EXPERIENCE-DEPENDENT ENHANCEMENT OF PITCH-SPECIFIC RESPONSES IN THE AUDITORY CORTEX IS LIMITED TO ACCELERATION RATES IN NORMAL VOICE RANGE

A. KRISHNAN, J. T. GANDOUR * AND C. H. SURESH

Department of Speech Language Hearing Sciences, Purdue University, USA

Abstract—The aim of this study is to determine how pitch acceleration rates within and outside the normal pitch range may influence latency and amplitude of cortical pitch-specific responses (CPR) as a function of language experience (Chinese, English). Responses were elicited from a set of four pitch stimuli chosen to represent a range of acceleration rates (two each inside and outside the normal voice range) imposed on the high rising Mandarin Tone 2. Pitch-relevant neural activity, as reflected in the latency and amplitude of scalp-recorded CPR components, varied depending on language-experience and pitch acceleration of dynamic, time-varying pitch contours. Peak latencies of CPR components were shorter in the Chinese than the English group across stimuli. Chinese participants showed greater amplitude than English for CPR components at both frontocentral and temporal electrode sites in response to pitch contours with acceleration rates inside the normal voice pitch range as compared to pitch contours with acceleration rates that exceed the normal range. As indexed by CPR amplitude at the temporal sites, a rightward asymmetry was observed for the Chinese group only. Only over the right temporal site was amplitude greater in the Chinese group relative to the English. These findings may suggest that the neural mechanism(s) underlying processing of pitch in the right auditory cortex reflect experience-dependent modulation of sensitivity to acceleration in just those rising pitch contours that fall within the bounds of one's native language. More broadly, enhancement of native pitch stimuli and stronger rightward asymmetry of CPR components in the Chinese group is consistent with the notion that long-term experience shapes adaptive, distributed hierarchical pitch processing in the auditory cortex, and reflects an interaction with higher order, extrasensory processes beyond the sensory memory trace. © 2015 IBRO. Published by Elsevier Ltd. All rights reserved.

Key words: auditory, pitch encoding, iterated rippled noise, cortical pitch response, pitch acceleration, Mandarin Chinese.

Pitch is an important information-bearing perceptual attribute that provides an excellent window for studying language-dependent effects on pitch processing at both subcortical and cortical levels. Tone languages are especially advantageous for investigating the linguistic use of pitch because variations in pitch patterns at the syllable level may be lexically significant (Laver, 1994, p. 465). Neural representation of pitch may be influenced by one's experience with language (or music) at subcortical as well as cortical levels of processing (for reviews, see Johnson et al., 2005; Kraus and Banai, 2007; Patel and Iversen, 2007; Zatorre and Gandour, 2008; Krishnan and Gandour, 2009; Tzounopoulos and Kraus, 2009; Krishnan and Gandour, 2014).

Experience-dependent enhancement of pitch-relevant phase-locked neural activity in the auditory brainstem has been observed for only specific portions of native pitch contours exhibiting high rates of pitch acceleration, irrespective of speech or nonspeech contexts (Krishnan et al., 2009). This heightened sensitivity to sections characterized by rapid changes in pitch is maintained even in severely degraded stimuli (Krishnan et al., 2010a). Particularly relevant here are our previous results (Krishnan et al., 2010b) that exhibit more robust brainstem representation of pitch-relevant information in Chinese listeners, relative to English, across a four-step acceleration rate continuum where the lowest rate was equivalent to Mandarin Tone 2 (T2) and the fastest rate fell well outside the normal range of dynamic pitch. Regardless of language group, neural periodicity strength was greater in response to acceleration rates within or proximal to natural speech relative to those beyond its range. Though both groups showed decreasing pitch strength with increasing acceleration rates, pitch representations of the Chinese group were more resistant to degradation. These findings indicate that perceptually salient pitch cues associated with lexical tone influence brainstem pitch extraction not only in the speech domain, but also in auditory signals that clearly fall outside the range of dynamic pitch that a native listener is exposed to.

Similar to the sensitivity to pitch acceleration observed in the brainstem, we recently reported that Na–Pb and Pb–Nb components of the cortical pitch response (CPR) are also sensitive to variations in pitch acceleration presented in three within-category variants of T2 (Krishnan et al., 2014a). In a follow-up study, we examined how language experience (Chinese, English) shapes the processing of different temporal attributes of pitch reflected in the CPR components using the same three variants of T2 (Krishnan et al., 2015a). Results showed that the magnitude of CPR components (Na–Pb and Pb–Nb) were larger for Chinese listeners in response to pitch stimuli, all of which fell within the range of T2 citation forms. It remains to be determined whether language-dependent effects on CPR components will extend across a range of acceleration rates, including those that exceed the normal pitch range as reported in the auditory brainstem (Krishnan et al., 2010b). Though it is true that both brainstem and CPR responses are of a sensory nature, they index pitch-relevant neural activity at lower and higher structural levels of the brain, respectively. By isolating a single pitch parameter (i.e., rate of pitch acceleration), the CPR provides us with an early, preattentive window on pitch processing that may involve an interaction between sensory and extrasensory effects whose relative weighting varies depending on language experience.

Using the same set of pitch stimuli as in Krishnan et al. (2010b), the aim of this study is to determine how pitch acceleration within and outside the normal pitch range may influence latency and amplitude of CPR components as a function of language experience (Chinese, English). We hypothesize that enhancement of pitch-relevant neural activity and presence of asymmetry in cortical pitch representation is language-dependent and limited to pitch contours with acceleration rates that fall within the normal pitch range. In the case of acceleration rates beyond the normal pitch range, CPR components are expected to index primarily auditory effects regardless of language group. Stimulus comparisons at frontal and temporal electrode sites allow us to assess the extent of asymmetry. We hypothesize that at the right temporal site, the pattern of changes in the CPR components reflect temporally distinct, differential weighting of sensory and extrasensory effects depending on language experience.

EXPERIMENTAL PROCEDURES

Participants

Fifteen native speakers of Mandarin Chinese (C; eight female) and native, monolingual speakers of American English (E; nine female) were recruited from the Purdue University student body to participate in the experiment. All were strongly right-handed (C: $92.2 \pm 11.8\%$; E: $95.8 \pm 9.2\%$) as measured by the laterality index of the Edinburgh Handedness Inventory (Oldfield, 1971). They were closely matched in age (C: 23.73 ± 3.59 years; E: 22.13 ± 1.68), years of formal education (C: 17.07 ± 2.49 years; E: 15.53 ± 0.99). All exhibited normal hearing sensitivity at audiometric frequencies between 500 and 4000 Hz and reported no previous

history of neurological or psychiatric illnesses. All Chinese participants were born and raised in mainland China. None had received formal instruction in English before the age of nine (12.07 ± 1.94 years). Self-ratings of their English language proficiency on a seven-point Likert-type scale ranging from 1 (very poor) to 7 (native-like) for speaking and listening abilities were, on average, 5.0 and 5.7, respectively (Li et al., 2006). Their daily usage of Mandarin and English, in order, was reported to be 64% and 36%. As determined by a music history questionnaire (Wong and Perrachione, 2007), all Chinese and English participants had less than two years of musical training (C, 0.64 ± 0.82 years; E, 0.71 ± 0.83) on any combination of instruments. No participant had any training within the past five years. Each participant was paid and gave informed consent in conformity with the 2013 World Medical Association Declaration of Helsinki and in compliance with an experimental protocol approved by the Institutional Review Board of Purdue University.

Stimuli

A set of four pitch stimuli were chosen to represent a range of acceleration rates imposed on Mandarin Tone 2 (Fig. 1; cf. Krishnan et al., 2010b, time-varying click trains). Two of them fell within the normal voice pitch range; two fell outside. These acceleration rates were derived from a study on the maximum speed at which a speaker can voluntarily change pitch (Xu and Sun, 2002). The maximum velocity in a rising direction was reported to be 61.3 semitones per second (st/s). In this study, we measured acceleration from turning point to tonal offset, an excursion time of 185 ms. As compared to 61.3 st/s, acceleration rate prototypical of Mandarin Tone 2 (A1) (25.4 st/s) falls well within the physiological limits of speed of rising pitch changes; faster acceleration rate of Tone 2 still within normal voice pitch range (A2) (51.94 st/s) similarly falls under the maximum, but is beginning to approach the maximum limit for changes in rising pitch; both faster acceleration rate of Tone 2 beyond normal voice pitch range (A3) (85.44 st/s) and faster acceleration rate of Tone 2 far outside normal voice pitch range (A4) (129.5 st/s) go far beyond the limit for changes in rising pitch. The first stimulus (A1) is represented by the f_0 contour of Mandarin Tone 2 as produced in citation form (Xu, 1997), and modeled by a fourth-order polynomial (Swaminathan et al., 2008). It exhibits an acceleration rate prototypical of Tone 2 (A1: Δ 31.2 Hz/0.38 octaves); Δ represents the difference in f_0 between turning point (65 ms), the minimum f_0 along the duration of the pitch contour, and offset (250 ms). The fourth stimulus is a scaled variant of Tone 2 that extends beyond the limits of the normal voice range (A4: Δ 299.2 Hz/1.99 octaves). Of the two intermediate f_0 contours, one represents a pitch pattern that approaches the upper margin of the normal voice range (A2: Δ 74.2 Hz/0.74 octaves); the other, a pitch pattern that does not occur within a normal voice range (A3: Δ 149.2 Hz/1.48 octaves).

