Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem

Gavin M. Bidelman, Jackson T. Gandour, and Ananthanarayan Krishnan

Abstract

■ Neural encoding of pitch in the auditory brainstem is known to be shaped by long-term experience with language or music, implying that early sensory processing is subject to experience-dependent neural plasticity. In language, pitch patterns consist of sequences of continuous, curvilinear contours; in music, pitch patterns consist of relatively discrete, stair-stepped sequences of notes. The primary aim was to determine the influence of domain-specific experience (language vs. music) on the encoding of pitch in the brainstem. Frequency-following responses were recorded from the brainstem in native Chinese, English amateur musicians, and English nonmusicians in response to iterated rippled noise homologues of a musical pitch interval (major third; M3) and a lexical tone (Mandarin tone 2; T2) from the music and language domains, respectively. Pitch-tracking accuracy (whole contour) and pitch strength (50 msec sections) were computed from the brainstem responses using autocorrelation algorithms. Pitch-tracking accuracy was higher in the Chinese and musicians than in the nonmusicians across domains. Pitch strength was more robust across sections in musicians than in nonmusicians regardless of domain. In contrast, the Chinese showed larger pitch strength, relative to nonmusicians, only in those sections of T2 with rapid changes in pitch. Interestingly, musicians exhibited greater pitch strength than the Chinese in one section of M3, corresponding to the onset of the second musical note, and two sections within T2, corresponding to a note along the diatonic musical scale. We infer that experience-dependent plasticity of brainstem responses is shaped by the relative saliency of acoustic dimensions underlying the pitch patterns associated with a particular domain. ■

INTRODUCTION

A longstanding debate in the cognitive neurosciences is whether language and music are processed by distinct and separate neural substrates or, alternatively, whether these two domains recruit similar and perhaps overlapping neural resources. Intimate ties between language and music have been advocated based on evidence from musicology (Feld & Fox, 1994), music theory and composition (Lerdahl & Jackendoff, 1983), acoustics (Ross, Choi, & Purves, 2007), and cognitive neuroscience (Jentschke, Koelsch, Sallat, & Friederici, 2008; Magne, Schon, & Besson, 2006; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Patel, Gibson, Ratner, Besson, & Holcomb, 1998).

Pitch provides an optimal window to study language and music as it is one of the most important information-bearing components shared by both domains (Plack, Oxenham, & Fay, 2005). In language, structure is based upon the hierarchical arrangement of morphemes, words, and phrases, whereas in music, structure relies primarily upon the hierarchical arrangement of pitch (McDermott & Hauser, 2005; Krumhansl, 1990). For comparison with music, tone languages provide a unique opportunity for investigating the linguistic use of pitch (Yip, 2003). In these languages, pitch variations at the syllable or word level are lexically significant. Mandarin Chinese has four lexical tones: ma^1 "mother" [T1], ma^2 "hemp" [T2], ma^3 "horse" [T3], ma^4 "scold" [T4].

There are important differences in how pitch is exploited in each domain. A great deal of music has pitch interval categories, a regular beat, and a tonal center; language does not. Musical melodies are typically organized in terms of pitch intervals governed by a fixed scale; linguistic melodies are not. Linguistic melodies are subject to *declination* and *coarticulation* (Xu, 2006); musical melodies are not. In natural speech, changes in pitch are *continuous* and *curvilinear*, a likely consequence of the physiologic capabilities of the human vocal apparatus as well as speech coarticulation. In music, on the other hand, changes in pitch are quintessentially *discrete* and *stair-stepped* in nature despite the capabilities of many instruments to produce continuous ornamental slides (i.e., glissando, bend, etc.).

It is an intriguing notion that domain-specific experience could positively benefit neural processing in another domain. Recent studies have shown that musical training

Purdue University, West Lafayette, IN

improves phonological processing (Slevc & Miyake, 2006; Anvari, Trainor, Woodside, & Levy, 2002). Indeed, English-speaking musicians show better performance in the identification of lexical tones than nonmusicians (Lee & Hung, 2008). Moreover, neurophysiologic indices show that music training facilitates pitch processing in language (Musacchia, Sams, Skoe, & Kraus, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007; Magne et al., 2006; Schon, Magne, & Besson, 2004). However, it remains an open question to what extent language experience can positively influence music processing (cf. Schellenberg & Peretz, 2008; Schellenberg & Trehub, 2008; Deutsch, Henthorn, Marvin, & Xu, 2006).

The neural representation of pitch may be influenced by one's experience with music or language at subcortical as well as cortical levels of processing (Krishnan & Gandour, 2009; Patel, 2008; Zatorre & Gandour, 2008; Kraus & Banai, 2007; Zatorre, Belin, & Penhune, 2002). As a window into *subcortical* pitch processing in the brainstem, we utilize the human frequency-following response (FFR). The FFR reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem (see Krishnan, 2006 for review of FFR characteristics and source generators). The response is characterized by a periodic waveform which follows the individual cycles of the stimulus waveform. Cross-language comparisons of FFRs show that native experience with a tone language enhances pitch encoding at the level of the brainstem irrespective of speech or nonspeech context (Krishnan, Swaminathan, & Gandour, 2009; Swaminathan, Krishnan, & Gandour, 2008b; Krishnan, Xu, Gandour, & Cariani, 2005). Cross-domain comparisons show that English-speaking musicians are superior to nonmusicians in pitch tracking of Mandarin lexical tones (Wong et al., 2007). Musicians also show more robust pitch encoding, relative to nonmusicians, in response to speech as well as music stimuli (Musacchia, Strait, & Kraus, 2008; Musacchia et al., 2007). Thus, musical training sharpens subcortical encoding of *linguistic* pitch patterns. However, the question remains whether tonal language experience enhances subcortical encoding of *musical* pitch patterns.

