported in previous publications from our laboratory (Krishnan, Gandour, et al., 2009; Krishnan, Swaminathan, et al., 2009). Participants reclined comfortably in an acoustically and electrically shielded booth. They were instructed to relax and refrain from extraneous body movements to minimize myogenic artifacts. FFRs were recorded from each participant in response to monaural stimulation of the right ear at  $\sim 82$  dB SPL at a repetition rate of 2.76/s. The presentation order of the stimuli was randomized both within and across participants. Control of the experimental protocol was accomplished by a signal generation and data acquisition system (Tucker-Davis Technologies, System III). The stimulus files were routed through a digital to analog module and presented through a magnetically shielded insert earphone (Bio-logic, ER-3A).

FFRs were recorded differentially between a non-inverting (positive) electrode placed on the midline of the forehead at the hairline (Fz) and inverting (reference) electrodes placed on (i) the



**Fig. 1.** Voice fundamental frequency ( $f_0$ ) contours of synthetic speech stimuli modeled after Mandarin (panel A, red) and Thai (panel A, blue)  $f_0$  contours that occur in natural speech productions of citation forms using 4th-order polynomial equations (Xu, Gandour, et al., 2006). Mandarin T4, T1, and T2 represent high falling, high level, and high rising tones, respectively. Thai TF, TM, and TR represent high falling, mid falling, and low falling-rising tones, respectively. All stimuli were matched in total duration and RMS amplitude. Acoustic spectral envelope (panel B), containing four steady-state formants ( $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_4$ ), was identical across all six stimuli.

ipsilateral mastoid (A2); (ii) the contralateral mastoid (A1); and (iii) the 7th cervical vertebra (C7). Another electrode placed on the mid-forehead (Fpz) served as the common ground. FFRs were recorded simultaneously from the three different electrode configurations, and subsequently averaged for each stimulus condition to yield a response with a higher signal-to-noise ratio (Krishnan, Gandour, et al., 2009). All inter-electrode impedances were maintained at or below 1 k $\Omega$ . The raw EEG inputs were amplified by 200,000 and band-pass filtered from 80 to 3000 Hz (6 dB/octave roll-off, RC response characteristics). Each FFR waveform represents the average of 3000 stimulus presentations over a 280 ms analysis window using a sampling rate of 24,414 kHz.

## 2.4. Data analysis

### 2.4.1. Pitch-tracking accuracy of whole tones

The ability of the FFR to follow pitch changes in the stimuli was evaluated by extracting the  $f_0$  contour from the FFRs using a periodicity detection short-term autocorrelation algorithm (Boersma, 1993). Essentially, the algorithm works by sliding a 40 ms window in 10 ms increments over the time course of the FFR. The autocorrelation function was computed for each 40 ms frame and the time-lag corresponding to the maximum autocorrelation value within each frame was recorded. The reciprocal of this time-lag (or pitch period) represents an estimate of  $f_0$ . The time-lags associated with autocorrelation peaks from each frame were concatenated together to give a running  $f_0$  contour. This analysis was performed on both the FFRs and their corresponding stimuli. Pitch-tracking accuracy is computed as the cross-correlation coefficient between the  $f_0$  contour extracted from the FFRs and the  $f_0$  contour extracted from the stimuli.

### 2.4.2. Pitch strength of tonal sections

To compute the pitch strength of the FFRs to *time-varying* stimuli, FFR responses were divided into six non-overlapping 40 ms sections (5–45; 45–85; 85–125; 125–165; 165–205; 205–245 ms). The normalized autocorrelation function (expressed as a value between 0 and 1) was computed for each of these sections, where 0 represents an absence of periodicity and 1 represents maximal periodicity. Within each 40 ms section, a response peak was selected which corresponded to the same location (time-lag) of the autocorrelation peak in the input stimulus (Krishnan, Gandour, et al., 2009; Krishnan, Swaminathan, et al., 2009; Swaminathan et al., 2008). The magnitude of this response peak represents an estimate of the pitch strength per section. All data analyses were performed using custom routines coded in MATLAB 7 (The MathWorks, Inc., Natick, MA).

