

# Experience-dependent Enhancement of Linguistic Pitch Representation in the Brainstem Is Not Specific to a Speech Context

Ananthanarayan Krishnan, Jayaganesh Swaminathan,  
and Jackson T. Gandour

## Abstract

■ Neural representation of pitch is influenced by lifelong experiences with music and language at both cortical and subcortical levels of processing. The aim of this article is to determine whether neural plasticity for pitch representation at the level of the brainstem is dependent upon specific *dimensions* of pitch contours that commonly occur as part of a native listener's language experience. Brainstem frequency following responses (FFRs) were recorded from Chinese and English participants in response to four Mandarin tonal contours presented in a nonspeech context in the form of iterated rippled noise. Pitch strength (whole contour, 250 msec; 40-msec segments) and pitch-tracking accuracy (whole contour) were extracted from the FFRs using autocorrelation algorithms. Narrow band spectrograms were used to extract spectral information. Results showed that the Chinese group exhibits smoother pitch tracking than the English group in

three out of the four tones. Moreover, cross-language comparisons of pitch strength of 40-msec segments revealed that the Chinese group exhibits more robust pitch representation of those segments containing rapidly changing pitch movements across all four tones. FFR spectral data were complementary showing that the Chinese group exhibits stronger representation of multiple pitch-relevant harmonics relative to the English group across all four tones. These findings support the view that at early preattentive stages of subcortical processing, neural mechanisms underlying pitch representation are shaped by particular dimensions of the auditory stream rather than speech per se. Adopting a temporal correlation analysis scheme for pitch encoding, we propose that long-term experience sharpens the tuning characteristics of neurons along the pitch axis with enhanced sensitivity to linguistically relevant variations in pitch. ■

## INTRODUCTION

Pitch is one of the most important information-bearing parameters of species-specific vocal signals (Plack, Oxenham, & Fay, 2005). In speech, voice pitch conveys information concerning emotion, attitude, and talker identity. In music, sequences of pitch comprise melodies. In tone languages (e.g., Mandarin), pitch variations on individual syllables convey part of word meaning. The study of the physiological mechanisms that underlie pitch perception can illuminate the neural basis of auditory processing in both linguistic and nonlinguistic domains. Most periodic complex sounds including speech evoke a sensation of low pitch associated with their voice fundamental frequency ( $f_0$ ) (Moore, 1989). The physiological bases of pitch perception are still a matter of debate. One view is that the auditory system uses the spatial discharge rate patterns along tonotopically organized neural maps to represent the spectrum of complex sounds (spectral pitch). Pitch is then extracted by pattern recognition mechanisms that de-

tect patterns of excitation produced by harmonically related components (Cohen, Grossberg, & Wyse, 1995; Goldstein, 1973).

Another view is that pitch extraction mechanisms are based on the timing of auditory nerve fiber activity (temporal pitch) irrespective of frequency organization (Cariani & Delgutte, 1996a, 1996b; Meddis & Hewitt, 1991). Most temporal models of pitch perception are based on the timing information available in the interspike intervals represented in the simulated (de Cheveigne, 1998; Meddis & O'Mard, 1997) or actual (Cariani & Delgutte, 1996a, 1996b) auditory nerve activity. Such temporal models derive a pitch estimate by pooling timing information across auditory nerve fibers without regard to the frequency-to-place mapping by means of an autocorrelation like analysis performed on the all-order interspike interval distribution (Shofner, 1999; Cariani & Delgutte, 1996a, 1996b). For a variety of periodic complex sounds, the largest peak in the population all-order interspike intervals corresponds closely to the perceived low pitch (*predominant interval hypothesis*). Temporal encoding schemes provide a unified and parsimonious way of explaining a diverse range of

Purdue University

pitch phenomena (Meddis & O'Mard, 1997). Thus, we may conclude that neural phase-locked activity related to  $f_0$  plays a dominant role in the encoding of pitch associated with complex sounds.

The scalp-recorded human frequency following response (FFR) reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem (Glaser, Suter, Dasheiff, & Goldberg, 1976; Smith, Marsh, & Brown, 1975; Marsh, Brown, & Smith, 1974). Most studies suggest that the longer latency scalp-recorded FFRs (>6 msec) are primarily generated in the inferior colliculus (IC) (Galbraith, Bagasan, & Sulahian, 2001; Sohmer & Pratt, 1977; Glaser et al., 1976; Smith et al., 1975; Marsh et al., 1974). The human FFR preserves steady-state and time-variant acoustic features that are present in the speech spectrum (Krishnan, 2002; Plyler & Ananthanarayan, 2001; Krishnan & Parkinson, 2000). Pitch information has been shown to be preserved in the phase-locked neural activity generating the FFR not only for steady-state complex tones (Greenberg, Marsh, Brown, & Smith, 1987) but also for time-varying pitch contours of lexical tones (Krishnan, Xu, Gandour, & Cariani, 2004). Thus, the FFR provides a noninvasive window to view neural processing of pitch at the level of the auditory brainstem.

In animals, it is already well established that experience-dependent neural plasticity is not limited to the cerebral cortex (Suga, Gao, Zhang, Ma, & Olsen, 2000; Suga, 1990). But it is also evident at the level of the human brainstem. For example, the latency of wave V of the scalp-recorded auditory brainstem-evoked response, also generated in the IC, is shorter in hearing-impaired listeners who use amplification as compared with those who do not (Philibert, Collet, Vesson, & Veuillet, 2005). FFR encoding improves after auditory training in children with learning impairments (Russo, Nicol, Zecker, Hayes, & Kraus, 2005). Abnormal brainstem timing in learning disabilities is related to higher incidence of reduced cortical sensitivity to acoustic change and deficient literacy skills (Banai, Nicol, Zecker, & Kraus, 2005). FFR encoding of pitch patterns in Mandarin is more robust in monolingual, English-speaking musicians compared with non-musicians (Wong, Skoe, Russo, Dees, & Kraus, 2007).

Tone languages arguably provide an optimal window for examining how language experience influences the way pitch is processed in the brain. Earlier functional neuroimaging studies of lexically relevant pitch contrasts in Mandarin have consistently shown that cortical processing of lexical tones is language specific (Wong, Parsons, Martinez, & Diehl, 2004; Hsieh, Gandour, Wong, & Hutchins, 2001; Klein, Zatorre, Milner, & Zhao, 2001). Using tonal chimeras, a double dissociation was observed in the left planum temporale, demonstrating that stronger activity is elicited in response to native as compared with nonnative tones (Xu, Gandour, Talavage, et al., 2006). These brain imaging experiments all employed discrimination tasks and thus likely reflect temporally aggre-

gated neural events at a relatively late attention-modulated stage of auditory processing. However, at an early preattentive *subcortical* stage of processing, FFRs elicited in response to Mandarin tones reveal stronger pitch representation and smoother pitch tracking in native versus nonnative listeners (Krishnan, Xu, Gandour, & Cariani, 2005). This experience-dependent effect occurs only when the speech stimuli reflect *curvilinear* dynamic contours representative of Mandarin tones as opposed to *linear* dynamic approximations (Xu, Krishnan, & Gandour, 2006). In our taxonomy, the overall shape of  $f_0$  trajectories can be divided into static (steady state) and dynamic. Of the latter, they can be further subdivided into linear and nonlinear (curvilinear). The term *contour* is reserved for dynamic curvilinear  $f_0$  trajectories only. These FFR findings together 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, and thus blocking access to the mental lexicon. To date, all brain imaging and auditory electrophysiological studies of pitch patterns in tonal languages have utilized speech stimuli only.