To generate auditory stimuli that preserve the perception of pitch, but do not have strict waveform periodicity or highly modulated stimulus envelopes, we employ iterated rippled noise (IRN) (Yost, 1996). A recent modification of the IRN algorithm makes it possible to generate time-variant, dynamic *curvilinear* pitch contours that are representative of those that occur in natural speech (Swaminathan, Krishnan, & Gandour, 2008a; Denham, 2005). Using such IRN homologues, it has been shown that experience-dependent enhancement of pitch encoding in the brainstem extends only to time-varying features of dynamic curvilinear pitch patterns that native speakers of a language are exposed to (Krishnan, Gandour, Bidelman, & Swaminathan, 2009). As far as we know, IRN homologues of music have yet to be exploited to study pitch processing at the brainstem level.

The aim of this study is to determine the nature of the effects of music and language experience on the processing of IRN homologues of pitch contours, as reflected by the FFR in the human auditory brainstem. Specifically, we are interested in whether long-term experience with pitch patterns specific to one domain may differentially shape the neural processing of pitch within another domain. We compare the encoding of prototypical pitch contours from both domains across three groups: native speakers of a tone language, English-speaking amateur musicians, and English-speaking nonmusicians. Prototypical pitch contours from the two domains include a *lexical tone* (mandarin tone 2; T2) and a *pitch interval* (melodic major third; M3). T2 is characteristic of the *continuous, curvilinear* pitch contours that occur in languages of the world, tonal or otherwise (Xu, 2006; Yip, 2003; Gandour, 1994). In contrast, M3 exemplifies the *discrete, stair-stepped* pitch contours that characterize music (Jackendoff, 2009, p.199; Patel, 2008; Peretz & Hyde, 2003, p.365; Zatorre et al., 2002, p. 39; Burns, 1999, p.217; Moore, 1995; Dowling, 1978). We assess pitch-tracking accuracy of Chinese and musically trained individuals in response to *both* music and language stimuli in order to determine whether subcortical pitch encoding in one domain transfers positively to another. We assess pitch strength of *subparts* of music and language stimuli to determine whether domain-dependent pitch processes transfer only to specific acoustic features that are perceptually salient in the listener's domain of pitch expertise. Regardless of domain of pitch expertise, we expect to find that early auditory processing is subject to neural plasticity that manifests itself in stimuli that contain perceptually salient acoustic features which occur within the listener's domain of experience.

METHODS

Participants

Fourteen adult native speakers of Mandarin Chinese (9 men, 5 women), hereafter referred to as Chinese (C), 14 adult monolingual native speakers of English with musical training (9 men, 5 women), hereafter referred to as musicians (M), and 14 adult monolingual native speakers of English without musical training (6 men, 8 women), hereafter referred to as English (E), participated in the FFR experiment. The three groups were closely matched in age (Chinese: $M = 23.8$, $SD = 2.5$; musicians: $M = 23.2$, $SD = 2.3$; English: $M = 24.7$, $SD = 2.9$), years of formal education (Chinese: $M = 17.2$, $SD = 2.1$; musicians: $M = 17.8$, $SD = 1.9$; English: $M = 18.2$, $SD = 2.7$), and were strongly right-handed (>83%) as measured by the Edinburgh Handedness Inventory (Oldfield, 1971). All participants exhibited normal hearing sensitivity (better than 15 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

Table 1. Musical Background of Amateur Musicians

Participant	Instrument(s)	Years of Training	Age of Onset
M1	saxophone/piano	12	11
M2	trombone	10	10
M3	piano/trumpet/flute	16	6
M4	piano/saxophone	11	8
M5	violin/piano	16	3
M6	trumpet/guitar	10	11
M7	saxophone/piano	12	6
M8	violin	9	8
M9	string bass/guitar	11	9
M10	trumpet/piano	14	10
M11	piano/trumpet/guitar	12	7
M12	piano	12	8
M13	violin	16	5.5
M14	piano	10	7

a language history questionnaire (Li, Sepanski, & Zhao, 2006). Native speakers of Mandarin were born and raised in mainland China and none had received formal instruction in English before the age of 9 ($M = 11.4$, $SD = 1.2$). Both English groups had no prior experience learning a tonal language. Each participant also completed a music history questionnaire (Wong & Perrachione, 2007). Musically trained participants were amateur instrumentalists who had at least 9 years of continuous training in the style of Western classical music on their principal instrument ($M = 12.2$, $SD = 2.4$), beginning at or before the age of 11 ($M = 7.8$, $SD = 2.3$) (Table 1). All musician participants had formal private or group lessons within the past 5 years and currently played their instrument(s). Chinese and English participants had no more than 3 years of formal music training ($M = 0.71$, $SD = 0.89$) on any combination of instruments and none had any training within the past 5 years. All participants 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.