The rationale for our choice of stimuli was based on phonetic, psychoacoustic, and previous empirical data
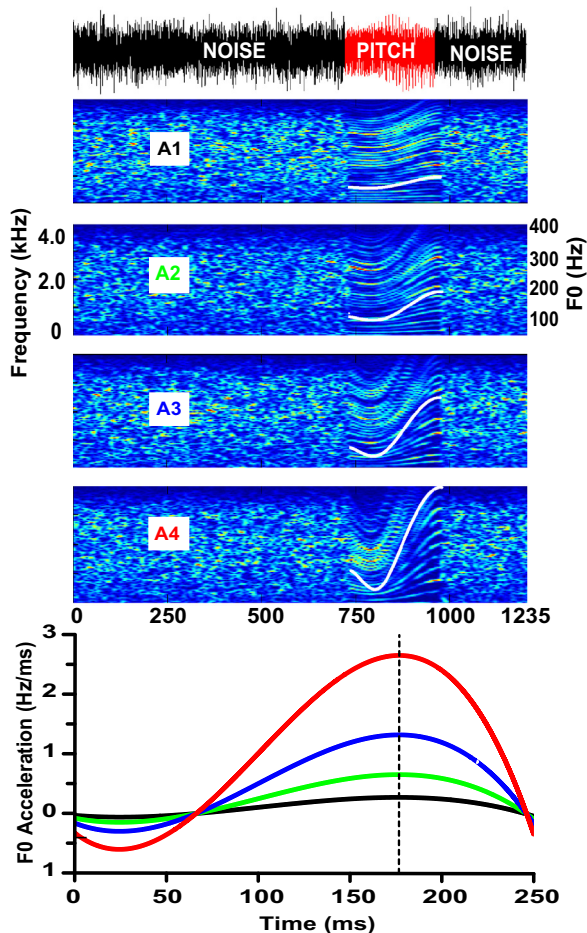
Fig. 1. Waveform (top: A1) and spectrograms (middle: A1, A2, A3, A4) of each of the four stimulus conditions illustrate the experimental paradigm used to acquire cortical responses. The waveform shows robust periodicity within the pitch segment (red) for A1, immediately preceded and followed by noise segments (black). f_0 contours (middle: white) and corresponding acceleration trajectories (bottom) are displayed for all four IRN stimuli on a logarithmic scale spanning two octaves – from 100.8 Hz, the minimum stimulus frequency, to 400 Hz. The vertical dashed line (177 ms) pinpoints the location of the maximum pitch acceleration. A1 (0.3 Hz/ms), A2 (0.7), A3 (1.3), and A4 (2.7), pitch stimuli; IRN, iterated rippled noise; f_0, voice fundamental frequency. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from cortical and subcortical processing. Our aim was to select a lexical tone from the Mandarin tonal inventory that exhibited a fundamentally, unidirectional change in pitch direction between onset and offset. This phonetic criterion ruled out Tone 1 (high level) and Tone 3 (bidirectional falling-rising), but admitted both T2 (unidirectional rising) and T4 (unidirectional falling). We chose T2 primarily because it has revealed experience-dependent plasticity at both brainstem and cortical levels (for reviews, Krishnan et al., 2012b; Krishnan and Gandour, 2014). These language-dependent effects point especially to those pitch attributes that are *perceptually salient* in a particular language. In a behavioral experiment using excised segments from F0 contours of Mandarin tones (Whalen and Xu, 1992), tonal recognition

was shown to be markedly better in the later segments of portions of Tone 2 (rising) and Tone 4 (falling). It is precisely those portions that coincide with a large change in F0. Just noticeable differences for detection of changes in slope of linear F0 ramps showed greatest sensitivity when one ramp is rising and the other is falling (Klatt, 1973). More recently, just noticeable differences for rising tones have been shown to be higher in English relative to Chinese listeners (Liu, 2013). Multidimensional scaling analyses have shown that the underlying perceptual dimension related to direction of pitch change in the stimulus space separates primarily rising versus non-rising F0 movements (Gandour and Harshman, 1978; Gandour, 1983).

Iterated rippled noise (IRN) was used to create these stimuli by applying polynomial equations that generate dynamic, curvilinear pitch patterns (Swaminathan et al., 2008). A high iteration step ($n = 32$) was chosen because pitch salience does not increase by any noticeable amount beyond this number of iteration steps. The gain was set to one. By using IRN, we create pitch-specific stimuli that preserve dynamic variations in pitch of auditory stimuli that lack a waveform periodicity, formant structure, temporal envelope, and recognizable timbre characteristic of speech. Each stimulus condition consisted of three segments (crossfaded with 7.5 ms $\cos^2$ ramps): a 250-ms pitch segment (A1, A2, A3, A4) preceded by a 750-ms noise segment and followed by a 250-ms noise segment (Fig. 1, top panel). The overall root-mean-square level of each segment was equated such that there was no discernible difference in intensity among the three segments. All stimuli were presented binaurally at 80-dB SPL through magnetically shielded tubal insert earphones (ER-3A; Etymotic Research, Elk Grove Village, IL, USA) with a fixed onset polarity (rarefaction) and a repetition rate of 0.56/s. Stimulus presentation order was randomized both within and across participants. All stimuli were generated and played out using an auditory-evoked potential system (SmartEP, Intelligent Hearing Systems; Miami, FL, USA).

CPR acquisition

Participants reclined comfortably in an electroacoustically shielded booth to facilitate recording of neurophysiologic responses. They were instructed to relax and refrain from extraneous body movement to minimize myogenic artifacts, and to ignore the stimuli as they watched a silent video (minus subtitles) of their choice throughout the recording session. The electroencephalogram (EEG) was acquired continuously (5000-Hz sampling rate; 0.3–2500-Hz analog bandpass) using ASA-Lab EEG system (ANT Inc., The Netherlands) utilizing a 32-channel amplifier (REFA8-32, TMS International BV) and WaveGuard (ANT Inc., The Netherlands) electrode cap with 32-shielded sintered Ag/AgCl electrodes configured in the standard 10–20-montage system. The high sampling rate of 5 kHz was necessary to recover the brainstem frequency following responses (not reported herein) in addition to the relatively slow cortical pitch components. Because the primary objective of this study was to characterize the

cortical pitch components, the EEG acquisition electrode montage was limited to nine electrode locations: Fpz, AFz, Fz, F3, F4, Cz, T7, T8, M1, M2. The AFz electrode served as the common ground and the common average of all connected unipolar electrode inputs served as default reference for the REFA8-32 amplifier. An additional bipolar channel with one electrode placed lateral to the outer canthi of the left eye and another electrode placed above the left eye was used to monitor artifacts introduced by ocular activity. Inter-electrode impedances were maintained below 10 k Ω. For each stimulus, EEGs were acquired in blocks of 1000 sweeps. The experimental protocol took about 2 h to complete.