To generate auditory stimuli that preserve the perception of pitch but do not have waveform periodicity or highly modulated stimulus envelopes that are characteristic of speech stimuli, we employ iterated rippled noise (IRN). Such stimuli preserve the temporal regularity of the stimulus without having to repeat the waveform in a periodic manner. They eliminate the lexical/semantic confound, thereby giving us an opportunity to investigate linguistic pitch at the brainstem level in a nonspeech context. An IRN stimulus is generated using a broadband noise that is delayed and added to itself repeatedly and therefore does not have a prominent modulated envelope (Patterson, Handel, Yost, & Datta, 1996; Yost, 1996a). However, by introducing temporal regularity into the fine structure of the noise, the delay-and-add process does change the envelope structure producing a "ripple" in the long-term power spectrum of the waveform. 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 psychophysical studies with humans, it is observed that pitch and pitch strength of IRN stimuli can be explained by temporal processing models based on autocorrelation analysis (Patterson et al., 1996; Yost, 1996a, 1996b). In brain imaging studies, IRN stimuli have been exploited to show that the processing of temporal pitch begins as early as the cochlear nucleus and continues up to auditory cortex (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Griffiths, Uppenkamp, Johnsrude, Josephs, & Patterson, 2001). Stimuli with dynamic spectral and temporal ripples, similar to IRN, have been employed to characterize the spectral and dynamic properties of auditory receptive fields (Chi, Ru,

& Shamma, 2005; Kowalski, Depireux, & Shamma, 1996) and to evaluate the auditory system's ability to extract perceptual information from a complex spectral profile (Yost & Moore, 1987). Most of the previous studies using IRN stimuli were limited to fixed delays that produce a constant pitch percept. The steady-state IRN algorithm was modified to allow the use of multiple delays to create IRN stimuli with linear time-variant pitch (Denham, 2005) and further to create IRN stimuli with curvilinear pitch contours that are ecologically representative of what occurs in natural speech (Swaminathan, Krishnan, & Gandour, 2008a). Using such IRN stimuli, it has been shown that experience-dependent neural mechanisms for pitch representation at the brainstem level are not specific to speech contexts but instead are sensitive to specific dimensions of pitch contours that native speakers of a tone language are familiar with (Swaminathan, Krishnan, & Gandour, 2008b). This enables us to investigate neural mechanisms underlying the processing of pitch contours that are of *behavioral relevance* in humans just as species-specific sounds are behaviorally relevant in other nonprimate and nonhuman primate animals (Suga, Ma, Gao, Sakai, & Chowdhury, 2003).

A complete understanding of the neural organization of language can only be achieved by viewing language processes as a set of hierarchical computations or mappings between representations at different stages of processing (Hickok & Poeppel, 2004). In speech perception, subcortical areas are not to be dismissed as "auditory areas" of no relevance to language processing. Rather, early stages of processing along the auditory pathway may perform computations that reflect experience-dependent sensitivity to specific features or dimensions that are linguistically relevant. To date, the bulk of research has been directed to the cerebral cortex, whereas the roles of subcortical structures in this hierarchical network for language processing have largely been neglected. 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 (Luo, Boemio, Gordon, & Poeppel, 2007; Bent, Bradlow, & Wright, 2006). To eliminate the task confound, we chose a passive listening paradigm to index pitch processing in the brainstem that involves no volitional memory or attention demands.

We hypothesized that pitch representation in the brainstem in response to IRN homologues of the four Mandarin tones, as reflected by pitch-tracking accuracy and pitch strength of FFRs, would be more robust in native speakers of Mandarin Chinese as compared with monolingual English speakers who had no prior knowledge of Mandarin or any other tone language. Although

differences in FFR responses may emerge from language experience, the effects of such experience are not specific to speech perception. By focusing on sections of a tonal contour instead of the whole tone, we were able to determine whether language-dependent effects are better conceptualized as applying across the board, that is, throughout the tonal contour, or alternatively, as applying to sections that exhibit certain acoustic dimensions irrespective of tonal category. Such experimental outcomes would 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.

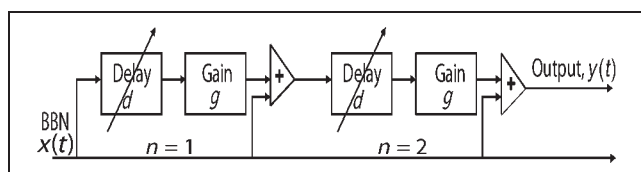
## METHODS

### Subjects

Twelve adult native speakers of Mandarin Chinese (5 men, 7 women) and 12 adult, monolingual native speakers of English (5 men, 7 women) participated in the FFR experiment. The two language groups were closely matched in age (Chinese:  $M = 27.6$ ,  $SD = 2.7$ ; English:  $M = 27.1$ ,  $SD = 3.2$ ) and years of formal education (Chinese:  $M = 18.8$ ,  $SD = 2.1$ ; English:  $M = 17.8$ ,  $SD = 2.3$ ) participated in the study. All Chinese subjects were born and raised in mainland China. None had received formal instruction in English before the age of 11. All Chinese subjects were classified as late onset Mandarin/English bilinguals who exhibited moderate proficiency in English, as determined by a language history questionnaire (Ping, Sepanski, & Zhao, 2006). They used their native language in the majority ( $M = 70\%$ ) of their daily activities combined. All subjects were nonmusicians, as determined by a music history questionnaire (Wong & Perrachione, 2007). None had more than 2 years of private lessons in any combination of instruments, and none had any musical training within the past 7 years. All subjects were strongly right-handed ( $\geq 96\%$ ) as measured by the Edinburgh Handedness Inventory (Oldfield, 1971). They had normal hearing sensitivity (i.e., pure-tone air conduction thresholds of 15 dB HL or better in both ears) at octave frequencies from 500 to 8000 Hz. All subjects were enrolled at Purdue University at the time of testing. They were paid for their participation. They gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

### Mandarin Tones

As a tone language, Mandarin is especially advantageous for investigating the processing of dynamic pitch movements using IRN stimuli. Its tonal inventory has four contrastive lexical tones ( $ma^1$  "mother" [T1],  $ma^2$  "hemp" [T2],  $ma^3$  "horse" [T3],  $ma^4$  "scold" [T4]). Tones 1 to 4 can be described phonetically as high level, high rising,



**Figure 1.** Block diagram to create dynamic IRN stimuli. The parameters involved in creating DIRN stimuli are delay  $d$  (in msec), gain  $g$  (set to 1), and number of iteration steps  $n$  (2 in this case). The arrow in the delay module suggests that each time sample of the original broad band noise (BBN) stimuli,  $x(t)$ , is delayed by a different amount and added back to itself. The DIRN stimulus allows for the use of parametrically defined time-varying delay functions that give rise to the perception of time-varying pitches.

low falling rising, and high falling, respectively (Howie, 1976). The dominant cue for tone recognition is  $f_0$  variation (Xu, 1997). All four tones exhibit  $f_0$  trajectories and harmonics that lie within the range of easily recordable FFRs (below about 2000 Hz).

### IRN Stimuli

IRN stimuli evoking time-varying pitch were generated from a noise waveform with Gaussian distribution of the instantaneous amplitudes. Instead of adding a copy of the original noise back at some fixed delay, a mapping function is created so that each point can be delayed by a different amount (Figure 1).