IRN Stimuli

IRN was used to create two stimuli with time-varying f_0 contours using procedures similar to those described by Swaminathan et al. (2008a). In the implementation of this algorithm, filtered Gaussian noise (10 to 3000 Hz) is delayed and added back on itself in a recursive manner. This procedure creates the perception of a pitch corresponding to the reciprocal of the delay (Yost, 1996). Instead of a sin-

gle static delay, time-varying delays can be used to create IRN stimuli with dynamic contours whose pitch varies as a function of time (Krishnan, Swaminathan, et al., 2009; Swaminathan et al., 2008a). By using IRN, we preserve dynamic variations in pitch of auditory stimuli that do not have waveform periodicity or highly modulated temporal envelopes characteristic of music or speech. We also remove instrumental quality and formant structure from our stimuli, thereby eliminating potential timbral and lexical/semantic confounds.

The f_0 contour of M3 was modeled with a step function by concatenating two steady-state trajectories together, resulting in the pitch interval of a major third (A♭2 to C3; 103.83–130.81 Hz, respectively). Using two static pitches is motivated by perceptual evidence showing that listeners hear musical notes as single fixed pitches even when they contain the natural embellishments found in acoustic music (e.g., vibrato) (Brown & Vaughn, 1996; d'Alessandro & Castellengo, 1994). Both notes of the interval were each 150 msec in duration (A♭2: 0–150 msec; C3: 150–300 msec). The curvilinear f_0 contour of T2 was modeled after its natural citation form as produced by a male speaker using a fourth-order polynomial equation (Xu, 1997). Its frequency range was then expanded by approximately 2 Hz so that it matched that of M3 (i.e., the span of a major third) (Boersma & Weenink, 2008). The duration of both stimuli was fixed at 300 msec including a 10-msec rise/fall time (cosine-squared ramps) added to minimize onset components and spectral splatter. Both stimuli were also matched in RMS amplitude. These normalizations ensured that our linguistic and musical pitch patterns differed only in f_0 contour (Figure 1).

The two f_0 contours, T2 and M3, were then passed through the IRN algorithm. A high iteration step (32) was used for
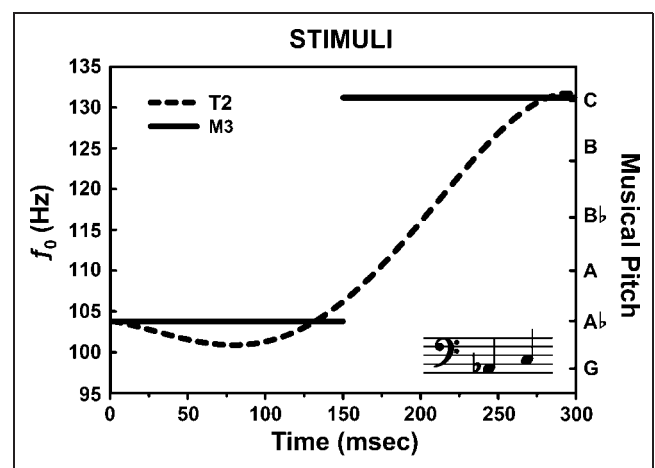

Figure 1. Fundamental frequency contours (f_0) of the IRN stimuli. M3 (solid) is modeled after the musical interval of a major third using two consecutive pitches as notated in the inset (A♭2 to C3; 103.83 to 130.81 Hz, respectively); T2 (dotted) is modeled after Mandarin Tone 2 using a fourth-order polynomial equation (Xu, 1997). Both stimuli are matched in total duration, RMS amplitude, and overall frequency range.

both stimuli with the gain set to 1. At a high iteration step, the IRN stimuli show clear bands ("ripples") of energy in their spectra at f_0 and its harmonics. However, unlike speech or music, they lack both a temporal envelope and a recognizable timbre.

Data Acquisition

Participants reclined comfortably in an acoustically and electrically shielded booth. They were instructed to relax and refrain from extraneous body movements to minimize movement artifacts. In fact, a majority of the participants fell asleep during the procedure. FFRs were recorded from each participant in response to monaural stimulation of the right ear at 80 dB SPL at a repetition rate of 2.44/sec. 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 (System III; Tucker-Davis Technologies, Gainesville, FL). The stimulus files were routed through a digital-to-analog module and presented through a magnetically shielded insert earphone (ER-3A; Etymotic Research, Elkgrove Village, IL).

FFRs were recorded differentially between a noninverting (positive) electrode placed on the midline of the forehead at the hairline (Fz) and inverting (reference) electrodes placed on (i) the right mastoid (A2); (ii) the left mastoid (A1); and (iii) the seventh cervical verterbra (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 were subsequently averaged for each stimulus condition to yield a response with a higher signal-to-noise ratio (Krishnan, Gandour, et al., 2009). All interelectrode impedances were maintained below 1 kΩ. The 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 response waveform represents the average of 3000 stimulus presentations over a 320-msec analysis window using a sampling rate of 25 kHz. The experimental protocol took about 100 min to complete.