Extraction of the CPR

CPR responses were extracted off-line from the EEG files. To extract the CPR components, EEG files were first down sampled from 5000 Hz to 2048 Hz. They were then digitally band-pass filtered (3–25 Hz, Butterworth zero phase shift filter with 24 dB/octave rejection rate) to enhance the transient components and minimize the sustained component. Sweeps containing electrical activity exceeding $\pm 50\ \mu V$ were rejected automatically. Subsequently, averaging was performed on all eight unipolar electrode locations using the common reference to allow comparison of CPR components at the right frontal (F4), left frontal (F3), right temporal (T8), and left temporal (T7) electrode sites to evaluate laterality effects. Given the poor spatial resolution of EEG even using multiple electrodes, the focus here is not to localize the source of the CPRs with just two electrodes, but to characterize the relative difference in the pitch-related neural activity over the widely separated right and left temporal electrode sites. We are quite confident that robust differences in CPR-related neural activity over the T7 and T8 electrode sites do represent a functional, experience-dependent rightward asymmetry that we have consistently observed across our crosslanguage CPR studies (Krishnan et al., 2014b, 2015a,b). The re-referenced electrode site, Fz-linked(T7/T8), was used to characterize the transient pitch response components. The Fz electrode site was chosen because both the MEG- and EEG-derived pitch responses are prominent at the fronto-central sites (Krumbholz et al., 2003; Bidelman and Grall, 2014; Krishnan et al., 2014b, 2015a,b). It also allows us to compare our CPR data with the Fz-derived POR data (Gutschalk et al., 2002, 2004; Krumbholz et al., 2003). In addition, this electrode configuration was exploited to improve the signal-to-noise ratio of the CPR components by differentially amplifying (i) the non-inverted components recorded at Fz-linked(T7/T8) and (ii) the inverted components recorded at the temporal electrode sites (T7 and T8). Finally, this identical electrode configuration makes it possible for us to compare these CPR responses with brainstem responses in subsequent experiments. For both averaging procedures, the analysis epoch was 1600 ms including the 100-ms pre-stimulus baseline.

Analysis of CPR

The evoked response to the entire three segment (noise-pitch-noise) stimulus is characterized by obligatory components (P1/N1) corresponding to the onset of energy in the precursor noise segment of the stimulus followed by several transient CPR components (Na, Pb, Nb) occurring after the onset of the pitch-eliciting segment of the stimulus and an offset component (Po) following the offset of the last noise segment in the stimulus. To characterize those attributes of the pitch patterns that are being indexed by the components of the CPR (e.g., pitch onset, pitch acceleration), we evaluated only the latency and magnitude of the CPR components. Peak latencies of pitch-specific response components (Na, Pb, Nb: time interval between pitch-eliciting stimulus onset and response peak of interest) and peak-to-peak amplitude of Na–Pb and Pb–Nb were measured to characterize the effects of changes in pitch acceleration on the components indexing the temporal course of the pitch contours. In addition, peak-to-peak amplitude of Na–Pb and Pb–Nb was measured separately at the frontal (F3/F4) and temporal (T7/T8) electrode sites to evaluate response asymmetry. To enhance visualization of the asymmetry effects along a spectrotemporal dimension, a joint time frequency analysis using a continuous wavelet transform was performed on the grand average waveforms derived from the frontal and temporal electrodes.

Statistical analysis

Separate, two-way, mixed model analysis of variances (ANOVAs) (SAS®; SAS Institute, Inc., Cary, NC, USA) were conducted on peak latency and peak-to-peak amplitude of the CPR components derived from the Fz-linked(T7/T8) electrode site; a three-way, mixed model ANOVA on peak-to-peak amplitude derived from the temporal electrode sites (T7 T8). Group (Chinese, English) functioned as the between-subject factor; subjects nested within group served as a random factor. Stimulus (A1, A2, A3, A4) and temporal site (T7 [left]/T8 [right]) were treated as within-subject factors. At the Fz-linked(T7/T8) electrode site, two-way ANOVAs were used to assess the effects of group and stimulus on each component of peak latency (Na, Pb, Nb) and peak-to-peak amplitude (Na–Pb, Pb–Nb). At the T7/T8 electrode sites, three-way ANOVAs were used to assess the effects of group, stimulus, and temporal site on each component of peak-to-peak amplitude (Na–Pb, Pb–Nb). Post hoc multiple comparisons were corrected with a Bonferroni significance level set at $\alpha = 0.05$. Partial eta-squared (η_p^2) values, where appropriate, were reported to indicate effect sizes.

RESULTS

Response morphology of CPR components

Fig. 2 (top) illustrates that Fz-linked(T7/T8)-derived CPR components (Na, Pb, Nb) of the pitch-eliciting segment (color) are clearly identifiable embedded within the
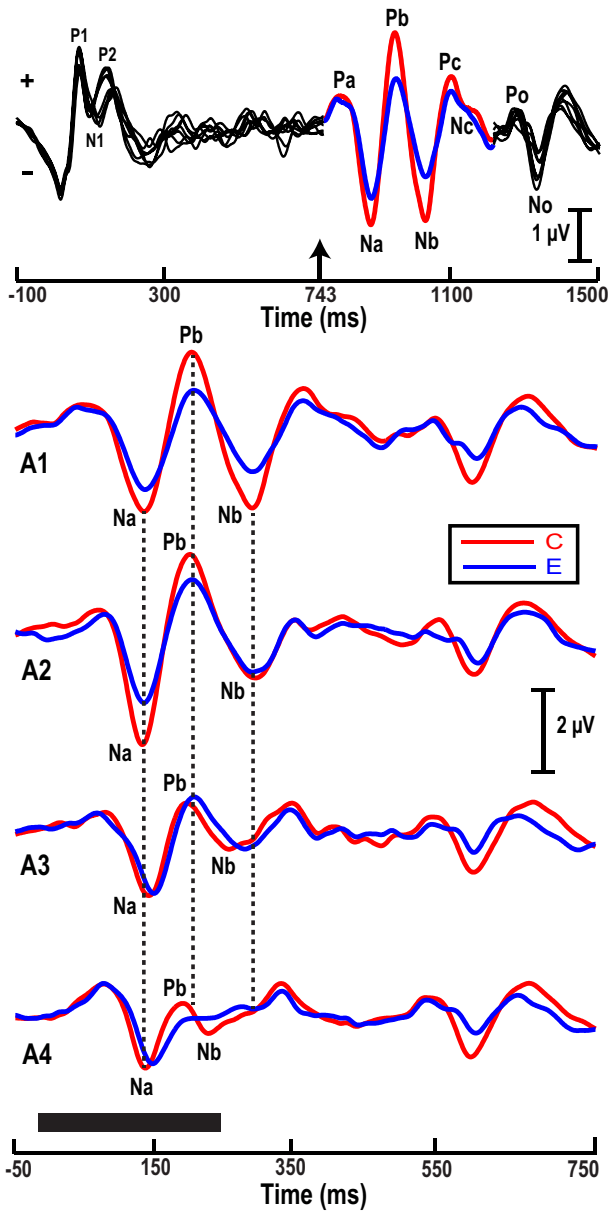
Fig. 2. Grand average waveforms of the Chinese (C) and English (E) groups at the Fz-linked(T7/T8) electrode site per stimulus condition (A1; A2; A3; A4). Na, Pb, and Nb (top: color segment) are the most robust pitch-relevant components. CPR waveforms elicited by the four stimuli (bottom) show that the amplitude of the pitch-relevant components (Na, Pb, Nb) appear to be more robust for the Chinese as compared to the English especially in response to pitch stimuli with slower acceleration rates (A1, A2). Solid black horizontal bar indicates the duration of each stimulus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

stimulus paradigm, which included preceding and following noise segments (black). Fig. 2 (bottom) displays only the time window showing the grand averaged CPR components per stimulus (A1, A2, A3, A4). The amplitude of the pitch-relevant components appears to be more robust for the Chinese group only in response to those stimuli that fall within the normal voice range (A1, A2).