## 2.5. Statistical analysis

### 2.5.1. Pitch-tracking accuracy of whole tones

Pitch-tracking accuracy was measured as the cross-correlation coefficient between the  $f_0$  contours extracted from the FFRs and tonal stimuli. A repeated measures, mixed model ANOVA (SAS<sup>®</sup>) – with group (Chinese, Thai, English) as the between-subject factor, tone (T1, T2, T4: TM, TR, TF) as the within-subject factor nested within the language identity factor (Mandarin, Thai), and subjects as a random factor nested within language group – was conducted on the cross-correlation coefficients to evaluate the effects of language experience on the ability of the FFR to track time-varying  $f_0$  information in Mandarin and Thai.

### 2.5.2. Pitch strength of tonal sections

Pitch strength (magnitude of the normalized autocorrelation peak) was calculated for each of the six sections within each of the three Mandarin and three Thai tones for every subject. These pitch-strength values were analyzed using a mixed model ANOVA with subjects as a random factor nested within language group (Chinese, Thai, English), and with two fixed, within-subject factors, tone (T1, T2, T4: TM, TR, TF) and section (5–45, 45–85, 85–125, 125–165, 165–205, 205–245 ms). Both tone and section were nested within the language identity factor (Mandarin, Thai). By focusing on 40-ms sections within these time-varying  $f_0$  contours, we were able to evaluate whether the effects of language

experience vary depending on velocity and/or acceleration per section irrespective of tonal category or language identity.

### 2.5.3. Pitch strength of tonal dimensions

For this purpose, Mandarin and Thai tones were concatenated together without any consideration for their language identity. We then identified four 40 ms  $f_0$  sections across the six tones that were differentiated on the basis of acceleration ( $\alpha$ , change in pitch per unit time) relative to the section with minimum acceleration, TM @ 85–125 ms ( $\alpha = 0.0000$ ) (cf. Fig. 1). The section with the highest negative acceleration coefficient, *maximum fall*, was extracted from T4 @ 165–205 ms ( $\alpha = -0.0181$ ); the acceleration coefficient midway between the maximum negative and minimum, *medium fall*, was extracted from TR @ 5–45 ms ( $\alpha = -0.0039$ ); the highest positive acceleration coefficient, *maximum rise*, was extracted from TR @ 165–205 ms ( $\alpha = 0.0184$ ); the acceleration coefficient midway between the maximum positive and minimum, *medium rise*, was extracted from T2 @ 85–125 ms ( $\alpha = 0.0049$ ). Pitch strength was calculated from the FFRs of every subject for these four  $f_0$  sections.

A discriminant analysis was conducted to determine the weighted linear combination of these four tonal dimensions that best discriminate between the three language groups (Chinese, Thai, English). The discriminant function was cross-validated using a  $k$ -fold cross-validation, i.e., pitch-strength weights from each subject were used as validation data for the function created from weights of the remaining subjects. This was repeated until every one of the 30 subjects was used for validation.

## 3. Results

### 3.1. Pitch-tracking accuracy of lexical tones

FFR pitch-tracking accuracy, as measured by the time-lag associated with the autocorrelation maximum per language group, is shown for each of the three Mandarin and Thai tones (Fig. 2). Overall pitch tracking is observed to be more accurate for the tone language groups (Chinese, Thai) compared to the non-tone language group (English). That is, on the whole,  $f_0$  contours derived from the FFR waveforms of the Chinese and Thai groups more closely approximate those of the original IRN stimuli.

An omnibus ANOVA on cross-correlation coefficients for Mandarin tones yielded significant main effects of group ( $F_{2,27} = 27.16$ ,  $p < 0.0001$ ) and tone ( $F_{2,54} = 21.75$ ,  $p < 0.0001$ ). The group  $\times$  tone interaction effect was not significant ( $F_{4,54} = 1.52$ ,  $p = 0.2101$ ). For T1, T2, and T4, a priori contrasts of