Data Analysis

Pitch-tracking Accuracy of Whole Stimuli

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-msec window in 10-msec increments over the time course of the FFR. The autocorrelation function was computed for each 40-msec 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.

Pitch Strength of Stimuli Sections

To compute the pitch strength of the FFRs to *time-varying* IRN stimuli, FFRs were divided into six nonoverlapping 50-msec sections (0–50, 50–100, 100–150, 150–200, 200–250, 250–300 msec). 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 50-msec 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., 2008b). 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).

Statistical Analysis

Pitch-tracking Accuracy of Whole Stimuli

Pitch-tracking accuracy was measured as the cross-correlation coefficient between the f_0 contours extracted from the FFRs and IRN homologues of M3 and T2. A mixed-model ANOVA (SAS), with subjects as a random factor nested within group (C, E, M), which is the between-subject factor, and domain (M3, T2), which is the within-subject factor, was conducted on the cross-correlation coefficients to evaluate the effects of domain-specific experience on the ability of the FFR to track f_0 contours in music and language.

Pitch Strength of Stimulus Sections

Pitch strength (magnitude of the normalized autocorrelation peak) was calculated for each of the six sections of M3 and T2 for every subject. For each domain separately, these pitch strength values were analyzed using an ANOVA with subjects as a random factor nested within group (C, E, M), and section (0–50, 50–100, 100–150, 150–200, 200–250, 250–300 msec) as a within-subject factor. By focusing on the pitch strength of 50-msec sections within these f_0 contours, we were able to determine whether the effects of music and language experience are uniform throughout the duration of the IRN stimuli, or whether they vary depending on specific time-varying f_0 properties within or between contiguous subparts of the stimuli.

RESULTS

Pitch-tracking Accuracy of M3 and T2

Mean stimulus–response correlation coefficients for the C (M3, 0.84; T2, 0.93), M (M3, 0.89; T2, 0.90), and E (M3, 0.62; T2, 0.41) groups are displayed in Figure 2. An omnibus ANOVA on cross-correlation coefficients of IRN homologues of M3 and T2 yielded a significant Group × Domain interaction effect [$F(2, 39) = 13.88$, $p < .0001$]. By group, post hoc Tukey–Kramer adjusted multiple comparisons ($\alpha = .05$) revealed no significant domain effects in either the Chinese or musician group, whereas pitch tracking of M3 was more accurate than T2 in the English group. Regardless of the domain, both the C and M groups were more accurate than E in pitch tracking. Yet neither M3 nor T2 elicited a significant difference in pitch-tracking accuracy between Chinese and musically trained individuals.

Pitch Strength of Sections within M3 and T2

FFR pitch strength, as measured by the average magnitude of the normalized autocorrelation peak per group, is shown for six sections within each of the IRN homologues of M3 and T2 (Figure 3).

Results from omnibus two-way ANOVAs of pitch strength in M3 and T2 revealed a significant interaction between
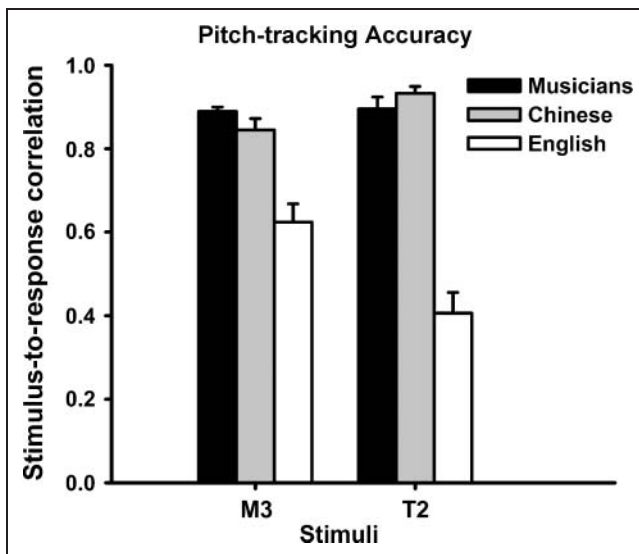

Figure 2. Cross-domain comparison of FFR pitch-tracking accuracy between groups. Bars represent the group means of the stimulus-to-response correlation coefficients of musicians (black), Chinese (gray), and nonmusicians (white), respectively. Error bars indicate one standard error of the mean. Both Chinese and musicians are superior in their tracking ability as compared to English nonmusicians, regardless of domain. Long-term experience with musical and linguistic pitch patterns transfer across domains. Musicians are comparable to Chinese in their ability to track T2; and likewise, Chinese are comparable to musicians in their ability to track M3.

group and section in both domains [M3: $F(10, 195) = 2.04$, $p = .0315$; T2: $F(10, 195) = 3.46$, $p = .0003$]. A priori contrasts of groups were performed using a Bonferroni adjustment ($\alpha = .0166$) per section. In the case of C versus E (Figure 3, top panels), pitch strength was greater for the Chinese group in all but the last section of M3, and in Sections 3 to 5 of T2. In the case of M versus E (Figure 3, middle panels), pitch strength was greater for the M group across the board irrespective of domain. In the case of M versus C (Figure 3, bottom panels), pitch strength was greater for the M group across domains but only in a limited number of sections, two in M3, and three in T2. The two sections (4 and 6) of M3 correspond to the onset and offset of the second note in the major third pitch interval, respectively. The three sections (1, 4, and 5) of T2, respectively, correspond to the onset and the portions of T2 where its curvilinear f_0 contour coincides with a pitch along the diatonic music scale (B♭: 116.54 Hz).