The signal can be described by Equation 1 (Swaminathan et al., 2008a):

$$y_i(t) = y_{i-1}(t) + g y_{i-1}(t - 1/f(t)); \text{ for } i = 1, 2, \dots, n \quad (1)$$

$y_0(t) = x(t)$ , the input signal,

where  $f(t)$  represents a polynomial equation of any degree modeling a linear/nonlinear  $f_0$  contour,  $n$  is the number of iteration steps, and  $g$  is the gain ( $-1 \leq g \leq 1$ ).

Mapping functions were chosen to be fourth-order polynomials representative of pitch contours occurring in natural speech productions of Mandarin Chinese (Xu, 1997). The gain was set to 1. The fourth-order polynomials of the IRN homologues of Mandarin tones are presented in Equation 2 (Figure 2):

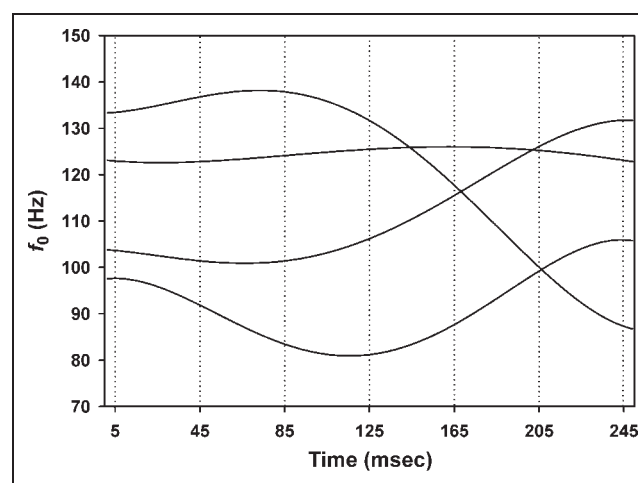
$$\begin{aligned} \text{Tone 1 : } f_1(t) &= 123.16 - (11.61/d)t + (67.94/d^2)t^2 \\ &\quad - (85.18/d^3)t^3 + (28.43/d^4)t^4 \\ \text{Tone 2 : } f_2(t) &= 103.85 - (8.45/d)t - (76.32/d^2)t^2 \\ &\quad + (297.91/d^3)t^3 - (185.34/d^4)t^4 \\ \text{Tone 3 : } f_3(t) &= 97.45 + (15.05/d)t - (382.13/d^2)t^2 \\ &\quad + (772.58/d^3)t^3 - (397.29/d^4)t^4 \\ \text{Tone 4 : } f_4(t) &= 133.26 + (4.36/d)t + (165.15/d^2)t^2 \\ &\quad - (504.84/d^3)t^3 + (288.71/d^4)t^4, \end{aligned} \quad (2)$$

where  $d$  is the duration of the stimuli. The duration of all four stimuli was fixed at 250 msec with 10-msec rise/fall time.

### Data Acquisition

Subjects reclined comfortably in an acoustically and electrically shielded booth. They were instructed to relax and refrain from extraneous body movements to minimize movement artifacts. Most subjects fell asleep during the data acquisition session. All stimuli were presented monaurally at 60 dB nHL at a repetition rate of 3.13/sec. The order of stimuli was randomized across subjects. All stimuli were controlled by a signal generation and data acquisition system (Tucker-Davis Technologies, System II). The stimulus files were routed through a digital to analog module and presented through magnetically shielded insert earphones (Biologic, TIP-300).

For all four test conditions, FFRs were recorded from each subject to monaural stimulation of the right ear. These evoked responses were recorded differentially between scalp electrodes placed on the midline of the forehead at the hairline and the ipsilateral mastoid. Another electrode placed on the contralateral mastoid served as the common ground. The interelectrode impedances were maintained below 3000  $\Omega$ . The EEG inputs were amplified by 200,000 and band-pass filtered from 60 to 3000 Hz (6 dB/octave roll-off, RC response characteristics). Each response waveform represents an average of 2000 stimulus presentations over a 300-ms analysis window using a sampling rate of 25 kHz.



**Figure 2.** IRN homologues modeled after average  $f_0$  contours of time-normalized (250 msec) Mandarin lexical tones (Xu, 1997). Numbers (1–4) denote each of the four Mandarin tonal categories (T1, T2, T3, and T4). Vertical dotted lines demarcate six 40-msec sections within each  $f_0$  contour: 5–45, 45–85, 85–125, 125–165, 165–205, and 205–245 msec.

## Data Analysis

### *Temporal and Spectral Analysis*

Short-term autocorrelation functions and running autocorrelograms (ACGs) were computed for the grand averaged FFRs to index variation in FFR periodicities over the duration of the response. The ACG represents the short-term autocorrelation function of windowed frames of a compound signal, that is,  $ACG(\tau, t) = X(t) \times X(t - \tau)$  for each time  $t$  and time lag  $\tau$ . It is a three-dimensional plot quantifying periodicity and pitch strength variations over time. The horizontal axis represents time, the vertical axis represents the time lags associated with the peaks of the autocorrelation function, and the intensity of each point in the image represents the amplitude of the autocorrelation function associated with a particular lag at a particular time. Thus, it represents the running distribution of all-order intervals present in the population response (Swaminathan et al., 2008a; Cariani & Delgutte, 1996a). Narrow-band spectrograms were obtained from each FFR waveform to evaluate the spectral composition.

### *Pitch-tracking Accuracy of Whole Tones*

The ability of the FFR to follow pitch change in the stimuli was evaluated by extracting the  $f_0$  contour from the grand averaged FFRs using a periodicity detection short-term autocorrelation algorithm (Boersma, 1993). Essentially, this algorithm performs a short-term autocorrelation analysis on a number of small segments (40-msec frames) taken from the signal (stimuli and FFR). The analysis window was shifted incrementally in 10-msec steps. The autocorrelation function was computed for each 40-msec frame after successive shifts. The time lag corresponding to the maximum autocorrelation value for each frame was recorded for both the stimulus and the FFR. Candidates below 75 Hz and above 200 Hz were not retained. The time lags associated with autocorrelation peaks in each frame were concatenated together to give a running  $f_0$  contour. The cross-correlation coefficient between the  $f_0$  contour extracted from the FFRs and the stimuli provides a measure of pitch-tracking accuracy.

### *Pitch Strength of Tonal Sections*

To compute the pitch strength of the FFR responses to time-varying IRN stimuli, FFR responses to IRN homologues of the four Mandarin tones were divided into six nonoverlapping 40-msec time frames (5–45, 45–85, 85–125, 125–165, 165–205, and 205–245 msec). The normalized autocorrelation function of the two language groups was derived from an analysis of each corresponding time frame of the IRN stimulus and the FFR response. Within each frame, the autocorrelation was normalized between 0 and 1 with higher values indicat-

ing more periodic temporal activity. The second author first identified visually the time of lag associated with the location of the autocorrelation peak per 40-msec frame from the input IRN stimuli. The time lag in the stimulus autocorrelation relates to the mean pitch for that section. The location of this time of lag information was then used to guide a visual search for the corresponding autocorrelation peak in the FFR response. Within each 40-msec frame, the response peak selected was the one that was closest to the location of the autocorrelation peak in the input stimulus. This response peak was taken to be an estimate of pitch strength per time frame. All waveforms were remeasured independently by a graduate student in electrical engineering. Interjudge measurement reliability was high (Pearson  $r = .99$ ).

## Statistical Analysis

### *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 the IRN stimuli. A repeated measures, mixed model ANOVA (SAS)—with group (Chinese, English) as the between-subject factor, tone (T1, T2, T3, T4) as the within-subject factor, 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 such degraded *nonspeech* stimuli.