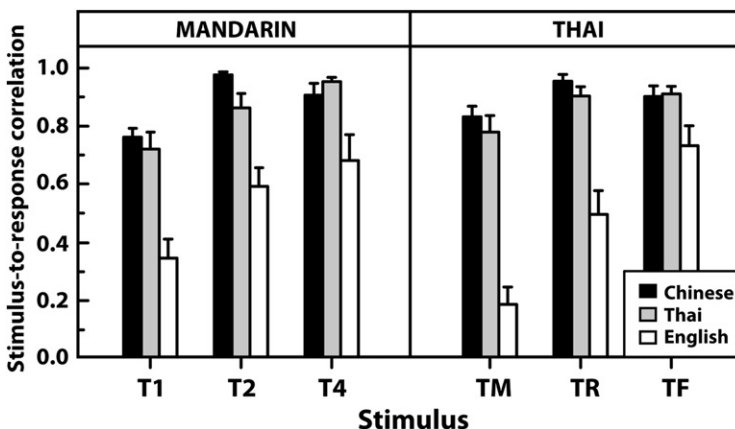


Fig. 2. A comparison of FFR pitch-tracking accuracy in response to Mandarin and Thai tones between language groups (Chinese, Thai, English). Bars represent the group means of the stimulus-to-response correlation coefficients; error bars =  $\pm 1$  SE. Both Chinese and Thai are better able to track pitch contours as compared to English regardless of their language identity, as well as to transfer their pitch-tracking abilities across tone languages.

groups ( $\alpha_{\text{Bonferroni}} = 0.0166$ ) revealed that pitch-tracking accuracy of the tone language groups was significantly higher than the English, whereas no significant difference in pitch-tracking accuracy was observed between the Chinese and Thai groups.

An omnibus ANOVA on cross-correlation coefficients for Thai tones yielded significant main effects of group ( $F_{2,27} = 49.57, p < 0.0001$ ) and tone ( $F_{2,54} = 22.58, p < 0.0001$ ), as well as a group  $\times$  tone interaction effect ( $F_{4,54} = 7.65, p < 0.0001$ ). In the case of a priori contrasts of groups ( $\alpha_{\text{Bonferroni}} = 0.0166$ ), pitch-tracking accuracy of TM and TR was significantly higher in the tone language groups as compared to the English. For TF, none of the between-group comparisons reached significance. In the case of a priori contrasts of Thai tones ( $\alpha_{\text{Bonferroni}} = 0.0166$ ), pitch-tracking accuracy of TR and TF was significantly higher than TM in the English group; TR was higher than TM in the Chinese group. None of the between-tone comparisons reached significance in the Thai group.

### 3.2. Pitch strength of sections within lexical tones

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 three Mandarin and Thai tones (Fig. 3). Except for section 1 of TF, pitch strength is observed to be greater across the board for the tone language groups (Chinese, Thai) compared to the non-tone language group (English).