Fz-linked(T7/T8): latency and amplitude of CPR components

Fig. 3 displays mean peak latency of CPR components (Na, Pb, Nb) elicited by each of the four stimuli. As reflected by Na (top panel), the ANOVA yielded a group × stimulus interaction ($F_{3,84} = 2.95$, $p = 0.0374$, $\eta_p^2 = 0.065$). In both language groups, those stimuli that fall outside the normal voice range (A3, A4) elicited longer peak latencies than those that fall inside (A1, A2) (C: A1 vs A3, $t_{84} = -7.32$, < 0.0001; A1 vs A4, $t = -5.12$; A2 vs A3, $t = -6.50$; A2 vs A4, $t = -4.30$, $p = 0.0003$; E: A1 vs A3, $t_{84} = -6.55$, < 0.0001; A1 vs A4, $t = -6.73$; A2 vs A3, $t = -8.26$; A2 vs A4, $t = -8.44$). By stimulus, response peak latencies were generally longer in the English than the Chinese group (A1, $t_{84} = -3.58$, $p = 0.0006$; A3, $t_{84} = -2.98$,
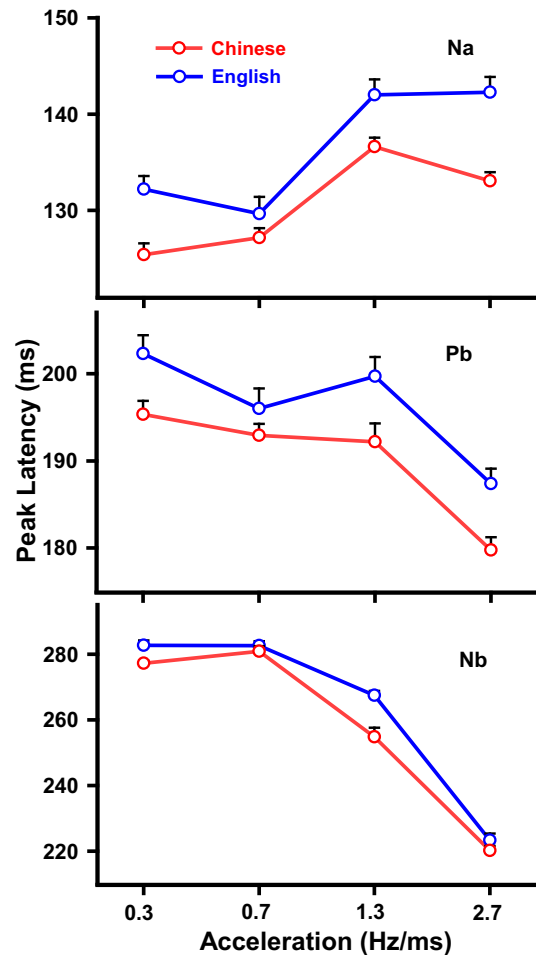

Fig. 3. Mean peak latency CPR components (Na, Pb, Nb) elicited by each of the four stimuli (A1, A2, A3, A4) at the Fz-linked(T7/T8) electrode site in both Chinese and English groups. Na peak latencies of the stimuli with the faster acceleration rates (A3, A4) are longer than those with slower rates (A1, A2) in both groups. Na peak latencies of A1, A3, and A4, however, are longer in English than Chinese. Peak latencies indexed by Pb are longer in English than Chinese across stimuli (A1–A4). Pb peak latencies of A1–A3 are longer than A4 across groups. Nb peak latencies of A1–A3 are longer than A4, and A1–A2 are longer than A3, in both groups. Nb peak latencies of A1 and A3, however, are longer in English than Chinese. A1 (0.3 Hz/ms), A2 (0.7), A3 (1.3), and A4 (2.7). Error bars = ±1 SE.

$p = 0.0038$; A4, $t = -4.84$, $p < 0.0001$). As reflected by Pb (middle panel), the ANOVA yielded main effects of group ($F_{1,28} = 12.36$, $p = 0.0015$, $\eta_p^2 = 0.076$) and stimulus ($F_{3,84} = 35.96$, $p < 0.0001$, $\eta_p^2 = 0.794$). Pooling across stimuli, peak latencies were longer in the English than in the Chinese group ($t_{28} = -3.52$, $p = 0.0015$). As reflected by Nb (bottom panel), the ANOVA yielded an interaction between group and stimulus ($F_{3,84} = 5.93$, $p = 0.0010$, $\eta_p^2 = 0.039$). Simple effects for both language groups showed that the stimulus with the highest acceleration rate (A4) was shorter in latency as compared to the other three (C: A1 vs A4, $t = 26.71$; A2 vs A4, $t = 28.65$; A3 vs A4, $t = 16.00$, $p < 0.0001$; E: A1 vs A4, $t = 27.48$; A2 vs A4, $t = 27.42$; A3 vs A4, $t = 20.44$, $p < 0.0001$). In addition, the other stimulus whose acceleration rate falls outside the normal voice range (A3) was longer in latency as compared to those that fall inside (A1, A2) (C: A1 vs A3, $t_{84} = 10.70$; A2 vs A3, $t = 12.64$; E: A1 vs A3, $t = 7.05$; A2 vs A3, $t = 6.98$, $p < 0.0001$). Simple effects of stimulus showed that the English group's latency was longer than Chinese in response to A1and A3 only (A1, $t = -2.59$, $p = 0.0114$; A3, $t = -6.14$, $p < 0.0001$).

Peak-to-peak amplitude of Na–Pb and Pb–Nb shows effects of both group and stimulus (Fig. 4). Two-way ANOVAs revealed a group × stimulus interaction for both Na–Pb ($F_{3,84} = 5.24$, $p = 0.0023$, $\eta_p^2 = 0.059$) and Pb–Nb ($F_{3,84} = 6.27$, $p = 0.0007$, $\eta_p^2 = 0.183$). For the Na–Pb component (top panel), the Chinese group's
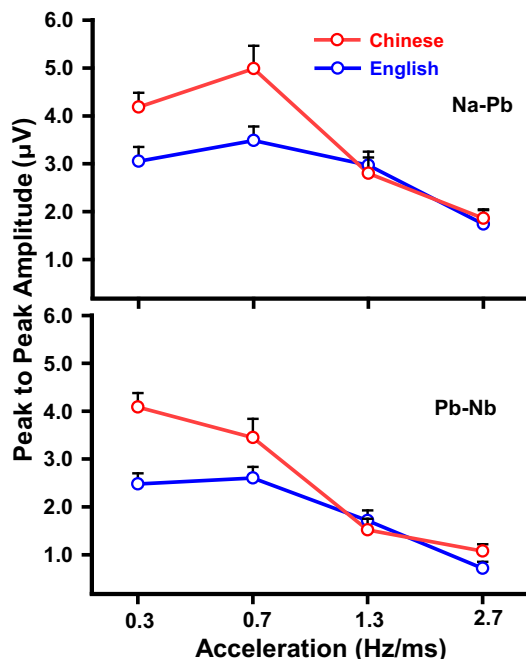
amplitude was larger for those stimuli whose acceleration rates fall within the normal voice range (A1, A2) relative to those whose rates exceed the normal voice range (A1 vs A3: $t_{84} = 3.97$, $p = 0.0009$; A1 vs A4, $t = 6.67$, $p < 0.0001$; A2 vs A3, $t = 6.28$, $p < 0.0001$; A2 vs A4, $t = 8.98$, $p < 0.0001$). In contrast, the English group's amplitude was not only larger for stimuli whose acceleration rates fall within the normal boundary (A1, A2) but also for one whose rate exceeds the normal voice range (A3) as compared to A4 (A1 vs A4: $t = 3.78$, $p = 0.0018$; A2 vs A4, $t = 5.02$, $p < 0.0001$; A3 vs A4, $t_{84} = 3.55$, $p = 0.0039$). Simple effects of stimulus showed that the Chinese group's amplitude was greater than the English in response only to those stimuli with acceleration rates falling within the normal voice range (A1: $t = 2.51$, $p = 0.0139$; A2: $t = 3.33$, $p = 0.0013$).

For the Pb–Nb component (bottom panel), simple effects of group similarly showed that for the Chinese group, those stimuli with acceleration rates falling within the normal voice range were larger in amplitude relative to those whose rates exceed the normal voice range (A1 vs A3, $t_{84} = 8.41$; A1 vs A4, $t = 9.87$; A2 vs A3, $t = 6.32$; A2 vs A4, $t = 7.78$, $p < 0.0001$). In contrast, the English group's amplitude was larger for stimuli whose acceleration rates spanned across the normal voice range boundary (A1, A2, A3) as compared to A4 (A1 vs A4, $t_{84} = 5.78$; A2 vs A4, $t = 6.18$, $p < 0.0001$; A3 vs A4, $t = 3.27$, $p = 0.0092$); A2 was also larger than A3 ($t = 2.91$, $p = 0.0278$). Simple effects of stimulus showed that the Chinese group's amplitude was greater as compared to the English in response only to those stimuli with acceleration rates falling within the normal voice range (A1: $t_{84} = 4.64$, $p < 0.0001$; A2: $t = 2.44$, $p = 0.0167$). In both components, language-dependent effects are likely related to the fact that A1 and A2 are representative of Tone 2 productions within the normal voice speech range for native Chinese speakers.

T7/T8: amplitude of CPR components

Grand average waveforms of the CPR components for each of the four stimuli per language group (left two columns) and their corresponding spectra (right two columns) are displayed in Fig. 5. CPR components in the Chinese group are greater in magnitude (left) and show a robust rightward asymmetry (right) for those stimuli with acceleration rates falling within the normal voice range (A1, A2) and a weaker asymmetry for those whose rates exceed the normal voice range (A3, A4).

