COGNITIVE NEUROSCIENCE

Pitch processing of dynamic lexical tones in the auditory cortex is influenced by sensory and extrasensory processes

Ananthanarayan Krishnan, Jackson T. Gandour and Chandan H. Suresh

Department of Speech Language Hearing Sciences, Purdue University, Lyles-Porter Hall, 715 Clinic Drive, West Lafayette, IN 47907, USA

Abstract

The aim is to evaluate how language experience (Chinese, English) shapes processing of pitch contours as reflected in the amplitude of cortical pitch response components. Responses were elicited from three dynamic curvilinear nonspeech stimuli varying in pitch direction and location of peak acceleration: Mandarin lexical Tone 2 (rising) and Tone 4 (falling), and a flipped variant of Tone 2, Tone 2′ (nonnative). At temporal sites (T7/T8), Chinese listeners' Na–Pb response amplitudes to Tones 2 and 4 were greater than those of English listeners in the right hemisphere only; a rightward asymmetry for Tones 2 and 4 was restricted to the Chinese group. In common to both Fz-to-linked T7/T8 and T7/T8 electrode sites, the stimulus pattern (Tones 2 and 4 > Tone 2′) was found in the Chinese group only. As reflected by Pb–Nb at Fz, Chinese subjects' amplitudes were larger than those of English subjects in response to Tones 2 and 4, and Tones 2 and 4 were larger than Tone 2′, whereas for English subjects, Tone 2 was larger than Tone 2′ and Tone 4. At frontal electrode sites (F3/F4), regardless of component or hemisphere, Chinese subjects' responses were larger in amplitude than those of English subjects across stimuli. For either group, responses to Tones 2 and 4 were larger than Tone 2′. No hemispheric asymmetry was observed at the frontal electrode sites. These findings demonstrate that cortical pitch response components are differentially modulated by experience-dependent, temporally distinct but functionally overlapping, weighting of sensory and extrasensory effects on pitch processing of lexical tones in the right temporal lobe and, more broadly, are consistent with a distributed hierarchical predictive coding process.

Introduction

Pitch processing is shaped by one's experience with language and music at the level of the auditory brainstem as well as the cerebral cortex (Patel, 2008; Alho *et al.*, 2012; Itoh *et al.*, 2012; Koelsch, 2012; Krishnan *et al.*, 2012). In tone languages, the primary auditory correlate of lexical tone is based on variations in pitch. As measured by the early, preattentive, mismatch negativity (MMN), Mandarin tones, relative to consonants, are lateralised to the right hemisphere (RH) (Luo *et al.*, 2006). Using sinusoidal tones, MMN has been shown to be comprised of temporally distinct auditory and cognitive mechanisms of frequency change detection in auditory cortex (Maess *et al.*, 2007). However, MMN itself is not a pitch-specific response.

Tone languages provide a physiological window for evaluating how neural representations of linguistic pitch emerge during early sensory processing. We have yet to achieve a precise characterisation of neural representation of specific attributes of dynamic pitch contours. Given that parallel processing of neurophysiological

indicators of psycholinguistic information occurs with near-simultaneity in the first 200–250 ms (Pulvermuller *et al.*, 2009), an early cortical pitch-specific response (CPR) is necessary to tease apart sensory and extrasensory influences on pitch.

Most of the previous studies measured cortical responses that were prominently obligatory responses to sound onset and not pitch-specific (Gutschalk *et al.*, 2004; Lutkenhoner *et al.*, 2006; Yrttiaho *et al.*, 2008). Pitch stimuli were steady-state, non-occurring in natural speech. Only the onset component (N100) was measured, a response to the onset of sound energy and not exclusively to pitch. The CPR, on the other hand, is characterised by multiple transient components that index different temporal attributes of dynamic pitch contours. To disentangle the pitch-specific from the obligatory onset response, a novel stimulus paradigm was constructed with two segments: an initial segment of noise with no pitch to evoke the onset components only, followed by a pitch-eliciting segment of iterated rippled noise (IRN) (Krumbholz *et al.*, 2003). Adapting this stimulus paradigm, the magnitude of CPR transient components elicited by dynamic pitch homologs of Mandarin Tone 2 (T2, rising) were larger in Chinese than English listeners (Krishnan *et al.*, 2015). By employing dynamic curvilinear pitch stimuli representing both native and nonnative contours we were able to tease apart sensory and extrasensory influences on experience-dependent pitch processing.

Correspondence: Jackson T. Gandour, as above.
E-mail: gandour@purdue.edu

Herein we examined a set of three dynamic, curvilinear, non-speech pitch stimuli that differ in pitch direction and location of peak acceleration. Two are homologous to Mandarin T2 and Tone 4 (T4; falling); the third is a flipped variant of T2 (T2′). We evaluated the effects of experience-dependent (Chinese, English) effects on latency and amplitude of CPR components. Stimulus comparisons allowed us to assess acoustic effects of location of peak acceleration and pitch direction. Stimulus comparisons at frontal and temporal electrode sites allowed us to assess hemispheric asymmetry. We hypothesised 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.