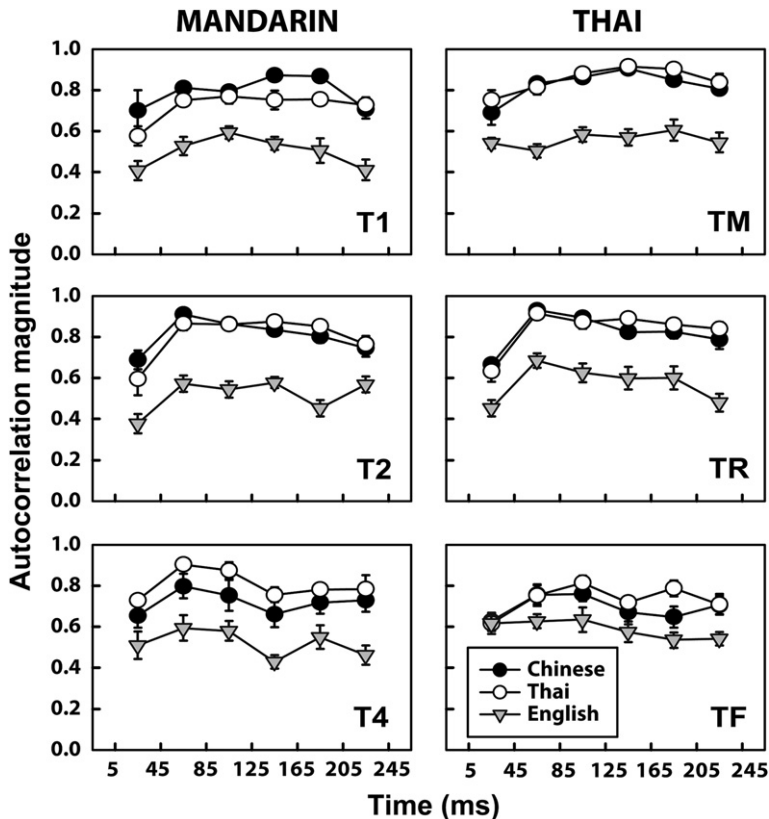


Fig. 3. Pitch strength of Mandarin (left) and Thai (right) dimensions within each of their six sections derived from the averaged FFR waveforms of the three language groups (Chinese, Thai, English). Pitch strength of the Chinese and Thai groups is greater than that of the English group virtually across the board regardless of the language identity of tones or location of sections therein. Moreover, long-term experience with either Mandarin or Thai tones appears to be mutually beneficial, transferring from one tone language to the other. Six 40-ms sections are demarcated as follows: 5–45, 45–85, 85–125, 125–165, 165–205, 205–245.

For each of the three Mandarin tones, omnibus ANOVAs on FFR pitch strength revealed significant main effects of group (T1,  $F_{2,27} = 57.53$ ,  $p < 0.0001$ ; T2,  $F = 55.73$ ,  $p < 0.0001$ ; T4,  $F = 15.07$ ,  $p < 0.0001$ ) and section (T1,  $F_{5,135} = 10.03$ ,  $p = 0.0017$ ; T2,  $F = 18.18$ ,  $p < 0.0001$ ; T4,  $F = 6.62$ ,  $p < 0.0001$ ). The group  $\times$  section interaction effect was not significant for any of the three Mandarin tones. For each of the six sections within T1, T2, and T4, a priori contrasts of groups ( $\alpha_{\text{Bonferroni}} = 0.0166$ ) revealed that pitch strength of the tone language groups was significantly greater than the English, whereas no significant differences in pitch strength were observed between the Chinese and Thai groups.

For each of the three Thai tones, omnibus ANOVAs on FFR pitch strength revealed significant main effects of group (TM,  $F_{2,27} = 47.54$ ,  $p < 0.0001$ ; TR,  $F = 42.40$ ,  $p < 0.0001$ ; TF,  $F = 5.47$ ,  $p = 0.0101$ ) and section (TM,  $F_{5,135} = 8.72$ ,  $p < 0.0001$ ; TR,  $F = 23.86$ ,  $p < 0.0001$ ; TF,  $F = 4.91$ ,  $p = 0.0004$ ). The group  $\times$  section interaction effect was not significant for any of the three Thai tones. For each of the six sections across the three Thai tones, a priori contrasts of groups ( $\alpha_{\text{Bonferroni}} = 0.0166$ ) revealed no significant differences in pitch strength between the Chinese and Thai groups. For TM and TR, pitch strength of the tone language groups was also significantly greater than the English. In the case of TF, three of the six sections (5–45, 45–85, 125–165 ms) showed that pitch strength of the English group was not significantly different from the tone language groups; two sections (85–125, 165–205) showed that pitch strength was greater in Thai than English; and one section (205–245) showed that pitch strength was greater for both tone language groups as compared to English.

### 3.3. Pitch strength of dimensions underlying lexical tones

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 four 40-ms time windows that were differentiated on the basis of direction and degree of slope: maximum rise, medium rise, maximum fall, medium fall. The classification matrix from this discriminant analysis is presented in Table 1. Overall, 83% of subjects were correctly classified into their respective language groups (Chinese, 70%; Thai, 80%; English 100%). Because we can expect to get only 33% of the classifications correct by chance, an overall 83% accuracy rate represents a considerable improvement (canonical correlation = 0.90). But considerably fewer correct classifications (63%) were made in the cross-validated analysis (Chinese, 3/10; Thai, 6/10; English, 9/10) in comparison to the original analysis. Of the misclassifications, 10 out of 12 were assigned incorrectly between the two tone language groups. The group centroids, i.e., average discriminant  $z$  scores, were 1.3052, 1.6011, and  $-2.9063$  for the Chinese, Thai, and English groups, respectively. A one-way ANOVA on discriminant scores revealed a significant group effect ( $F_{2,27} = 63.57$ ,  $p < 0.0001$ ). Post hoc Tukey–Kramer comparisons indicated that the two tone language groups were significantly ( $p < 0.0001$ ) different from English, but not from one another ( $p = 0.7875$ ).