Spectral f_0 Magnitudes within Region of Interest of T2

We further examined each FFR response *within* Sections 4 and 5 of T2 to determine whether the musicians' advantage over Chinese is attributable to the musical scale. Running FFTs were computed using a 50-msec analysis window incremented by 5 msec, and zero-padding was implemented to obtain high-frequency resolution (~1 Hz). f_0 was defined as the dominant component in the short-term FFT falling within the frequency range of the stimulus (100–130 Hz). f_0 magnitude of musicians is greater than either Chinese or nonmusicians in the portion of T2 corresponding to the musical pitch B♭ (Figure 4; cf. Figure 1, ~200 msec). Comparing the two groups with domain-specific pitch expertise, we further observed that f_0 magnitude at B♭ is 6 dB greater in musicians than Chinese.

A one-way ANOVA was performed on the spectral f_0 magnitude of three frequencies within this 15-Hz span of T2. One frequency corresponds to a prominent note on the diatonic musical scale (B♭ = 116.5 Hz); the other two do not (cf. Figure 4; down arrows at 111.5, 121.5 Hz). Results revealed a significant interaction between group and frequency [$F(4, 54) = 4.30$, $p = .0043$]. By frequency, post hoc multiple comparisons ($\alpha_{Bonferroni} = .0166$) revealed that spectral f_0 magnitude within this region of interest was greater in musicians than Chinese for B♭ only.

DISCUSSION

Using IRN homologues of musical and linguistic pitch contours, the major findings of this cross-language, cross-domain study demonstrate that experience-dependent neural mechanisms for pitch representation at the brainstem level, as reflected in pitch-tracking accuracy and pitch strength, are more sensitive in Chinese and amateur
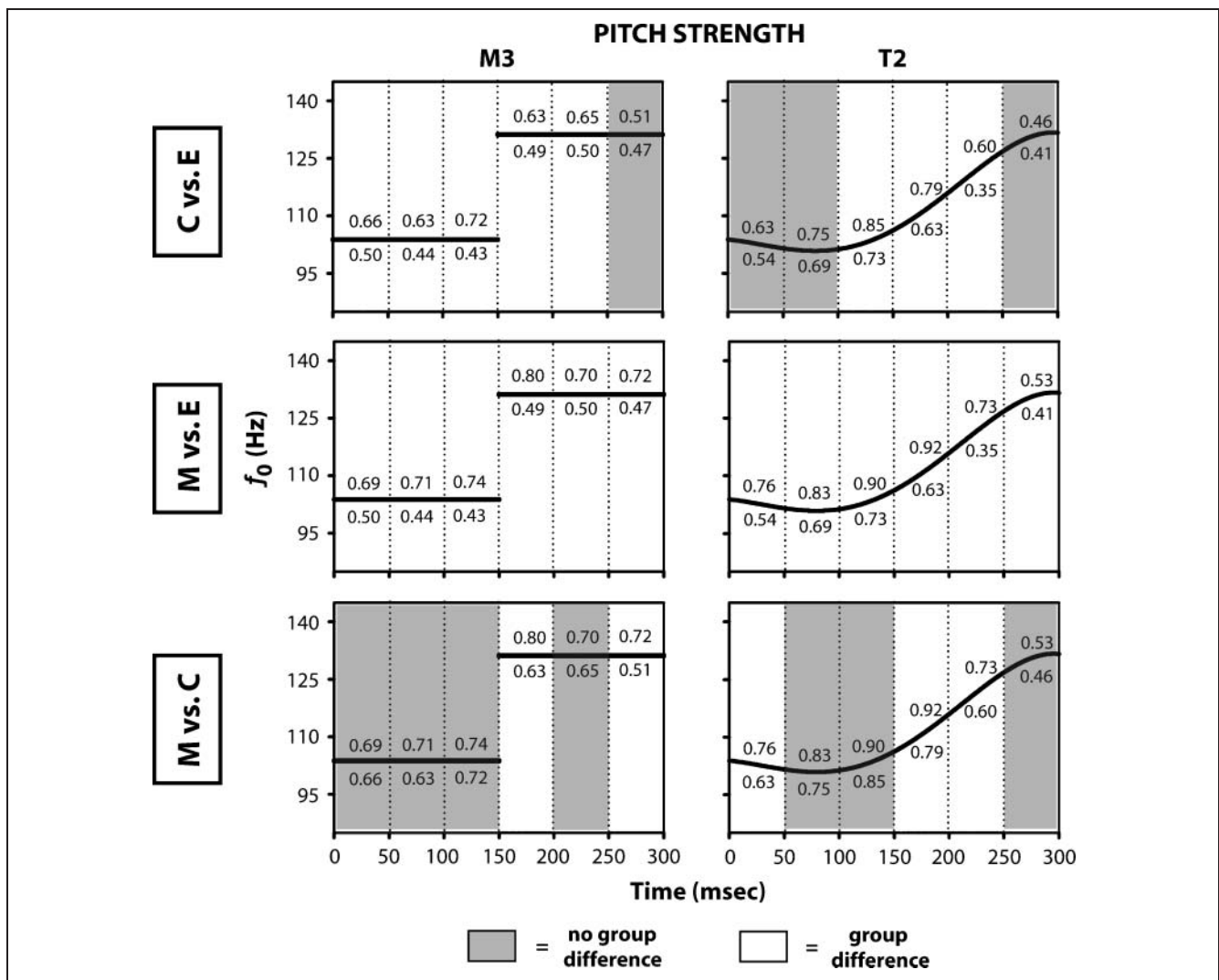
PITCH STRENGTH