### *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 four IRN homologues of Mandarin 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, English) and with two fixed within-subject factors: tone (T1, T2, T3, and T4) and section (5–45, 45–85, 85–125, 125–165, 165–205, and 205–245 msec). The section factor was nested within tone. By focusing on 40-msec 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*.

### *Pitch Strength of Tonal Dimensions*

For this purpose, the four IRN homologues of Mandarin tones were concatenated without any consideration for their linguistic significance. We then identified three 40-msec  $f_0$  frames at large that were maximally differentiated based on acceleration ( $\alpha = dv/dt$ : maximum +  $\alpha$ ,

maximum  $-\alpha$ , minimum  $\alpha$ ) by sliding a 40-msec window in 10-msec increments across the entire duration of the four IRN stimuli. The  $f_0$  frame labeled *flat*, the entry with the lowest acceleration coefficient, was extracted from the 45- to 85-msec time frame of T2 ( $\alpha = .0001$ ; Table 2, S2, T2); *falling*, the entry with highest negative acceleration coefficient, was extracted from the 165- to 205-msec time frame of T4 ( $\alpha = -.0181$ ; Table 2, S5, T4); and *rising*, the entry with highest positive acceleration coefficient, was extracted from the 165- to 205-msec time frame of T3 ( $\alpha = .0118$ ; Table 2, S5, T3). Pitch strength was calculated from the FFRs of every subject for these three  $f_0$  frames.

A discriminant analysis was conducted to determine the weighted linear combination of these three dimensions that best discriminate between the two language groups (Chinese, English). The discriminant function was cross-validated using a  $k$ -fold cross-validation, that is, pitch strength weights from each subject were used

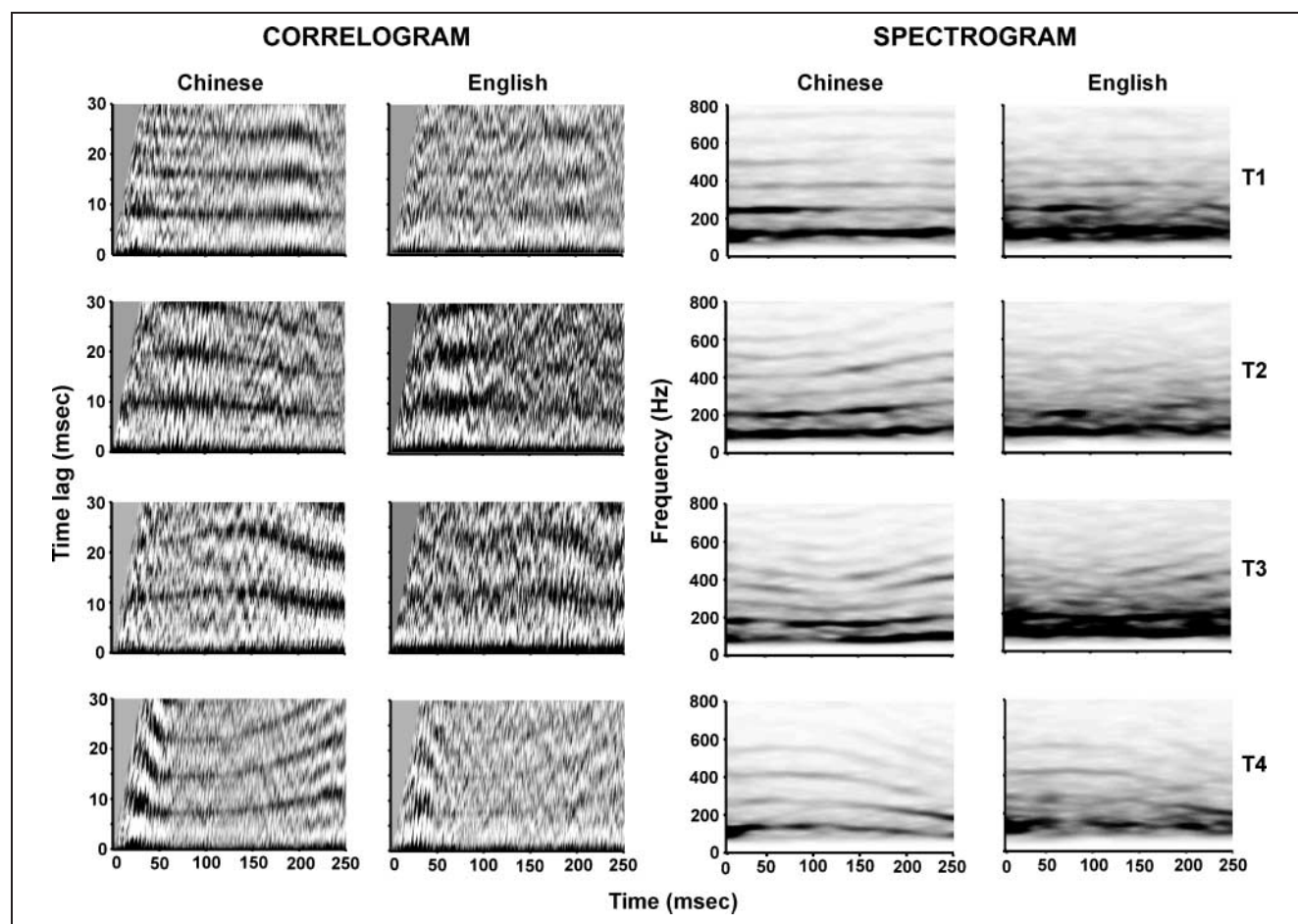
as validation data for the function created from weights of the remaining subjects. This was repeated until every one of the 22 subjects was used for validation. Finally, comparisons of the mean subject weights per language group on each of the three dimensions were performed by means of a one-way ANOVA.

## RESULTS

### Temporal and Spectral Properties of Whole Tones

ACGs (left panels) and narrow band spectrograms (right panels) derived from the grand average FFR waveforms in response to the IRN homologues of Mandarin tones are shown in Figure 3 for the Chinese and English groups.

In the Chinese group, the ACGs show clear dark bands of phase-locked activity at the fundamental period ( $f_0$ ) and its multiples throughout the entire duration of the



**Figure 3.** Correlograms (columns 1 and 2) and spectrograms (columns 3 and 4) derived from grand averaged FFR waveforms of Chinese and English groups in response to the four IRN homologues of Mandarin tones (T1, row 1; T2, row 2; T3, row 3; and T4, row 4). Across all four tones, correlograms of the Chinese group (column 1) show clearer bands (black) of temporal regularity in the phase-locked activity in the FFR response at the fundamental period ( $1/f_0$ ) and its multiples as compared with that of the English group (column 2). Similarly, the spectrograms of the Chinese group (column 3) show a markedly improved (black) spectral band at the  $f_0$  and its harmonics as compared with the English group (column 4) for all four tones.

FFR response (column 1). In contrast, the bands are less distinct and more diffuse in the English group (column 2), indicating that time-varying temporal correlations in falling/rising  $f_0$  movements of dynamic IRN stimuli are better represented in the Chinese group. Although phase-locked activity to the fundamental period is present regardless of language experience, the bands of phase-locked intervals for the Chinese group are narrower than those for the English group. These narrower bands suggest that phase-locked activity for Chinese listeners is, not only more robust, but also more accurate than that of English listeners.

The FFR spectrograms reveal energy bands corresponding to  $f_0$  and its harmonics for both the Chinese (column 3) and English (column 4) groups. We observe, however, that spectral variations are strongly represented up to the fifth harmonic in the Chinese group, but only up to the third or fourth harmonic in the English group.