Of the two discriminant functions in this three-group analysis, the first accounted for 89.8% of the total variance. Its pooled within-class standardized canonical coefficients of the four pitch dimensions were 0.8252 (medium rise), 0.3655 (maximum fall), 0.3466 (maximum rise), and  $-0.0284$  (medium fall). The fact that the first dimension was 2.26 times as important as the second indicates that medium

**Table 1**

Classification matrix for three-group discriminant analysis as a function of pitch strength of four trajectories as defined by direction and degree of slope.

Actual group	<i>n</i>	Predicted group		
		Chinese	Thai	English
Chinese	10	7	3	0
Thai	10	2	8	0
English	10	0	0	10
Total	30	9	11	10
Priors		0.3333 <sup>a</sup>	0.3333	0.3333

Note. Numbers on the diagonal represent correct classifications; off-diagonal numbers represent incorrect classifications.

<sup>a</sup> Prior probabilities based on actual group size.



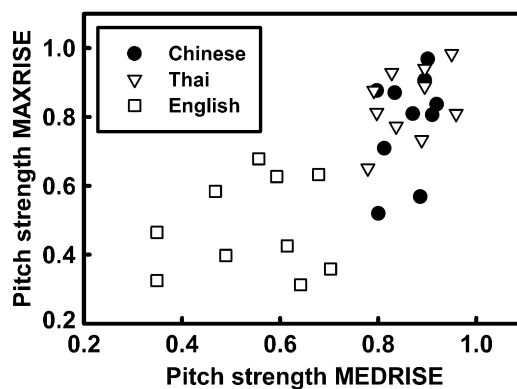
rise was the most important variable for differentiating the English, Chinese, and Thai groups. The second function failed to reach significance, and was comparatively weak in its ability to separate the groups, accounting for only 7.8% of the variance of the total variance. Using only one discriminant function yielded the same number of correct classifications. The importance of the medium rise dimension in discriminating participants by language affiliation is consistent with individual subjects' weighting of pitch-strength dimensions (Fig. 4).

#### 4. Discussion

Using synthetic speech stimuli that contain  $f_0$  contours representative of citation forms of Mandarin and Thai lexical tones, the major finding of this study demonstrates that experience-dependent brainstem mechanisms for pitch representation, as reflected in pitch-tracking accuracy and pitch strength, are more sensitive in tone (Chinese, Thai) than non-tone (English) language speakers. No matter the degree of phonetic similarity between corresponding tones from the two languages, Chinese and Thai are both able to transfer their abilities in pitch encoding across languages. Brainstem neurons appear to be differentially sensitive to changes in pitch without regard to their language identity as long as they occur in a language with a comparable phonological system. As reflected in a discriminant analysis of variables derived from tonal sections differing in magnitude of rising and falling pitch, moderate rises turn out to be the most important pitch dimension for distinguishing tone language from non-tone language speakers. This finding suggests that instead of tonal categories, pitch extraction in the brainstem is driven by specific acoustic dimensions or features of pitch contours.

##### 4.1. Experience-dependent neural plasticity in the brainstem

These cross-language comparisons support the idea that experience-driven adaptive subcortical mechanisms sharpen response properties of pitch-extraction neurons that are sensitive to the prosodic needs of a particular language. In the case of Mandarin and Thai, pitch variations are lexically relevant in monosyllabic words. The stimuli herein represent prototypical tonal contours as produced in citation form. The data acquisition protocol for eliciting FFRs effectively presents a sequence of identical tonal contours in isolation ( $\approx 3000$ ). Such pitch contours are linguistically relevant for native speakers of either Mandarin or Thai; they are not relevant for native speakers of English. It is unlikely that lexical identity can account for the language group outcomes on pitch-tracking accuracy and pitch strength.



**Fig. 4.** Pitch-strength values projected in two dimensions (medium rise, medium fall) derived from individual Chinese and English subjects' FFR waveforms in response to two sections defined in terms of direction and degree of slope. In order of importance, medium rise ranks first out of the four variables that were entered into the discriminant function; medium fall ranks last. Chinese and Thai individuals tend to cluster together in the rightmost third of the subject space, indicating relatively higher pitch strength in response to moderate increases in pitch as compared to the English individuals. On the other hand, considerable overlap in the distribution of subjects is observed on the medium fall dimension, indicating that moderate decreases in pitch are much less effective in discriminating between speakers of tone and non-tone languages.