T7/T8 peak-to-peak amplitude of Na–Pb (top) and Pb–Nb (bottom) are displayed in Fig. 6. As reflected by Na–Pb, a three-way (group × stimulus × temporal site) ANOVA revealed a group × stimulus interaction ($F_{3,84} = 5.60$, $p = 0.0015$, $\eta_p^2 = 0.167$) and group × temporal site interaction ($F_{1,112} = 16.07$, $p = 0.0001$, $\eta_p^2 = 0.125$). Regarding the group × stimulus interaction, the Chinese group exhibited larger amplitude for those stimuli with acceleration rates falling within the normal voice range (A1, A2) relative to the other two stimuli whose rates exceed the normal voice range (A1 vs A3:

Fig. 4. Mean peak-to-peak amplitude of CPR components (Na–Pb, Pb–Nb) elicited by each of the four stimuli (A1, A2, A3, A4) at the Fz-linked(T7/T8) electrode site in both Chinese and English groups. For both Na–Pb and Pb–Nb, Chinese exhibit greater amplitude than English in response to those pitch contours with acceleration rates (A1, A2) that fall within the bounds of the normal voice pitch range. A1 (0.3 Hz/ms), A2 (0.7), A3 (1.3), and A4 (2.7). Error bars = ±1 SE.

Fig. 5. Grand average waveforms (left) and their corresponding spectra (right) of the CPR components for the two language groups (Chinese, red; English, blue) recorded at electrode sites T7 (dashed) and T8 (solid) for each of the four stimuli (A1, 0.3 Hz/ms; A2, 0.7; A3, 1.3; A4, 2.7). CPR waveforms appear to show a right-sided preference (T8 > T7) for the Chinese group especially in response to pitch stimuli with a slower acceleration rate characteristic of natural speech (A1, A2). The robust rightward preference for A1 and A2 is clearly evident in the spectrotemporal plots. No remarkable asymmetries are apparent in response to A3 and A4, representative of acceleration rates that fall outside the upper bound of maximum speed of pitch change. The zero on the *x*-axis of spectrotemporal plots denotes the time of onset of the pitch-eliciting segment of the four stimuli. Na–Pb and Pb–Nb time windows are demarcated by two vertical, white dashed lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

t_{84} = 4.16, p = 0.0005; A1 vs A4, t = 6.44, p < 0.0001; A2 vs A3, t = 5.94, p < 0.0001; A2 vs A4, t = 8.21, p < 0.0001). In contrast, the English group had greater amplitude for those stimuli with acceleration rates falling within the normal voice range (A1, A2) when compared to the stimulus with the highest acceleration rate only (A1 vs A4: t = 3.24, p = 0.0102; A2 vs A4: t = 3.82, p = 0.0015). Simple effects of stimulus showed that amplitude of the Chinese group was greater than that of the English only in response to those stimuli with acceleration rates falling within the normal voice range (A1: t = 2.47, p = 0.0155; A2: t = 3.37, p = 0.0011). Regarding the group × temporal site interaction, post hoc comparisons at each level of group showed an advantage at the right temporal electrode site (T8 > T7) for the Chinese group only (t_{112} = −7.41, p < 0.0001). At each level of temporal site, the Chinese group's larger Na–Pb amplitude relative to that of the English was restricted to the right temporal electrode site (t = 3.42, p = 0.0009).

As reflected by Pb–Nb, a three-way (group × stimulus × temporal site) ANOVA revealed a main effect of temporal site ($F_{1,112}$ = 52.10, p < 0.0001, η_p^2 = 0.317) and group × stimulus interaction ($F_{3,84}$ = 6.47, p = 0.0005, η_p^2 = 0.188). The main effect revealed a strong asymmetry for the right temporal electrode site irrespective of group or stimulus. In the case of the group × stimulus interaction, the Chinese group had larger amplitude for those stimuli with acceleration rates falling within the normal voice range (A1, A2) relative to the other two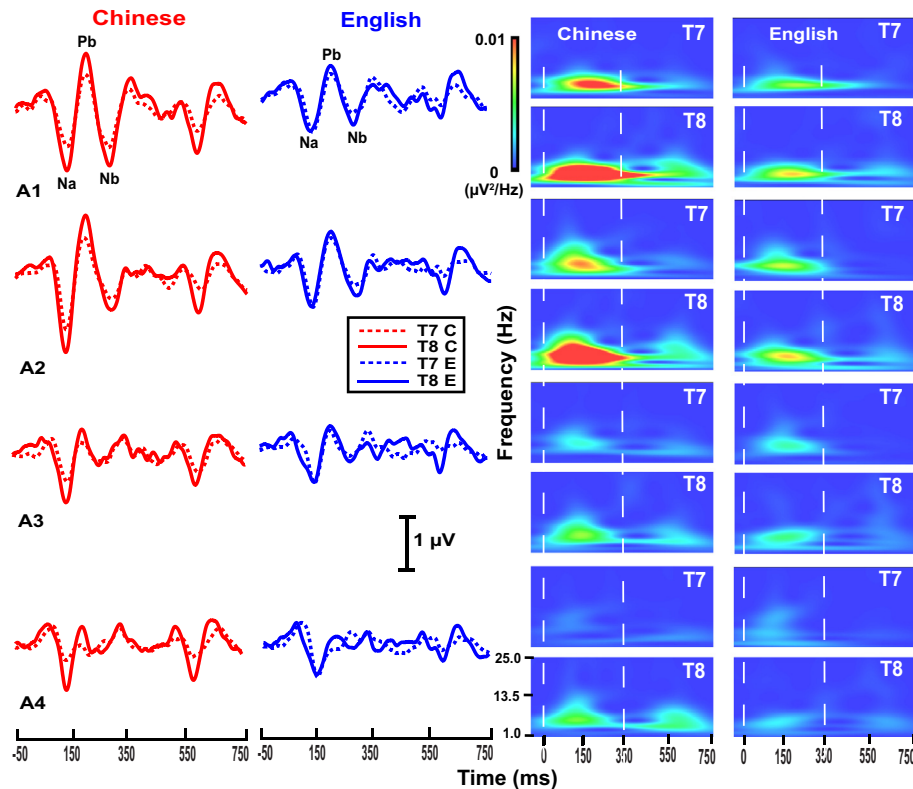 stimuli whose rates exceed the normal voice range (A1 vs A3, t_{84} = 8.24; A1 vs A4, t = 9.02; A2 vs A3, t = 5.94; A2 vs A4, t = 6.71, p < 0.0001). For English, on the other hand, stimuli A1, A2, and A3 were larger in amplitude than A4 (A1, t = 5.23; A2, t = 5.33, p < 0.0001; A3, t = 3.06, p = 0.0179). Simple effects of stimulus showed that amplitude of the Chinese group was greater than that of the English in response only to those stimuli with acceleration rates falling within the normal voice range (A1: t = 4.50, p < 0.0001; A2: t = 2.40, p = 0.0185).

DISCUSSION

The major findings of this study demonstrate that pitch-relevant neural activity as reflected in the latency and amplitude of scalp-recorded CPR components show distinct changes that are specific to language-experience (Chinese vs English); and to changes in the acoustic attribute (pitch acceleration) of dynamic, time
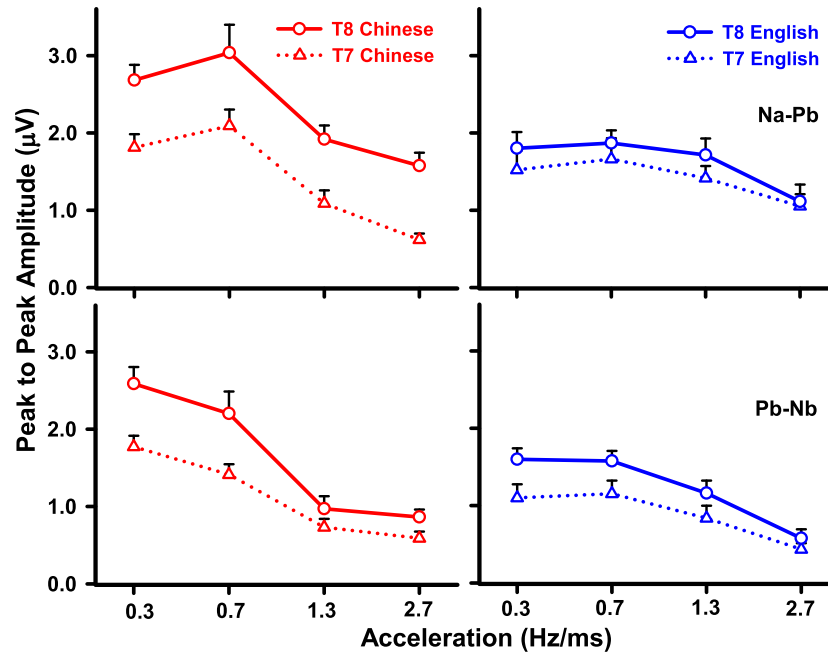
Fig. 6. Mean peak-to-peak amplitude of CPR components (Na–Pb, top row; Pb–Nb, bottom row) extracted from T7/T8 in the temporal lobe as a function of language group (Chinese, red; English, blue), stimulus (A1, 0.3 Hz/ms; A2, 0.7; A3, 1.3; A4, 2.7), and temporal site (left, diagonal; right, solid). Both Na–Pb and Pb–Nb amplitude show a language-dependent effect (C > E) over the right temporal site (T8) elicited by A1 and A2, the two pitch stimuli with native-like acceleration rates. Error bars = ±1 SE. C, Chinese; E, English. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