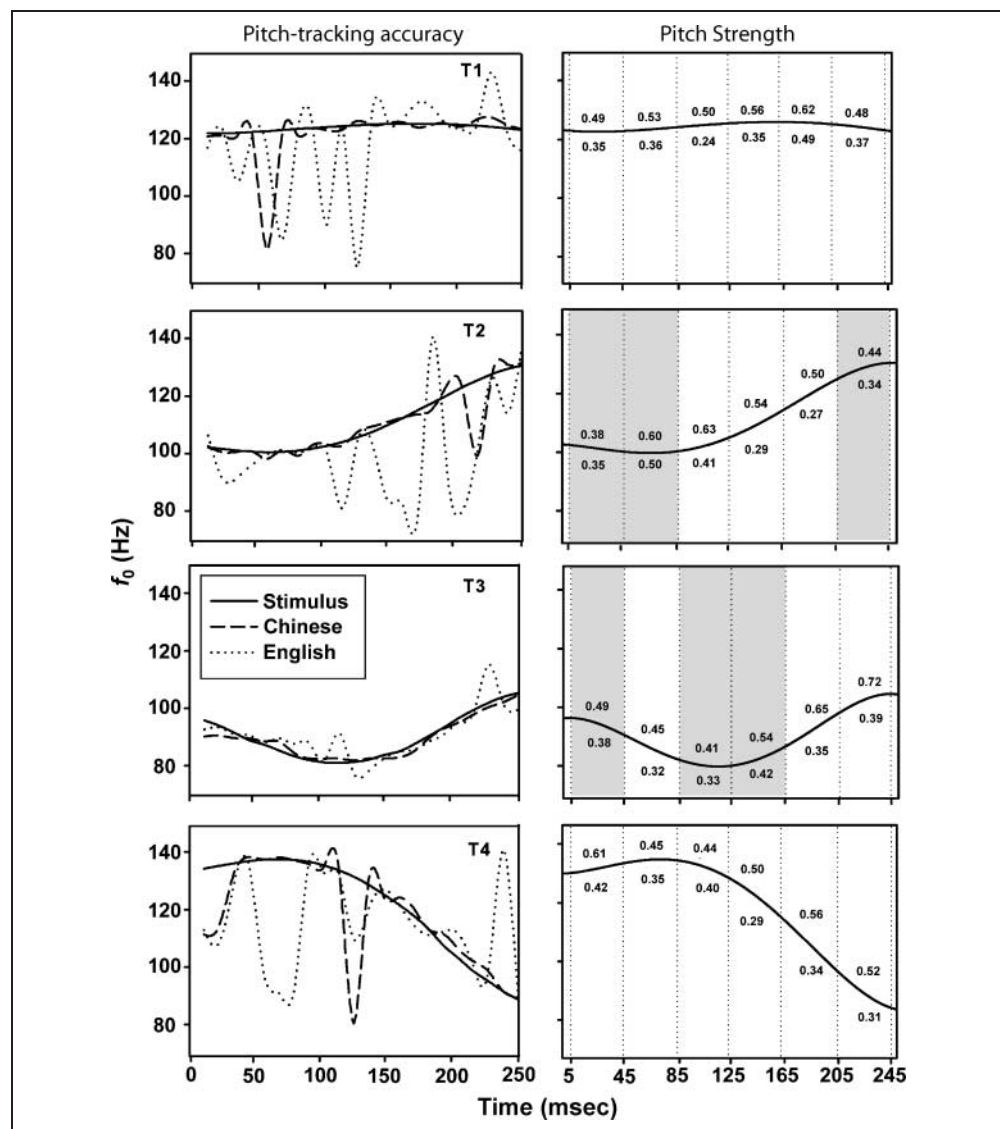
The better representation of multiple pitch-relevant harmonics in the Chinese group points to more accurate pitch tracking as well as stronger pitch representations.

### Pitch-tracking Accuracy of Whole 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 four IRN homologues of Mandarin tones (Figure 4, left panels). Overall, pitch tracking is observed to be less variable for the Chinese group (dashed line) compared with 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.

Table 1 presents mean cross-correlation coefficients of the Chinese and English groups for IRN homologues of

**Figure 4.** Pitch-tracking accuracy of IRN homologues 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 by 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 with 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-msec sections within each  $f_0$  contour: 5–45, 45–85, 85–125, 125–165, 165–205, and 205–245 msec. Sections that yielded significantly larger pitch strength for the Chinese group relative to English are unshaded; those that did not are shaded in gray (cf. Table 2).



**Table 1.** Cross-correlation Coefficient Values of Pitch-tracking Accuracy per Language Group and IRN Homologues of Mandarin Tones

| Tone            | Group       |             |
|-----------------|-------------|-------------|
|                 | Chinese     | English     |
| T1              | 0.34 (0.04) | 0.27 (0.05) |
| T2 <sup>a</sup> | 0.60 (0.05) | 0.23 (0.04) |
| T3 <sup>a</sup> | 0.52 (0.07) | 0.31 (0.09) |
| T4 <sup>a</sup> | 0.62 (0.06) | 0.25 (0.04) |

Values are expressed as  $M$  ( $SD$ ).

<sup>a</sup>Statistically significant difference between language groups in pitch-tracking accuracy.

Mandarin tones. An omnibus ANOVA on cross-correlation coefficients yielded a significant main effect of group,  $F(1, 22) = 25.92, p < .0001, \eta^2_{\text{partial}} = .541$ ; cf. tone,  $F(3, 66) = 2.67, p = .0545$ . The Group  $\times$  Tone interaction effect was also significant,  $F(3, 66) = 3.93, p = .0121, \eta^2_{\text{partial}} = .151$ . Only the Chinese group showed significant differences in cross-correlation coefficients among the four tones,  $F(3, 66) = 6.20, p = .0009$ ; cf. English,  $F(3, 66) = 0.40, p = .7520$ . Post hoc Tukey–Kramer-adjusted comparisons ( $\alpha = .05$ ) revealed that the cross-correlation coefficient of T1 was significantly smaller than either T2 or T4. Cross-language comparisons showed that cross-correlation coefficients were significantly larger in the Chinese than the English group for T2,  $F(1, 66) = 21.30, p < .0001$ , T3,  $F(1, 66) = 6.76, p < .0115$ , and T4,  $F(1, 66) = 20.76, p < .0001$ , but not T1,  $F(1, 66) = 0.74, p = .3928$ .

### Pitch Strength of Tonal Sections

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 homologues of Mandarin tones (Figure 4, right panels). Pitch strength values for the Chinese and the English groups, respectively, appear above and below the  $f_0$  contour.

For all four tones, omnibus ANOVAs on FFR pitch strength revealed significant main effects of Group [T1,  $F(1, 22) = 22.68, p < .0001, \eta^2_{\text{partial}} = .517$ ; T2,  $F(1, 22) = 12.59, p = .0018, \eta^2_{\text{partial}} = .364$ ; T3,  $F(1, 22) = 25.70, p < .0001, \eta^2_{\text{partial}} = .539$ ; T4,  $F(1, 22) = 23.58, p < .0001, \eta^2_{\text{partial}} = .517$ ] and Section [T1,  $F(5, 110) = 4.15, p = .0017, \eta^2_{\text{partial}} = .159$ ; T2,  $F(5, 110) = 8.13, p < .0001, \eta^2_{\text{partial}} = .270$ ; T3,  $F(5, 110) = 6.27, p < .0001, \eta^2_{\text{partial}} = .222$ ; T4,  $F(5, 110) = 2.71, p = .0239, \eta^2_{\text{partial}} = .110$ ]. The Group  $\times$  Section interaction effect was significant for T2,  $F(5, 110) = 2.62, p = .0279, \eta^2_{\text{partial}} = .106$ , and T3,  $F(5, 110) = 3.81, p = .0032, \eta^2_{\text{partial}} = .148$ , but not T1 or T4. In T2 and T3, post hoc Tukey–Kramer-adjusted multiple

comparisons ( $\alpha = .05$ ) showed that pitch strength of the Chinese group was greater than the English group in 3 of 6 tonal sections (Figure 4, right panels, unshaded). In T1 and T4, pitch strength of the Chinese group was greater than the English group in all tonal sections. For each language group, pitch strength was found to vary across sections of T2. In T3, however, pitch strength varied across sections for the Chinese group only.