M3 T2

f_0 (Hz)

C vs. E

M vs. E

M vs. C

Time (msec)

☐ (gray) = no group difference ☐ (white) = group difference

Figure 3. Group comparisons of pitch strength derived from the FFR waveforms in response to sections of musical (M3) and linguistic (T2) f_0 contours. Chinese (C) vs. English nonmusicians (E), row 1; musicians (M) vs. E, row 2; M vs. C, row 3. Vertical dotted lines demarcate six 50-msec sections within each f_0 contour: 0–50, 50–100, 100–150, 150–200, 200–250, 250–300 msec. Sections that yielded significantly larger pitch strength for the C and the M groups relative to E are unshaded; those that did not are shaded in gray. Top row: C (values above solid line) exhibits greater pitch strength than E (values below solid line) in nearly all sections of M3, and in those sections of T2 that exhibit rapid changes in f_0 movement. Middle row: M (above) exhibits greater pitch strength than E (below) across the board, irrespective of domain. Bottom row: M (above) exhibits greater pitch strength than C (below), most notably in those sections that are highly relevant to musical pitch perception, regardless of the domain of the f_0 contour. Although musicians have larger pitch strength than Chinese in the final section of M3 and the beginning section of T2, stimulus ramping and the absence of a preceding/following note preclude firm conclusions regarding group differences in onset/offset encoding of the stimuli.

musicians as compared to nonmusicians across domains. Despite the striking differences in the nature of their pitch experience, Chinese and musicians, relative to nonmusicians, are both able to transfer their abilities in pitch encoding across domains, suggesting that brainstem neurons are differentially sensitive to changes in pitch without regard to the domain or context in which they are presented. As reflected in pitch strength, a direct comparison of Chinese and musicians reveals that pitch encoding is superior in musicians across domains, but only in those subparts of the musical pitch interval (M3) and the lexical high rising tone (T2) that can be related to perceptually salient notes along the musical scale.

Experience-dependent Plasticity of Brainstem Mechanisms underlying Pitch Extraction

Our findings provide further evidence for experience-dependent plasticity induced by long-term experience with ecologically relevant pitch patterns found in language and music. Pitch encoding is stronger in Chinese and musicians as compared to individuals who are untrained musically and who are unfamiliar with the use of pitch in tonal languages (i.e., English nonmusicians). This finding demonstrates that the sustained phase-locked activity in the rostral brainstem is enhanced after long-term experience with pitch regardless of domain. Whether

lexical tones or musical pitch intervals, these individuals' brainstems are tuned to extract dynamically changing interspike intervals that cue linguistically or musically relevant features of the auditory signal. As such, our findings converge with previous FFR studies which demonstrate that subcortical pitch processing is enhanced for speakers of a tonal language (Krishnan et al., 2005) and individuals with extensive musical training (Musacchia et al., 2007, 2008; Wong et al., 2007).

As a function of pitch experience across languages, Chinese exhibit more robust pitch strength than English nonmusicians, but only in those dynamic segments of T2 exhibiting higher degrees of pitch acceleration (i.e., more rapid pitch change; Figure 3, Sections 3–5). In agreement with previous FFR studies (Krishnan, Gandour, et al., 2009; Krishnan, Swaminathan, et al., 2009; Swaminathan et al., 2008b; Wong et al., 2007), this finding reinforces the view that the advantage of tone language experience does not necessarily apply across the board, and is mainly evident in just those sections of an f_0 contour that exhibit rapid changes of pitch. We infer that the FFRs of the Chinese group reflect a processing scheme that is streamlined for dynamic pitch changes over relatively short time intervals. Such a scheme follows as a consequence of their long-term experience linguistically relevant pitch patterns that occur at the syllable level. Indeed, speech production data has shown that f_0

Figure 4. Group comparisons of spectral f_0 magnitudes in a region of interest spanning the most rapid changes of pitch in T2. Despite the continuous nature of T2, musicians show enhanced pitch encoding relative to Chinese and nonmusicians in that portion localized to the musical pitch B♭. These group differences suggest that musically trained individuals extract pitch information in relation to the discrete musical scale at the level of the brainstem. Each point represents the mean FFT magnitude (raw microvolt amplitudes were normalized between 0 and 1) per group computed at a particular frequency. Shaded regions show ±1 *SE*. Downward arrows denote the two "off" frequencies used for statistical comparison to B♭.

patterns in Mandarin have a greater amount of dynamic movement as a function of time and number of syllables than those found in English (Eady, 1982).