No matter that Mandarin T1 (high level) and Thai TM (mid falling) are dissimilar in meaning and form; we still observe comparable sensitivity to either stimulus regardless of tone language background of the listener.

From a neuroethological perspective, this subcortical adjustment in pitch extraction of lexical tones by native speakers of Mandarin or Thai 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 begin well before the signal reaches the cerebral cortex.

The fact that pitch experience in one language can positively benefit neural processing in another with a similar prosodic system is not surprising. Similar benefits of pitch experience have been observed even across domains. English-speaking musicians show better performance in the identification of Mandarin tones than non-musicians (Lee & Hung, 2008). At the level of the cerebral cortex, electrophysiological indices show that music training facilitates pitch processing in language (Magne, Schon, & Besson, 2006; Schon, Magne, & Besson, 2004). At the level of the brainstem, pitch-tracking accuracy of Mandarin tones is more accurate in nonnative musicians than non-musicians (Wong, Skoe, Russo, Dees, & Kraus, 2007). Indeed, pitch-tracking accuracy improves even in English-speaking non-musicians after undergoing short-term training using Mandarin tones in word identification (Song, Skoe, Wong, & Kraus, 2008).

#### 4.2. Neural mechanisms underlying experience-dependent pitch encoding in the brainstem

##### 4.2.1. Corticofugal modulation

The observed experience-dependent enhancement of pitch representation in the tonal groups could, in part, reflect corticofugal influence. Several theoretical constructs have been proposed that invoke corticofugal influence to modulate subcortical sensory processing. The reverse hierarchy theory (Ahissar & Hochstein, 2004; Nahum, Nelken, & Ahissar, 2008) provides a representational hierarchy to describe the interaction between sensory input and top-down processes to guide plasticity in primary sensory areas. This theory suggests 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 percept. This theory has been invoked as a plausible explanation for top-down influences on subcortical sensory processing (Banai, Abrams, & Kraus, 2007).

Another proposed circuitry mediating learning-induced plasticity is the cortico-colliculo-thalamo-cortico-collicular (CTCC) loop (Xiong, Zhang, & Yan, *in press*). This framework with its bottom-up (tonotopic colliculo-thalamic and thalamo-cortical projections) and top-down (corticofugal projections) forms a tonotopic CTCC loop that presumably is the only neural substrate carrying accurate auditory information (cf. Krishnan & Gandour, 2009). Additionally, the CTCC loop incorporates several neuromodulatory inputs forming a core neural circuit mediating sound specific plasticity associated with perceptual learning. Auditory stimuli and neuromodulatory inputs are believed to induce large scale frequency-specific plasticity in the loop.

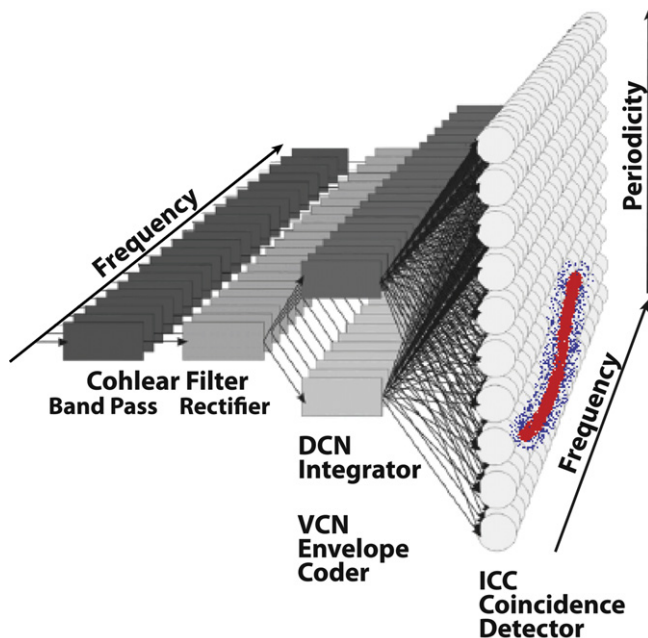
Top-down guided plasticity may also be mediated by memory based theories (Goldinger, 1998; Pasternak & Greenlee, 2005). Pasternak and Greenlee (2005) propose sensory specific memory guiding plasticity instead of prefrontal and parietal, to increase the probability of creating accurate sensory memory traces by enhancing encoding at the sensory level. Long-term stored episodes of native pitch patterns (similar to pitch templates) may also guide plasticity by engaging the corticofugal pathways (Goldinger, 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, online 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, Gao, Zhang, Ma, & Olsen, 2000; Suga & Ma, 2003; Suga et al., 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 permanent corticofugal influence. It is important to note here that the corticofugal influence will likely come into play in degraded listening conditions and during training protocols.