Table 2 presents the acceleration and deceleration values of the six sections of each of the four IRN homologues of Mandarin tones. Pooling across tones, a positive correlation was observed between the pitch strength ratios of the two language groups (Chinese/English) and acceleration (absolute) values of the input IRN stimuli per section (Pearson  $r = .45, p = .0270$ ). Within each tone (cf. Table 2; Figure 4, right panels), we chose three maximally accelerating or decelerating sections as a measure of the extent to which rapidly changing pitch may account for language-dependent effects *relative to the tone itself*. The first section was conservatively omitted in recognition of the possibility of onset artifacts. Of these 12 sections, 92% (11 of 12) yielded a significant difference in pitch strength between the native and nonnative language groups.

### Pitch Strength of Tonal Dimensions

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-msec temporal intervals that were maximally differentiated in terms of slope (flat, rising, and falling). The classification matrix from this discriminant analysis is presented in Table 3. Overall, about 83% of subjects were correctly classified into their respective language groups (Chinese, 83.33%; English, 83.33%). Because we can expect to get only 50%

**Table 2.** Acceleration Values of the Six Sections from Each of the Four IRN Homologues of Mandarin Tones

| Tone | Section |         |         |         |         |         |
|------|---------|---------|---------|---------|---------|---------|
|      | S1      | S2      | S3      | S4      | S5      | S6      |
| T1   | −0.0002 | 0.0013  | 0.0014  | 0.0005  | −0.0008 | −0.0022 |
| T2   | −0.0023 | 0.0001  | 0.0049  | 0.0095  | 0.0108  | 0.0058  |
| T3   | −0.0059 | −0.0086 | −0.0023 | 0.0066  | 0.0118  | 0.0068  |
| T4   | 0.0034  | 0.0011  | −0.0063 | −0.0143 | −0.0181 | −0.0131 |

S1, S2, S3, S4, S5, and S6 represent the six 40-msec sections within each  $f_0$  contour: 5–45, 45–85, 85–125, 125–165, 165–205, and 205–245 msec. T1, T2, T3, and T4 stand for the four Mandarin tones. Values represent the degree of acceleration/deceleration within each section. For a 40-msec time frame, acceleration was computed as the difference in pitch values at the offset and onset divided by the duration of the frame. Positive and negative signs represent rising and falling  $f_0$  trajectories, respectively. Values shaded in gray represent the three maximally accelerating or decelerating sections, excluding S1, within each tone.



**Table 3.** Classification Matrix for Two-group Discriminant Analysis as a Function of Pitch Strength of Three Maximally Differentiated Slopes as Defined by Acceleration

| Actual Group | n  | Predicted Group  |            |
|--------------|----|------------------|------------|
|              |    | Chinese          | English    |
| Chinese      | 12 | 10 (83.33)       | 2 (16.67)  |
| English      | 12 | 2 (16.67)        | 10 (83.33) |
| Total        | 24 | 0.5 <sup>a</sup> | 0.5        |

Values in parentheses are expressed in percentages. Numbers on the diagonal represent correct classifications; off-diagonal numbers represent incorrect classifications.

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

of the classifications correct by chance, an overall 83% accuracy rate represents a considerable improvement (canonical correlation = .70). Only 4.2% fewer correct classifications (Chinese, 9/12; English, 10/12) were made in the cross-validated analysis in comparison to the original analysis. The group centroids, that is, average discriminant  $z$  scores, were significantly different between the Chinese (0.95) and the English ( $-0.95$ ) groups ( $F(3, 20) = 6.56, p = .0029$ ). The pooled within-class standardized canonical coefficients for flat, rising, and falling, respectively, were .0473, .7064, and .4280, indicating that of these three variables, pitch strength of the rising  $f_0$  trajectory was the most important in discriminating listeners by language affiliation. The importance of dynamic changes in pitch in discriminating the two language groups is consistent with individual subjects' pitch strength loadings on these three  $f_0$  trajectories (Figure 5). Univariate tests of pitch strength of these flat (relatively), rising, and falling  $f_0$  trajectories further confirmed that more dynamic changes in pitch had a greater influence on the FFR responses of the Chinese group relative to the English [rising:  $F(1, 22) = 18.79, p = .0003, \eta^2_{\text{partial}} = .461$ ; falling:  $F(1, 22) = 14.89, p = .0009, \eta^2_{\text{partial}} = .404$ ], whereas less dynamic changes in pitch did not yield a language group effect,  $F(1, 22) = 1.60, p = .2186$ .

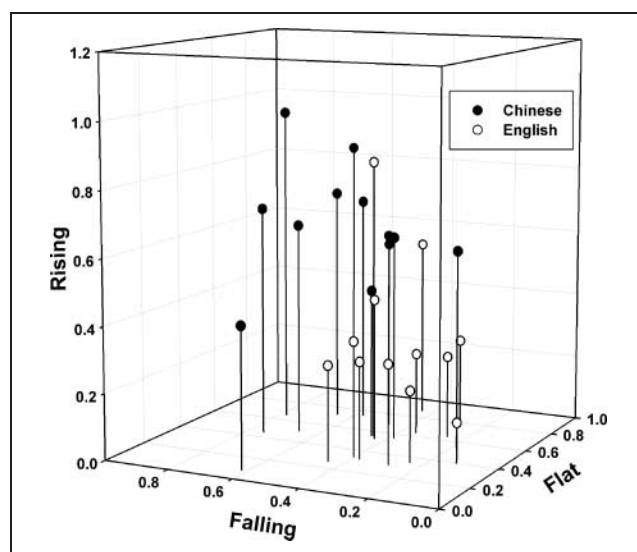
## DISCUSSION

### Experience-dependent Plasticity of Brainstem Mechanisms Underlying FFR Pitch Extraction

Using IRN stimuli, the results of this study demonstrate that experience-dependent brainstem reorganization for pitch representation in the Chinese group is dependent upon specific *dimensions* of pitch contours that commonly occur in a native listener's language experience. This finding suggests that brainstem neurons are differentially sensitive to changes in pitch without regard to the context in which they are presented. As regards to sustained phase-locked activity in the brainstem, we infer that cross-language differences reflect 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, we adopt the temporal correlation analysis model described by Langner (1992, 1997). 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-msec 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,



**Figure 5.** Pitch strength values projected in the three dimensions (flat, falling, and rising) derived from individual Chinese and English subjects' FFR waveforms in response to three 40-msec temporal intervals at large characterized by maximally differentiated slopes. Chinese individuals tend to cluster in the upper (rising) left (falling) quadrant of the 3-D space, indicating relatively higher pitch strength of dynamic pitch movements regardless of pitch direction, as compared with the English individuals. Most individuals, Chinese and English alike, tend to cluster toward the back (flat) quadrant, indicating that slowly changing pitch movements are less effective in discriminating between speakers of tone and nontone languages.

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. It is also possible that corticofugal mechanisms facilitate the experience-dependent reorganization for pitch in the brainstem.