As a function of pitch experience across domains, musicians exhibit greater pitch strength than Chinese in only two of the six 50-msec sections of M3 (Figure 3; Sections 4 and 6). These two sections correspond to the onset and offset of the second musical note within the major third pitch interval. The fact that amateur musicians have enhanced encoding for instantaneous changes in pitch height of this magnitude (4 semitones) is a consequence of their extensive experience with the discrete nature of musical melodies. Pitch changes within the fixed hierarchical scale of music are more demanding than those found in language (Andrews & Dowling, 1991; Dowling & Bartlett, 1981). To cope with these demands, musicians may develop a more acute, and possibly more adaptive, temporal integration window (Warrier & Zatorre, 2002).

One unexpected finding is that musicians show greater pitch strength than Chinese in two consecutive sections of T2 (Figure 3; Sections 4 and 5). The greater pitch strength of musicians in these sections may be the result of their superior ability to accurately encode rapid, fine-grained changes in pitch. This is consistent with a musician's capacity for detecting minute variations in pitch (e.g., in tune vs. out of tune). Another plausible explanation is based on the intriguing fact that these two sections straddle a time position where the curvilinear pitch contour of T2 passes directly through a note along the diatonic musical scale (B♭: 116.54 Hz; Figure 1, 200 msec). Despite the unfamiliarity with T2, musicians seemingly exploit local mechanisms in the auditory brainstem to extract pitch in relation to a fixed, hierarchical musical scale (Figure 4). No such pitch hierarchy is found in language. In this experiment, T2 spans a frequency range of a major third (A♭2 to C3). Musicians show enhanced encoding of the intermediate diatonic pitch B♭2 by "filling in" the major third (i.e., do-RE-mi). No enhancement was observed in the two other chromatic pitches within this range (A♮ or B♮) because these notes are less probable in the major/minor musical context examined here (key of A♭).

We hypothesize that the pitch axis of a musician's brainstem is arranged in a piano-like fashion, showing more sensitivity to pitches that correspond to discrete notes along the musical scale than to those falling between them. These enhancements are the result of many years of active engagement during hours of practice on an instrument. The musician's brainstem is therefore tuned by long-term exposure to the discrete pitch patterns inherent to instrumental scales and melodies. Work is currently underway in our lab to rigorously test this hypothesis by presenting musicians with a continuous frequency sweep spanning a much larger musical interval (e.g., perfect fifth) over a much larger frequency range (e.g., hundreds of Hz). We expect to see local enhancement for those frequencies which correspond to notes along the diatonic musical scale relative to those which do not.
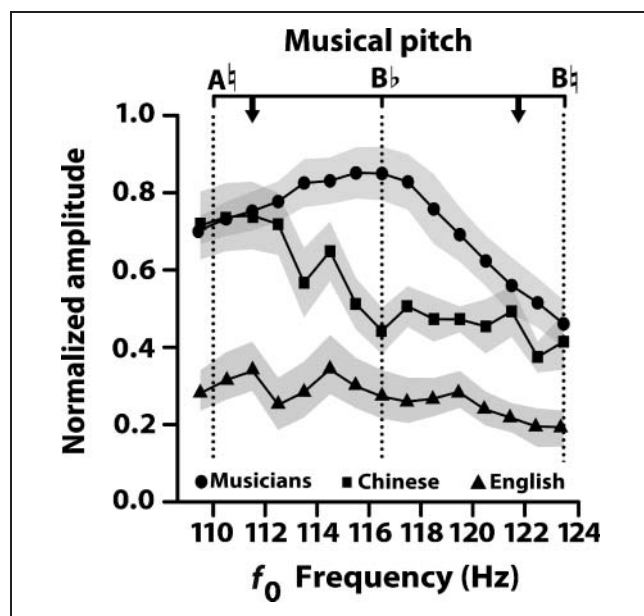
Corticofugal vs. Local Brainstem Mechanisms underlying Experience-dependent Pitch Encoding

We utilize an empirically driven theoretical framework to account for our data showing experience-dependent pitch representation in the brainstem (Krishnan & Gandour, 2009). The corticofugal system is crucially involved in the experience-driven reorganization of subcortical neural mechanisms. It can lead to enhanced subcortical processing of behaviorally relevant parameters in animals (Suga, Ma, Gao, Sakai, & Chowdhury, 2003). In humans, it likely shapes the reorganization of brainstem mechanisms for enhanced pitch extraction at earlier stages of language development and music learning. Once this reorganization is complete, however, local mechanisms in the brainstem are sufficient to extract relevant pitch information in a robust manner without permanent corticofugal influence (Krishnan & Gandour, 2009). We infer that the enhanced pitch representation in native Chinese and amateur musicians reflect an enhanced tuning to interspike intervals that correspond to the most relevant pitch segments in each domain. Long-term experience appears to sharpen the tuning characteristics of the best modulation frequency neurons along each pitch axis with particular sensitivity to acoustic features that are most relevant to each domain.