varying pitch contours. Peak latencies of Na, Pb, and to a lesser extent, Nb, are shorter in the Chinese across stimuli. Chinese show greater amplitude for Na–Pb and Pb–Nb at both Fz-linked(T7/T8) and temporal (T7/T8) electrode sites in response to pitch contours (A1, A2) with acceleration rates that fall within the normal voice pitch range as compared to pitch contours (A3, A4) with acceleration rates that exceed the normal range. As indexed by Na–Pb and Pb–Nb amplitude at the temporal sites, a rightward asymmetry is observed for the Chinese group only. Moreover, it is only over the right temporal site that Chinese amplitude is greater than English. These findings suggest that the neural mechanism(s) underlying processing of pitch in the right auditory cortex reflect experience-dependent modulation of sensitivity to acceleration in just those rising pitch contours that fall within the bounds of one's native language.

CPR changes reflect sensory and extrasensory processes underlying pitch representation

Changes in the latency and amplitude of CPR components across stimuli (A1, A2, A3, A4) and between groups (C, E) allow us to distinguish between *experience-independent* (sensory) and *experience-dependent* processes (extrasensory) that modulate sensitivity to changes in pitch acceleration of time-variant pitch contours in the auditory cortex.

Response latency (Na, Pb, Nb). Na reflects neural activity synchronized to pitch onset. It presumably represents the integration of pitch information across

frequency channels and/or the calculation of the initial pitch value and pitch strength in Heschl's gyrus (HG) (Gutschalk et al., 2004). Source analyses (Gutschalk et al., 2002, 2004; Krumbholz et al., 2003), and EEG pitch responses (Bidelman and Grall, 2014), corroborated by human depth-electrode recordings (Schonwiesner and Zatorre, 2008; Griffiths et al., 2010) indicate that Na is localized to the anterolateral portion of HG, the putative site of pitch processing (Zatorre, 1988; Griffiths et al., 1998; Johnsrude et al., 2000; Patterson et al., 2002; Penagos et al., 2004; Bendor and Wang, 2005). Na's longer latency at higher pitch acceleration rates (A3, A4) for both groups may reflect frequency-specific neural adaptation, spread of this adaptation across frequency, and recovery from this adaptation that disrupts both synaptic efficiency and neural synchrony (Herrmann et al., 2013). Because Na reflects integration of pitch information across frequency channels, it is also possible that the longer latency at the higher pitch acceleration rates reflects longer integration times for the initial pitch estimate due to desynchronizing of neural activity at faster pitch accelerations.

In fact, the latency of the pitch onset response (Na) has been shown to be differentially sensitive to several acoustical attributes of the pitch-eliciting stimulus. For example, Na's sensitivity is indicated by its longer latency for stimuli (i) with weaker, relative to stronger, pitch salience (Gutschalk et al., 2002, 2004; Krumbholz et al., 2003; Krishnan et al., 2012a); (ii) with a rapidly rising portion of the pitch contour that occurs early, relative to late, after stimulus onset (Krishnan et al., 2015b); (iii) with a falling compared to rising pitch contour (Krishnan et al., 2015b); and (iv) with linear ramps or steady-state

pitch compared to curvilinear T2 (Krishnan et al., 2014b). In this study, a similar pattern of Na latency change is observed with increasing pitch acceleration in both groups. These data support the notion of shared properties of an underlying, language-universal pitch mechanism. However, Na latency evoked by A1, A3, and A4 is shorter in Chinese than English, suggesting relatively shorter integration time to estimate pitch acceleration. This language group difference leads us to posit that experience-dependent, extrasensory modulation of sensitivity to pitch acceleration is overlaid on the shared sensory effects.

In contrast to Na, the latency of Pb and Nb evoked by A4, the stimulus with the highest acceleration rate, is shorter than A3 regardless of group. This finding may reflect an adaptive increase in sampling rate to optimally capture the increasingly rapid changes in pitch. This finding is consistent with previous work that suggests Pb and Nb may be indexing pitch-relevant neural activity associated with more rapidly-changing portions of the pitch contour (Krishnan et al., 2014a, 2015a). This inference is further supported by a strong correlation between pitch acceleration and changes in Pb and Nb (Krishnan et al., 2014a). In the case of Pb, the shorter latency for Chinese as compared to English across stimuli reflects experience-dependent modulation of sensitivity to pitch acceleration. Pitch mechanisms for Chinese may be better able to adjust the pitch integration time scales to achieve optimal representation of rapidly changing pitch. That is, latency may be reflecting the dynamics of the temporal windows being utilized to process the various temporal attributes of pitch.

Response amplitude (Na–Pb, Pb–Nb). Fz-linked (T7/T8) amplitude of CPR components (Na–Pb, Pb–Nb) also reflect both sensory and extrasensory modulation of sensitivity to changes in pitch acceleration. Both groups show nearly identical amplitude decrements in response to the two stimuli with higher pitch acceleration (A3, A4) relative to those with lower pitch acceleration (A1, A2). This reduced sensitivity to rapid pitch acceleration presumably reflects degradation of neural activity associated with rapidly changing pitch segments (e.g., neural desynchronization, synaptic inefficiency). This finding agrees with previous observations of amplitude reduction for these components at higher pitch acceleration rates regardless of language group (Krishnan et al., 2014a, 2015a). We infer that reduced sensitivity to pitch acceleration rates that exceed the normal speech range is a language-universal property of early sensory level processing of dynamic pitch in the auditory cortex, similar to these components' differential sensitivity to location of the acceleration peak (early vs late) (Krishnan et al., 2015b).

Yet we also observe a language-dependent enhancement of Fz-linked(T7/T8) amplitude of Na–Pb and Pb–Nb evoked by pitch contours with acceleration rates falling within the Mandarin voice pitch range (A1, A2). For the Chinese group, those stimuli that exhibit native acceleration rates evoke larger amplitude as compared to those that fall outside the normal voice pitch range (A3, A4). The English group, in contrast, fails to show a selectivity to native acceleration rates (A1, A2, A3 > A4). These findings point to experience-dependent enhancement of pitch attributes that varies depending on their functional roles in a particular language. More broadly, they are also consistent with earlier studies that have revealed experience-dependent neural plasticity in pitch processing at both cortical and subcortical levels of the brain (for reviews, Zatorre and Gandour, 2008; Zatorre and Baum, 2012; Krishnan et al., 2012b; Gandour and Krishnan, 2014). In this study, we infer that extrasensory processes are overlaid on sensory processes to modulate long-term, experience-driven, adaptive pitch mechanisms at early sensory levels of pitch processing in the auditory cortex. This is accomplished by sharpening response properties of neural elements to enable optimal representation of temporal attributes of pitch contours that are behaviorally relevant. That is, amplitude may be reflecting the robustness of the underlying pitch-relevant neural activity within the temporal windows being utilized to process the various temporal attributes of pitch.