We further 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 (Figure 3). This finding complements our data on voice pitch representation (Krishnan et al., 2004). In the Chinese group, stronger pitch and more accurate pitch tracking cooccur 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 (second to about the fifth harmonic) for pitch (Schwartz & Purves, 2004; Cariani & Delgutte, 1996a).

### Do Pitch Representations in the Brainstem Reflect Speech, Language, Tone, and/or Something Else?

Our findings force us to reconceptualize how linguistically relevant pitch patterns are processed at the brainstem level. We argue that language, speech, and tone, albeit relevant to the observed effects, are epiphenomenal. Instead, we infer that Chinese listeners' native language experience has changed the way they process linguistically relevant pitch patterns regardless of the stimulus context (nonspeech) in which these patterns are embedded. Although the basis for cross-language differences in FFR pitch extraction may emerge from language experience (Mandarin), the effects of such experience are not specific to speech perception (cf. Xu, Gandour, & Francis, 2006, categorical perception of pitch direction), nor is pitch extraction at the brainstem level necessarily specific to the domain in which pitch patterns occur.

Enhanced FFR pitch extraction of lexical tones may be induced by long-term exposure to musical pitch patterns as well as native language experience (Wong et al., 2007). In musicians, brainstem responses are enhanced in response to both speech and music stimuli (Musacchia, Sams, Skoe, & Kraus, 2007). Taken together, these findings suggest that experience-dependent enhancement of pitch representation in the brainstem has more to do with similarities in pitch patterns that listeners are exposed to than with the context (speech, nonspeech) or domain (music, language) in which these pitch patterns occur.

Of the four tonal categories (T1, T2, T3, and T4), pitch-tracking accuracy of the whole contour is more

accurate in the native than the nonnative group for all but T1 (Figure 4, left panels; Table 1). Wong et al. (2007) similarly showed that English-speaking musicians exhibit more robust pitch tracking of T2 and T3, but not T1, than nonmusicians. In the native Chinese group, pitch tracking of T1 is less accurate than T2 or T4. We therefore attribute the absence of a language group effect in response to T1 to diminished accuracy in the Chinese group (Table 1). This finding may be due to the fact that the stimuli used in this study were citation forms, that is, tones produced on isolated monosyllables. In connected speech, T1 actually exhibits considerably more dynamic pitch than observed in citation forms (Xu, 1997, 2006).

Pitch tracking alone is insufficient because of its restriction to the *whole* rather than to the parts of a pitch contour. Although the two groups do not differ in pitch tracking of T1 (Figure 4, left panel), they do differ in pitch strength across all six sections of T1 (Figure 4, right panel). Moreover, we can now see that between-group differences in pitch strength are not necessarily uniform throughout the duration of IRN homologues of the four Mandarin tones (Figure 4, right panels). In T2 and T3, only three of six sections yield a group effect. This finding suggests 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.

These findings lead us to the next logical question: 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 (Figure 4, right panels; Table 2). 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 at the brainstem only when stimuli reflect *curvilinear* dynamic contours representative of Mandarin tones as opposed to *linear* dynamic approximations (Xu, Krishnan, et al., 2006). In the Xu, Krishnan, et al. (2006) study, linear rising or falling ramps, crude approximations of T2 and T4, respectively, elicited homogeneous FFR pitch strength and pitch-tracking accuracy values in native (Chinese) and nonnative (English) listeners. 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. Although they are dynamic, acceleration (T2) or deceleration (T4) is a constant. As such, they are not ecologically representative of what native listeners are familiar with. Thus, pitch extraction at the brainstem level is critically dependent on *specific dimensions* of pitch contours that native speakers have been exposed to in natural speech contexts. The logical next step in this line of research requires a direct comparison of Chinese and English listeners' FFR responses to *both* linear and curvilinear IRN pitch patterns, which is already underway in our laboratory.

The discriminant analysis of the three 40-msec  $f_0$  frames at large (flat, rising, and falling) that were maximally differentiated based on acceleration suggests that sensitivity of the FFR to rising  $f_0$  trajectories is heavily weighted in separating listeners based on language experience (Figure 5). Both psychoacoustic (Schouten, 1985; Collins & Cullen, 1978; Nabelek, 1978; Klatt, 1973) and physiological studies (Krishnan & Parkinson, 2000; Shore, Clopton, & Au, 1987; Shore & Nuttall, 1985) indicate better sensitivity for rising versus falling tones. Multidimensional scaling analyses show that the perceptual dimension related to direction of pitch change is spatially distributed primarily in terms of rising versus 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 toward 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. Consistent with this notion is relatively more robust FFR representation of periodicities corresponding to harmonics in the dominant region for pitch in the stimuli with rising versus falling contours (see spectrograms in right panels of Figure 3). This observation suggests that the greater pitch salience for the rising stimulus may be due to its more selective activation of frequencies in the dominant pitch region (Plomp, 1967; Ritsma, 1967; Flanagan & Guttman, 1960a, 1960b).

We further observe that the pitch movement is rising in 11 of 18 tonal sections in which pitch strength is larger in the Chinese than the English group (Figure 4, right panels). In only 2 of 13 rising sections do we fail to see a language-dependent advantage. Despite relatively shallow rising pitch movement across three sections (S2–S4) of T1, the Chinese group maintains its advantage over the English group in terms of pitch strength. Although the two groups are found to differ in pitch strength when pitch movement is falling, the slopes generally must be steeper than rising to induce

language-dependent effects (cf. S3–S6 of T4 with S1 of T2 and T3). Thus, we hypothesize that innate differential sensitivity to rising and falling pitch movements in the auditory system is enhanced at the level of the human brainstem depending upon the prosodic needs of a particular language.

Because pitch is a *multidimensional* (e.g., height, direction, magnitude of change) perceptual attribute, we have been able to distinguish between the effects of categories versus dimensions on pitch representations in the brainstem. We infer that brainstem neural activity underlying the generation of the FFR is tuned to optimally shape specific, linguistically relevant dimensions of the signal that are fed forward to the auditory cortex. The role of the brainstem is to shape those dimensions that are to be subsequently processed separately at early cortical stages of auditory processing. How information about pitch dimensions extracted from the FFR is integrated into the auditory stream at higher stages of the auditory pathway is a topic for future research.

### Using Dynamic Curvilinear IRN as a Window on Pitch Processing in the Brain

Any periodic or quasi-periodic stimuli like speech have a highly modulated envelope structure. Most behavioral, physiological, and brain imaging studies have used non-speech stimuli that retain the envelope structure of the original unprocessed speech: for example, FM sweeps (Luo et al., 2007), rotated speech (Scott, Blank, Rosen, & Wise, 2000), and sine wave speech (Liebenthal, Binder, Piorkowski, & Remez, 2003). IRN stimuli, however, do not have waveform periodicity or highly modulated stimulus envelopes characteristic of speech stimuli, thus eliminating any potential lexical bias for native listeners. Herein, dynamic curvilinear IRN stimuli exhibit pitch contours ecologically representative of what occurs in natural speech. These stimuli are of theoretical interest because they allow us to investigate sensitivity to speech-like pitch contours parametrically without confounding psycholinguistic variables. In view of the importance of prosody in speech communication, they promise to be a useful tool for assessing the efficacy of different signal processing strategies for cochlear implants.