Emergence of Domain-relevant Representations at Subcortical Stages of Processing

Although music and language have been shown to recruit common neural resources in cerebral cortex, it is important to bear in mind the level of representation and the time course in which such overlaps occur. For either music or language, neural networks likely involve a series of computations that apply to representations at different stages of processing (Poeppel, Idsardi, & van Wassenhove, 2008; Hickok & Poeppel, 2004). We argue that our FFR data provide a window on the nature of intermediate, subcortical pitch representations at the level of the midbrain which, in turn, suggests that higher-level abstract representations of speech and music are grounded in lower-level sensory features that emerge very early along the auditory pathway.

The auditory brainstem is domain general insomuch as it mediates pitch encoding in both music and language. As a result, both Chinese and musicians show positive transfer and parallel enhancements in their subcortical representation of pitch. Yet the emergence of domain-dependent extraction of pitch features (e.g., M3: Section 4; T2: Sections 4–5) highlight the fact that their pitch extraction mechanisms are not homogeneous. Indeed, how pitch information is extracted depends on the *interactions* between specific features of the input signal, their corresponding output representations, and the domain of pitch experience of the listener (cf. Zatorre, 2008, p. 533). Such insights into the neural basis of pitch processing across domains are made possible by means of a cross-cultural study of music and language.

Conclusions

Cross-domain effects of pitch experience in the brainstem vary as a function of stimulus and domain of expertise. Experience-dependent plasticity of the FFR is shaped by the relative saliency of acoustic dimensions underlying pitch patterns associated with a particular domain. Pitch experience in either music or language can transfer from one domain to the other. Music overrides language in pitch encoding in just those phases exhibiting rapid changes in pitch that are perceptually relevant on a musical scale. Pitch encoding from one domain of expertise may transfer to another as long as the latter exhibits acoustic features overlapping those with which individuals have been exposed to from long-term experience or training.

REFERENCES

Andrews, M. W., & Dowling, W. J. (1991). The development of perception of interleaved melodies and control of auditory attention. *Music Perception, 8,* 349–368.

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing and early reading ability in preschool children. *Journal of Experimental Child Psychology, 83,* 111–130.

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) [Computer program].* Amsterdam: Institute of Phonetic Sciences. Available from www.praat.org/.

Brown, J. C., & Vaughn, K. V. (1996). Pitch center of stringed instrument vibrato tones. *Journal of the Acoustical Society of America, 100,* 1728–1735.

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 215–264). San Diego, CA: Academic Press.

d'Alessandro, C., & Castellengo, M. (1994). The pitch of short-duration vibrato tones. *Journal of the Acoustical Society of America, 95,* 1617–1630.

Denham, S. (2005). Pitch detection of dynamic iterated rippled noise by humans and a modified auditory model. *Biosystems, 79,* 199–206.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a

speech-related critical period. *Journal of the Acoustical Society of America, 119,* 719–722.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85,* 341–354.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1,* 30–49.

Eady, S. J. (1982). Differences in the F0 patterns of speech: Tone language versus stress language. *Language and Speech, 25,* 29–42.

Feld, S., & Fox, A. (1994). Music and language. *Annual Review of Anthropology, 23,* 25–53.

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.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92,* 67–99.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception, 26,* 195–204.

Jentschke, S., Koelsch, S., Sallat, S., & Friederici, A. D. (2008). Children with specific language impairment also show impairment of music-syntactic processing. *Journal of Cognitive Neuroscience, 20,* 1940–1951.

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP Study. *Journal of Cognitive Neuroscience, 17,* 1565–1577.

Kraus, N., & Banai, K. (2007). Auditory-processing malleability: Focus on language and music. *Current Directions in Psychological Science, 16,* 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, MD: Lippincott Williams & Wilkins.

Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language, 110,* 135–148.

Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience-dependent neural representation of dynamic pitch in the brainstem. *NeuroReport, 20,* 408–413.

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,* 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,* 161–168.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University 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,* 3235–3248.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A Web-based interface for bilingual research. *Behavioral Research Methods, 38,* 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,* 199–211.

McDermott, J., & Hauser, M. D. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception, 23,* 29–59.

Moore, B. C. J. (1995). *Hearing.* San Diego, CA: 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.

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Research, 241,* 34–42.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

Patel, A. D. (2008). *Music, language, and the brain.* New York: Oxford University Press.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10,* 717–733.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences, 7,* 362–367.

Plack, C. J., Oxenham, A. J., & Fay, R. R. (Eds.) (2005). *Pitch: Neural coding and perception* (Vol. 24). New York: Springer.

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,* 1071–1086.

Ross, D., Choi, J., & Purves, D. (2007). Musical intervals in speech. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 9852–9857.

Schellenberg, E. G., & Peretz, I. (2008). Music, language and cognition: Unresolved issues. *Trends in Cognitive Sciences, 12,* 45–46.

Schellenberg, E. G., & Trehub, S. E. (2008). Is there an Asian advantage for pitch memory? *Music Perception, 25,* 241–252.

Schon, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41,* 341–349.

Slevc, R. L., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science, 17,* 675–681.

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.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception and Psychophysics, 64,* 198–207.

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.

Yip, M. (2003). *Tone*. New York: Cambridge University Press.

Yost, W. A. (1996). Pitch of iterated rippled noise. *Journal of the Acoustical Society of America, 100,* 511–518.

Zatorre, R. J. (2008). Musically speaking. *Neuron, 26,* 532–533.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6,* 37–46.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 363,* 1087–1104.