#### 4.2.2. Local brainstem mechanisms

The enhanced pitch representation in the tonal groups reflects an enhanced tuning to interspike 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 (Fig. 5), we adopt the temporal correlation analysis model described by Langner (1992, 2004). Coincidence detection neurons in the IC perform a correlation analysis on the delayed and un-delayed 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 demonstrated by our Mandarin and Thai speakers, and reflected by smoother tracking of whole pitch contours and greater pitch strength of 40-msec sections thereof. Cross-language comparisons further reveal that this encoding scheme is more sensitive to dynamic segments of pitch contours in the tonal groups (Chinese and Thai) 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



**Fig. 5.** 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. DCN, dorsal cochlear nucleus; VCN, ventral cochlear nucleus; ICC, inferior colliculus (Langner, 2004, adapted with kind permission of Springer Science and Business Media).

an active facilitation/disinhibition of the pitch intervals corresponding to the dynamic segments and inhibition of other pitch periods. Neuromodulatory inputs (Xiong et al., *in press*) could also influence the balance between excitation and inhibition.

Several lines of evidence support the invocation of local mechanisms in the brainstem mediating the language experience-dependent plasticity observed in our data. First, the corticofugal egocentric selection is short term and takes time (latency) to be activated whereas the FFR response latency is only about 6–9 ms. Even considering the sustained portion of our FFR response (260 ms data acquisition window), 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 a dynamic pitch pattern over its entire duration. Also, the consideration of the more caudal MOCB reflex would require us to invoke bottom-up processes that can influence the plasticity seen at the IC level. Second, 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).

Third, 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, in contrast to curvilinear  $f_0$  contours (Krishnan, Gandour, et al., 2009; Krishnan, Swaminathan, et al., 2009; 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, Gandour, & Francis (2006) and 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 native 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.

#### 4.3. *Pitch dimensions at multiple stages of processing*

Our finding that medium rises in pitch turn out to be the most important dimension for distinguishing tone language from non-tone language speakers suggests that pitch extraction at the level of the brainstem is driven by specific dimensions or features of pitch contours. This preeminence of rising pitch movements is consistent with the extant literature on rising and falling pitch movements. In the brainstem, the FFR amplitude for falling tonal sweeps is smaller than for rising (Krishnan & Parkinson, 2000). In the cochlea, microphonic potentials show that displacements of the cochlear partition occur closer together in time for rising than for falling tonal sweeps (Shore & Cullen, 1984). Specifically, the eighth-nerve compound action potential evoked by a rising tonal sweep is larger in amplitude than falling (Shore & Nuttall, 1985). This asymmetry is further supported by differential temporal response patterns to rising and falling tones for single units in the ventral cochlear nucleus (Shore, Clopton, & Au, 1987). Our findings lead us to infer that differential weighting of rising and falling  $f_0$  contours for tone vs. non-tone language speakers reflects a language-dependent enhancement of universal, temporal response patterns to rising and falling pitch among the neural elements generating the FFR. Consistent with this notion is our earlier observation of more robust FFR representation of harmonics in the dominant region for pitch in response to rising as compared to falling contours (Krishnan et al., 2005).