Although the Mandarin pitch contours are presented in a nonspeech context, it does not necessarily follow that they are nonlinguistic pitch shifts. To the contrary, all four IRN stimuli are homologues of phonetic exemplars of Mandarin tones. Experience-dependent reorganization of pitch in the brainstem of native listeners may be triggered by linguistically relevant pitch contours. IRN stimuli simulate signal degradation in the noisy environment of everyday life. Its robust representation of pitch suggests neural mechanisms are resistant to signal degradation to extract behaviorally relevant dimensions. To the extent this reorganization cuts across domains, it might facilitate robust representation of other stimuli

(e.g., music) with dynamic pitch similar to the pitch contours experienced by the native listeners.

## Conclusion

Our findings demonstrate 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. 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. The dynamic IRN stimuli used herein give us a tool to investigate neural mechanisms underlying pitch patterns representative of those that occur in natural speech without a semantic confound.

## Acknowledgments

This research was supported in part by the National Institutes of Health R01 DC008549-01 (A. K.) and the College of Liberal Arts (A. K., J. G.).

Reprint requests should be sent to Ananthanarayan Krishnan, Department of Speech Language Hearing Sciences, Purdue University, 1353 Heavilon Hall, West Lafayette, IN 47907-2038, or via e-mail: rkrish@purdue.edu.

## REFERENCES

- 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*, 223–226.
- Anathanarayan, A. K., & Gerken, G. M. (1987). Response enhancement and reduction of the auditory brainstem response in a forward-masking paradigm. *Electroencephalography and Clinical Neurophysiology*, *66*, 427–439.
- Banai, K., Nicol, T., Zecker, S. G., & Kraus, N. (2005). Brainstem timing: Implications for cortical processing and literacy. *Journal of Neuroscience*, *25*, 9850–9857.
- 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*, 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.
- Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, *76*, 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*, 1717–1734.
- Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *Journal of the Acoustical Society of America*, *118*, 887–906.
- Cohen, M. A., Grossberg, S., & Wyse, L. L. (1995). A spectral network model of pitch perception. *Journal of the Acoustical Society of America*, *98*, 862–879.
- Collins, M. J., & Cullen, J. K., Jr. (1978). Temporal integration of tone glides. *Journal of the Acoustical Society of America*, *63*, 469–473.
- de Cheveigne, A. (1998). Cancellation model of pitch perception. *Journal of the Acoustical Society of America*, *103*, 1261–1271.
- Denham, S. (2005). Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model. *Biosystems*, *79*, 199–206.
- Eady, S. J. (1982). Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech*, *25*, 29–42.
- Flanagan, J. L., & Guttman, N. (1960a). On the pitch of periodic pulses. *Journal of the Acoustical Society of America*, *32*, 1308–1319.
- Flanagan, J. L., & Guttman, N. (1960b). Pitch of periodic pulses without fundamental component. *Journal of the Acoustical Society of America*, *32*, 1319–1328.
- Galbraith, G. C., Bagasan, B., & Sulahian, J. (2001). Brainstem frequency-following response recorded from one vertical and three horizontal electrode derivations. *Perceptual and Motor Skills*, *92*, 99–106.
- Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–175.
- Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*, 1–33.
- Glaser, E. M., Suter, C. M., Dasheiff, R., & Goldberg, A. (1976). The human frequency-following response: Its behavior during continuous tone and tone burst stimulation. *Electroencephalography and Clinical Neurophysiology*, *40*, 25–32.
- Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America*, *54*, 1496–1516.
- 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*, 91–114.
- 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*, 633–637.
- Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.
- Howie, J. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language*, *76*, 227–252.
- 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*, 8–16.
- Klein, D., Zatorre, R., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *Neuroimage*, *13*, 646–653.
- Kowalski, N., Depireux, D. A., & Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex: II. Prediction of unit responses to arbitrary dynamic spectra. *Journal of Neurophysiology*, *76*, 3524–3534.

- Krishnan, A. (2002). Human frequency-following responses: Representation of steady-state synthetic vowels. *Hearing Research, 166*, 192–201.
- Krishnan, A., & Parkinson, J. (2000). Human frequency-following response: Representation of tonal sweeps. *Audiology and Neuro-otology, 5*, 312–321.
- Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*, 161–168.
- 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*, 1–12.
- Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research, 60*, 115–142.
- Langner, G. (1997). Neural processing and representation of periodicity pitch. *Acta Oto-laryngologica, Supplement, 532*, 68–76.
- Liebenthal, E., Binder, J. R., Piorkowski, R. L., & Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *Journal of Cognitive Neuroscience, 15*, 549–558.
- Luo, H., Boemio, A., Gordon, M., & Poeppel, D. (2007). The perception of FM sweeps by Chinese and English listeners. *Hearing Research, 224*, 75–83.
- Marsh, J. T., Brown, W. S., & Smith, J. C. (1974). Differential brainstem pathways for the conduction of auditory frequency-following responses. *Electroencephalography and Clinical Neurophysiology, 36*, 415–424.
- Meddis, R., & Hewitt, M. J. (1991). Virtual pitch and phase-sensitivity studies using a computer model of auditory periphery: I. Pitch identification. *Journal of the Acoustical Society of America, 89*, 2866–2882.
- Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America, 102*, 1811–1820.
- Moore, B. C. (1989). *Introduction to the psychology of hearing* (3rd ed.). London: Academic Press.
- 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, U.S.A., 104*, 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*, 751–763.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.
- 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 Acoustical Society of America, 100*, 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*, 767–776.
- Philibert, B., Collet, L., Vesson, J. F., & Veillet, E. (2005). The auditory acclimatization effect in sensorineural hearing-impaired listeners: Evidence for functional plasticity. *Hearing Research, 205*, 131–142.
- Ping, L., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavior Research Methods, 38*, 202–210.
- Plack, C. J., Oxenham, A. J., & Fay, R. R. (Eds.) (2005). *Pitch: Neural coding and perception* (Vol. 24). New York: Springer.
- Plomp, R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America, 41*, 1526–1533.
- 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*, 523–533.
- Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *Journal of the Acoustical Society of America, 42*, 191–198.
- 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*, 95–103.
- Schouten, M. E. (1985). Identification and discrimination of sweep tones. *Perception and Psychophysics, 37*, 369–376.
- Schwartz, D. A., & Purves, D. (2004). Pitch is determined by naturally occurring periodic sounds. *Hearing Research, 194*, 31–46.
- Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*, 2400–2406.
- Shofner, W. P. (1999). Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms. *Journal of Neurophysiology, 81*, 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*, 471–478.
- 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*, 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*, 465–472.
- Sohmer, H., & Pratt, H. (1977). Identification and separation of acoustic frequency following responses (FFRs) in man. *Electroencephalography and Clinical Neurophysiology, 42*, 493–500.
- Suga, N. (1990). Biosonar and neural computation in bats. *Scientific American, 262*, 60–68.
- 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, U.S.A., 97*, 11807–11814.
- 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*, 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*, 281–287.
- Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008b). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport, 19*, 1163–1167.
- 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*, 9153–9160.
- Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28*, 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*, 420–422.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.
- Xu, Y. (2006). Tone in connected discourse. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed., Vol. 12, pp. 742–750). Oxford, UK: Elsevier.
- Xu, Y., Gandour, J., & 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*, 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*, 173–183.
- Xu, Y., Krishnan, A., & Gandour, J. T. (2006). Specificity of experience-dependent pitch representation in the brainstem. *NeuroReport*, *17*, 1601–1605.
- Yost, W. A. (1996a). Pitch of iterated rippled noise. *Journal of the Acoustical Society of America*, *100*, 511–518.
- Yost, W. A. (1996b). Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America*, *100*, 3329–3335.
- Yost, W. A., & Moore, M. J. (1987). Temporal changes in a complex spectral profile. *Journal of the Acoustical Society of America*, *81*, 1896–1905.