As indexed by Na–Pb and Pb–Nb amplitude over the temporal electrode sites (T7/T8), a rightward asymmetry is limited to the Chinese group. It is only over the right temporal site that the amplitude of these components are larger for the Chinese compared to English. The observed electrophysiological responses are putatively specific to pitch. Stimuli are reduced to the pitch parameter only. The experimental paradigm is free of task demands. Our findings converge with an extant literature that supports the role of the right hemisphere in processing linguistic as well as nonlinguistic pitch (Zatorre et al., 2002; Friederici and Alter, 2004; Hyde et al., 2008; Meyer, 2008; Zatorre and Gandour, 2008; Friederici, 2011). This experience-dependent effect demonstrates that extrasensory components may predominate over sensory components in their influence within a given temporal integration window or, in other words, mask purely sensory effects. Purely sensory effects should produce a similar pattern of change for both groups as observed for A3 and A4. The English pattern may be attributed simply to differences in auditory sensitivity to pitch acceleration (A1, A2, A3 > A4). The Chinese pattern is different (A1, A2 > A3, A4). Its sensitivity to pitch acceleration is modulated by extrasensory factors. Specifically, it separates those stimuli representative of T2 with acceleration rates in the normal voice pitch range (A1, A2) from T2 variants with acceleration rates exceeding the normal range (A3, A4).

As indexed by Na–Pb and Nb–Pb amplitude over the frontal electrode sites (F3/F4), the response patterns of Chinese and English across stimuli mirror those derived from the Fz-linked(T7/T8) and T7/T8 sites. The Chinese pattern (A1, A2 > A3, A4) reflects the influence of extrasensory factors, whereas the English pattern does not (A1, A2, A3 > A4). Chinese amplitude is greater than English for just those pitch contours that fall within the normal voice pitch range (A1, A2). The absence of a group × frontal site interaction over F3/F4, however,

stands in stark contrast to T7/T8. Over the temporal sites, Chinese exhibit a rightward asymmetry; they also have larger amplitude than English. Our failure to observe language-dependent, asymmetry at the F3/F4 electrode sites is consistent with our previous findings (Krishnan et al., 2015a,b). This suggests that the temporal electrodes proximal to and located over the auditory cortices, the putative regions for pitch processing, are better situated to capture the experience-dependent, preferential recruitment of the right auditory cortex for pitch processing.

Thus, CPR components may capture both experience-dependent extrasensory influences as well experience-independent sensory effects (Krishnan et al., 2015a,b). By extrasensory, we mean neural processes at a higher hierarchical level beyond the purely sensory processing of acoustic attributes of the stimulus. One likely candidate for fine-grained stored representations of pitch attributes at this early sensory cortical level of processing is *analyzed sensory memory* (Cowan, 1984, 1987, cf. Xu et al., 2006). This memory store is to be distinguished from the initial, *sensory memory trace* and later cognitive processes with their associated *short-* and *long-term memory* stores. It contains fine-grained, analyzed sensory codes including information about pitch height, time-varying pitch direction and acceleration, and event timing of pitch onset and offset. Its lifetime is on the order of seconds. In this four-store model of memory, information is encoded in a hierarchical order but short-term memory and analyzed sensory memory can be processed in parallel. Because our experimental task does not require attention, short-term or long-term memory, or decision-making, it is unnecessary to invoke short-term memory. But any language-dependent effect is also a memory effect. As soon as a sensory signal makes contact with a stored representation, there is an interesting interaction based on the nature of that representation. In our case, that stored representation resides in analyzed sensory memory. It cannot be attributed simply to the sensory memory trace. Otherwise we would expect homogeneous effects across language groups.

With respect to influence of memory in shaping processing, Hasson et al. (2015) propose a fundamentally different framework for information processing in cortical neural circuits that integrates past information (activated in neural circuits) with incoming information. They utilize a biologically motivated *process memory* framework, in contrast to traditional encapsulated stored memory, to explain information processing in cortical neural circuits. *Process memory* refers to the integration of active traces of past information that are used by a neural circuit to process incoming information in the present moment. To influence ongoing processing, the prior information must be in an active state, i.e., neural activity composed of the stimulus information accumulated in a given circuit throughout the event. In our case, that would include analyzed sensory memory and/or long-term representation(s) of the stimulus that is activated during processing of the incoming information. Active refers to either sustained elevation in firing rates or information sustained in a neural circuit by short-term calcium mediated synaptic facilitation in the absence of recurrent activity (Mongillo et al.,

2008). The integration time scales, i.e., temporal receptive windows in which prior information from an ongoing stimulus can influence the processing of newly arriving information vary in a hierarchical fashion across the cerebral cortex, with shorter (milliseconds to seconds) timescales in sensory regions and a gradient of lengthening timescales (seconds to minutes) in higher order cortices. Our CPR responses would be activated in the early stages of this processing memory hierarchy and utilize short temporal receptive windows where the neural dynamics are more rapid.

Effects of acoustic properties at different hierarchical levels of pitch processing

Collectively, our findings suggest that fundamental neural mechanisms of pitch at early stages in the auditory cortex are the same for Chinese and English listeners alike, but that Chinese are more sensitive to perceptually relevant pitch attributes by virtue of their long-term experience with a tonal language. Because enhanced sensitivity to pitch acceleration is already present in neural activity at the level of the brainstem (Krishnan et al., 2010b), it seems plausible that cortical pitch mechanisms may be reflecting, at least in part, this enhanced pitch input from the brainstem (cf. Bidelman et al., 2014). However, stimulus effects derived from the CPR components differ from those observed at the level of the brainstem. Using the same acceleration rates as those in this study, Chinese showed stronger pitch representation than English across all four stimuli (Krishnan et al., 2010b). These findings suggest that at the brainstem level, perceptually salient pitch cues associated with lexical tone influence pitch extraction not only in the normal voice pitch range, but also in auditory signals that clearly fall outside the range of dynamic pitch that a native listener is exposed to. At the cortical level, the restriction of experience-dependent effects to A1 and A2 likely represents a transformation in pitch processing characterized by enhanced selectivity to linguistically relevant pitch contours.

Predictive coding may underlie experience-dependent processing of pitch

Growing evidence shows pitch-related neural activity in both the primary auditory cortex in the medial HG as well as in the adjacent more lateral non-primary areas in the HG, suggesting that pitch-relevant information is available in multiple areas of the auditory cortex: functional imaging plus direct cortical recording (Patterson et al., 2002; Penagos et al., 2004; Griffiths et al., 2010; Puschmann et al., 2010), patients with focal excisions (Zatorre, 1988; Zatorre and Samson, 1991; Johnsrude et al., 2000), and magnetoencephalography (Gutschalk et al., 2002, 2004; Krumbholz et al., 2003). Lateral HG also appears to be important for computations relevant to extraction of pitch of complex sounds (Zatorre and Belin, 2001; Hall et al., 2002; Schonwiesner et al., 2005).

A hierarchical processing framework for coordinated interaction between these areas is provided by

application of predictive coding model of perception to depth-electrode recordings of pitch-relevant neural activity along HG (Rao and Ballard, 1999; Kumar et al., 2011; Kumar and Schonwiesner, 2012). Essentially, higher level areas in the hierarchy contributing to pitch (lateral HG) use stored information of pitch to make a pitch prediction. This prediction is passed to the lower areas in the processing hierarchy (medial and middle HG) via top-down connection(s). The lower areas then compute a prediction error. The strength of the top-down and bottom-up connections is continually adjusted in a recursive manner in order to minimize predictive error and to optimize representation at the higher level. Consistent with the predictions of the model, Kumar et al. (2011) showed that strength of connectivity varies with pitch salience such that the strength of the top down connection from lateral HG to medial and middle HG increased with pitch salience, whereas the strength of the bottom-up connection from middle HG to lateral HG decreased. It is likely that lateral HG has more pitch-specific mechanisms, and therefore plays a relatively great role in pitch perception.

Applied to our data, this framework suggests that CPR changes attributable wholly to acoustic properties of the stimulus invoke a recursive process in the representation of pitch (initial pitch prediction, error generation, error correction). At this level, the hierarchical flow of processing and its connectivity strengths along the HG are essentially the same regardless of one's language background. However, the initial pitch prediction at the level of the lateral HG is more precise for Chinese because of their access to stored information about native pitch contours (A1 and A2) with a smaller error term. Consequently, the top-down connection from lateral HG to medial and middle HG is stronger than the bottom-up connection. The opposite would be true for English because of their less precise initial prediction. Language experience therefore alters the nature of the interaction between levels along the hierarchy of pitch processing by modulating connection strengths. The hierarchical processing memory framework is broadly consistent with predictive coding.

Pitch processing in the auditory cortex is influenced by inputs from subcortical structures that are themselves subject to experience-dependent plasticity. It is likely that top-down connections in the hierarchy provide feedback to adjust the effective time scales of processing at each stage to optimally control the temporal dynamics of pitch processing. Language-dependent changes in the CPR by Chinese may reflect interplay between sensory and extrasensory processing. This expanded model represents a unified, physiologically plausible, theoretical framework that includes both cortical and subcortical components in the hierarchical processing of pitch.