This pitch direction asymmetry is further supported by human psychophysical data. The just-noticeable difference for change in frequency of sweep-tone stimuli are smaller for rising than for falling tones (Shore et al., 1987), and detection thresholds for rising tones are lower than those for falling (Collins & Cullen, 1978; Cullen & Collins, 1979; Nabelek, 1978). Discrimination of falling sweeps requires longer durations and/or higher sweep rates than rising sweeps (Schouten, 1985). Detecting

changes in slope of linear  $f_0$  ramps show the greatest sensitivity when one ramp is rising and the other is falling (Klatt, 1973). Based on multidimensional scaling analyses of tonal stimuli varying in height and direction, the underlying perceptual dimension in the stimulus space related to direction separates rising vs. non-rising  $f_0$  movements (Gandour, 1983; Gandour & Harshman, 1978).

In speech perception, it remains an open question as to how the incoming auditory signal is transformed into an abstract lexical representation (see Poeppel, Idsardi, & van Wassenhove, 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 the nature of subcortical representations, our FFR data lead us to hypothesize that they are based on scalar or  $n$ -ary features. In spite of their continuous nature, there is considerable evidence to support the notion that distinctive features in the lexicon are grounded in lower-level sensory dimensions or features that emerge along the auditory pathway (Krishnan & Gandour, 2009). 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).

## Acknowledgements

Research supported by NIH R01 DC008549 (A.K.) and NIDCD predoctoral traineeship (G.B.). Thanks to Bruce Craig and Yang Zhao for their assistance with statistical analysis (Department of Statistics). Reprint requests should be addressed to Ananthanarayan Krishnan, Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, IN, USA 47907-2038, or via email: rkrish@purdue.edu.

## Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at [doi:10.1016/j.jneuroling.2009.09.001](https://doi.org/10.1016/j.jneuroling.2009.09.001)

## References

- Abramson, A. S. (1962). *The vowels and tones of standard Thai: Acoustical measurements and experiments*. Bloomington, Indiana, USA: Research Center in Anthropology, Folklore, and Linguistics. Pub. 20.
- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10), 457–464.
- 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.
- Ananthanarayan, 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.
- Ananthanarayan, 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.
- 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.
- Boersma, P., & Weenink, D. (2008). *Praat: Doing phonetics by computer (Version 5.0.40)*. Amsterdam: Institute of Phonetic Sciences. [Computer program].
- Collins, M. J., & Cullen, J. K., Jr. (1978). Temporal integration of tone glides. *Journal of the Acoustical Society of America*, 63(2), 469–473.
- Cullen, J. K., & Collins, M. J. (1979). Rate effects in the detection of short duration tonal glides. *Journal of the Acoustical Society of America*, 65(Suppl. 1), S38.
- 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.
- Galbraith, G. C. (2008). Deficient brainstem encoding in autism. *Clinical Neurophysiology*, 119(8), 1697–1700.
- Galbraith, G. C., 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. (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 & linguistics*, Vol. 6 (pp. 3116–3123). New York: Pergamon Press.

- 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.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.
- 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.
- 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.
- 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.
- 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.
- Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67(3), 971–995.
- Klatt, D., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87(2), 820–857.
- 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.
- 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. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*. doi:10.1016/j.bandl.2009.03.005.
- 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. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *Journal of Cognitive Neuroscience*, 21(6), 1092–1105.
- 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.
- Ladefoged, P. (2006). *A course in phonetics* (5th ed.). Boston, MA: Thomson Wadsworth.
- 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.
- Lee, C. Y., & Hung, T. H. (2008). Identification of mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America*, 124(5), 3235–3248.
- Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: a web-based interface for bilingual research. *Behavioral Research Methods*, 38(2), 202–210.
- Magne, C., Schon, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18(2), 199–211.
- 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), 126.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, 6(2), 97–107.
- 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.
- Schon, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*, 41(3), 341–349.
- Schouten, M. E. (1985). Identification and discrimination of sweep tones. *Perception and Psychophysics*, 37(4), 369–376.
- 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.
- Song, J. H., Skoe, E., Wong, P. C., & 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. (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. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *Neuroreport*, 19(11), 1163–1167.
- Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(4), 565–585.
- 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. The neurobiology of sound-specific auditory plasticity: a core neural circuit. *Neuroscience and Biobehavioral Reviews*, in press.
- Xu, Y. (1997). Contextual tonal variations in mandarin. *Journal of Phonetics*, 25, 61–83.
- Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, 120(2), 1063–1074.
- Xu, Y., Gandour, J. T., Talavage, T., Wong, D., Dziedzic, M., Tong, Y., et al. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, 27(2), 173–183.
- 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.