CONCLUSIONS

The strong sensitivity to pitch acceleration, reflected by latency and amplitude of CPR components Na–Pb and Pb–Nb, enables us to evaluate differential sensitivity to language-universal (acoustic) and overlaid language-dependent (linguistic) temporal attributes of pitch processing during early sensory level processing in the auditory cortex. Enhancement of native pitch stimuli and stronger rightward asymmetry of CPR components in the Chinese group is consistent with the notion that long-term experience shapes adaptive, distributed hierarchical pitch processing in the auditory cortex, and reflects an interaction with higher order, extrasensory processes beyond the sensory memory trace. Within a given temporal integration window, pitch processing involves a hierarchy of both sensory and extrasensory effects whose relative weighting varies depending on language experience.

REFERENCES

Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. Nature 436:1161–1165.

Bidelman GM, Grall J (2014) Functional organization for musical consonance and tonal pitch hierarchy in human auditory cortex. Neuroimage 101:204–214.

Bidelman GM, Weiss MW, Moreno S, Alain C (2014) Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. Eur J Neurosci 40:2662–2673.

Cowan N (1984) On short and long auditory stores. Psychol Bull 96:341–370.

Cowan N (1987) Auditory sensory storage in relation to the growth of sensation and acoustic information extraction. J Exp Psychol Hum Percept Perform 13:204–215.

Friederici AD (2011) The brain basis of language processing: from structure to function. Physiol Rev 91:1357–1392.

Friederici AD, Alter K (2004) Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang 89:267–276.

Gandour JT (1983) Tone perception in Far Eastern languages. J Phonetics 11:149–175.

Gandour JT, Harshman RA (1978) Cross language differences in tone perception: a multidimensional scaling investigation. Lang Speech 21:1–33.

Gandour JT, Krishnan A (2014) Neural bases of lexical tone. In: Winskel H, Padakannaya P, editors. Handbook of South and Southeast Asian psycholinguistics. Cambridge, UK: Cambridge University Press. p. 339–349.

Griffiths TD, Buchel C, Frackowiak RS, Patterson RD (1998) Analysis of temporal structure in sound by the human brain. Nat Neurosci 1:422–427.

Griffiths TD, Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, Brugge JF, Howard MA (2010) Direct recordings of pitch responses from human auditory cortex. Curr Biol 20:1128–1132.

Gutschalk A, Patterson RD, Rupp A, Uppenkamp S, Scherg M (2002) Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. Neuroimage 15:207–216.

Gutschalk A, Patterson RD, Scherg M, Uppenkamp S, Rupp A (2004) Temporal dynamics of pitch in human auditory cortex. Neuroimage 22:755–766.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. Cereb Cortex 12:140–149.

Hasson U, Chen J, Honey CJ (2015) Hierarchical process memory: memory as an integral component of information processing. Trends Cogn Sci 19:304–313.

Herrmann B, Henry MJ, Obleser J (2013) Frequency-specific adaptation in human auditory cortex depends on the spectral variance in the acoustic stimulation. J Neurophysiol 109:2086–2096.

Hyde KL, Peretz I, Zatorre RJ (2008) Evidence for the role of the right auditory cortex in fine pitch resolution. Neuropsychologia 46:632–639.

Johnson KL, Nicol TG, Kraus N (2005) Brain stem response to speech: a biological marker of auditory processing. Ear Hear 26:424–434.

Johnsrude IS, Penhune VB, Zatorre RJ (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 123:155–163.

Klatt D (1973) Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. J Acoust Soc Am 53:8–16.

Kraus N, Banai K (2007) Auditory-processing malleability: focus on language and music. Curr Dir Psychol Sci 16:105–110.

Krishnan A, Gandour JT (2009) The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang 110:135–148.

Krishnan A, Gandour JT (2014) Language experience shapes processing of pitch relevant information in the human brainstem and auditory cortex: electrophysiological evidence. Acoust Aust 42:166–178.

Krishnan A, Swaminathan J, Gandour JT (2009) Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. J Cogn Neurosci 21:1092–1105.

Krishnan A, Gandour JT, Bidelman GM (2010a) Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. Brain Res 1313:124–133.

Krishnan A, Gandour JT, Smalt CJ, Bidelman GM (2010b) Language-dependent pitch encoding advantage in the brainstem is not limited to acceleration rates that occur in natural speech. Brain Lang 114:193–198.

Krishnan A, Bidelman GM, Smalt CJ, Ananthakrishnan S, Gandour JT (2012a) Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. Neuropsychologia 50:2849–2859.

Krishnan A, Gandour JT, Bidelman GM (2012b) Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. Neuroreport 23:498–502.

Krishnan A, Gandour JT, Ananthakrishnan S, Vijayaraghavan V (2014a) Cortical pitch response components index stimulus onset/offset and dynamic features of pitch contours. Neuropsychologia 59:1–12.

Krishnan A, Gandour JT, Suresh CH (2014b) Cortical pitch response components show differential sensitivity to native and nonnative pitch contours. Brain Lang 138:51–60.

Krishnan A, Gandour JT, Ananthakrishnan S, Vijayaraghavan V (2015a) Language experience enhances early cortical pitch-dependent responses. J Neurolinguistics 33:128–148.

Krishnan A, Gandour JT, Suresh CH (2015b) Pitch processing of dynamic lexical tones in the auditory cortex is influenced by sensory and extrasensory processes. Eur J Neurosci 41:1496–1504.

Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lutkenhoner B (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb Cortex 13:765–772.

Kumar S, Schonwiesner M (2012) Mapping human pitch representation in a distributed system using depth-electrode recordings and modeling. J Neurosci 32:13348–13351.

Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, Howard 3rd MA, Friston KJ, Griffiths TD (2011) Predictive coding and pitch processing in the auditory cortex. J Cogn Neurosci 23:3084–3094.

Laver J (1994) Principles of phonetics. New York, NY: Cambridge University Press.

Li P, Sepanski S, Zhao X (2006) Language history questionnaire: a web-based interface for bilingual research. Behav Res Methods 38:202–210.

Liu C (2013) Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners. J Acoust Soc Am 134:3011–3020.

Meyer M (2008) Functions of the left and right posterior temporal lobes during segmental and suprasegmental speech perception. Z Neuropsychol 19:101–115.

Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. Science 319:1543–1546.

Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113.

Patel AD, Iversen JR (2007) The linguistic benefits of musical abilities. Trends Cogn Sci 11:369–372.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776.

Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J Neurosci 24:6810–6815.

Puschmann S, Uppenkamp S, Kollmeier B, Thiel CM (2010) Dichotic pitch activates pitch processing centre in Heschl's gyrus. Neuroimage 49:1641–1649.

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87.

Schonwiesner M, Zatorre RJ (2008) Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. Exp Brain Res 187:97–105.

Schonwiesner M, Rubsamen R, von Cramon DY (2005) Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur J Neurosci 22:1521–1528.

Swaminathan J, Krishnan A, Gandour JT, Xu Y (2008) Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. IEEE Trans Biomed Eng 55:281–287.

Tzounopoulos T, Kraus N (2009) Learning to encode timing: mechanisms of plasticity in the auditory brainstem. Neuron 62:463–469.

Whalen DH, Xu Y (1992) Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica 49:25–47.

Wong PC, Perrachione TK (2007) Learning pitch patterns in lexical identification by native English-speaking adults. Appl Psycholinguist 28:565–585.

Xu Y (1997) Contextual tonal variations in Mandarin. J Phonetics 25:61–83.

Xu Y, Sun X (2002) Maximum speed of pitch change and how it may relate to speech. J Acoust Soc Am 111:1399–1413.

Xu Y, Gandour JT, Francis AL (2006) Effects of language experience and stimulus complexity on the categorical perception of pitch direction. J Acoust Soc Am 120:1063–1074.

Zatorre RJ (1988) Pitch perception of complex tones and human temporal-lobe function. J Acoust Soc Am 84:566–572.

Zatorre RJ, Baum SR (2012) Musical melody and speech intonation: singing a different tune. PLoS Biol 10:e1001372.

Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946–953.

Zatorre RJ, Gandour JT (2008) Neural specializations for speech and pitch: moving beyond the dichotomies. Philos Trans R Soc Lond B Biol Sci 363:1087–1104.

Zatorre RJ, Samson S (1991) Role of the right temporal neocortex in retention of pitch in auditory short-term memory. Brain 114(Pt 6):2403–2417.

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends in cognitive sciences 6:37–46.