Materials and methods

Participants

Fourteen native speakers of Mandarin Chinese (seven male, seven female) and English (ten male, four female) were recruited from the Purdue University student body to participate in the experiment. All exhibited normal hearing sensitivity at audiometric frequencies between 500 and 4000 Hz and reported no previous history of neurological or psychiatric illnesses. They were closely matched in age (Chinese, 24.14 ± 3.28 years; English, 22.36 ± 1.08), years of formal education (Chinese, 16.86 ± 2.11 years; English, 16.18 ± 1.44), and were strongly right handed (Chinese, 92.2 ± 11.8%; English, 95.8 ± 9.2%) as measured by the laterality index of the Edinburgh Handedness Inventory (Oldfield, 1971). 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). As determined by a music history questionnaire (Wong & Perrachione, 2007), all Chinese and English participants had < 2 years of musical training (Chinese, 0.64 ± 0.82 years; English, 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

Three nonspeech stimuli were constructed to investigate CPR responses to curvilinear, time-varying pitch that differed in changes of direction (Fig. 1, top panel, right) and rates of acceleration (Fig. 1, bottom panel). Two of the stimuli represented lexical tones of Mandarin, modeled after productions of citation forms on isolated monosyllables of T2 (rising) and T4 (falling) (Howie, 1976; Moore & Jongman, 1997; Xu, 1997). T2 and T4 shared an average F_0 of 111 Hz and pitch acceleration rate trajectory that reached its peak at 70% (175.69 ms) of total duration. They differed in the direction of pitch during those portions characterised by larger changes in F_0 that are known to contribute importantly to tonal recognition (Whalen & Xu, 1992). The third stimulus (T2′, rising) does not exist in citation form in the Mandarin tonal space. Though T2′ represented a flipped variant of T2, it shared F_0 onset/offset (103/131 Hz) in common with T2 as well as direction of pitch change (rising); its average F_0 was 123 Hz. Despite differences in pitch direction, T2 and T4 shared the same acceleration trajectory throughout their duration. Both were 180° out of phase with T2′. T2 and T4 reached a late peak of acceleration at 70% of duration; T2′, in contrast, had an early peak at ~ 10%
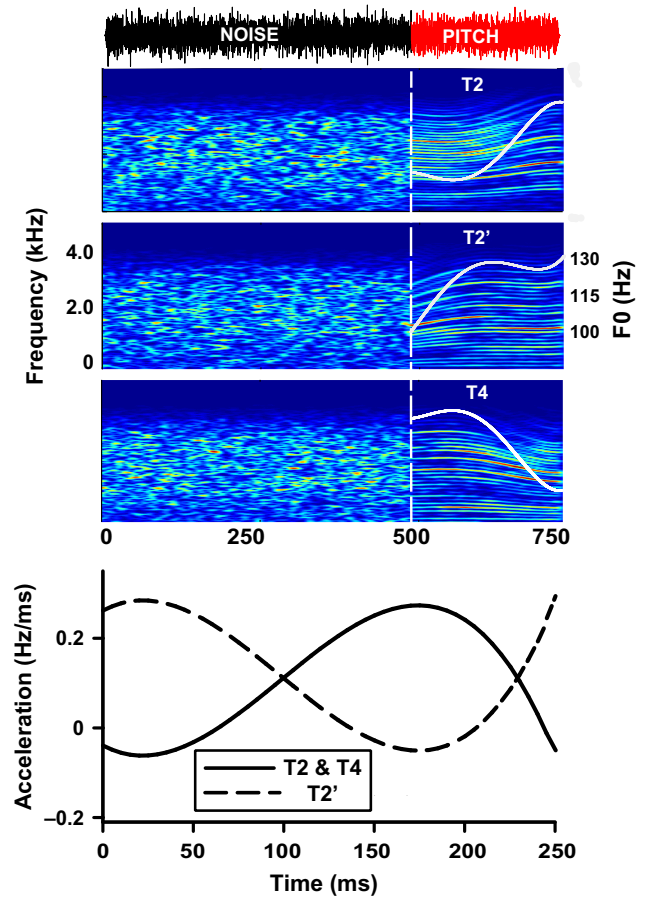

FIG. 1. (Top) Waveform (T2) and spectrograms of each of the three stimulus conditions (T2, T2′, T4) illustrate the experimental paradigm used to acquire cortical responses. The waveform shows robust periodicity within the pitch segment for T2. The vertical white dashed line at 500 ms demarcates the transition from the initial noise segment to the final pitch segment. CPRs were extracted from evoked responses beginning with pitch onset. F_0 contours (white) are superimposed on their respective pitch segments. T2 (rising) and T4 (falling), respectively, represent exemplary citation forms of Mandarin Tone 2 and Tone 4. T2′ (flipped variant of T2) exhibits a rising pitch contour similar in pitch direction to T2, although T2′ does not occur in the Mandarin tonal space Across stimuli, spectrograms show clear resolution of dynamic, rising or falling spectral bands corresponding to the harmonics of the fundamental frequency. (Bottom) In addition to changes in pitch direction, pitch contours are differentiated phonetically by changes in the location of peak acceleration (early vs. late). T2 and T4 (solid line) share the same acceleration trajectory throughout their duration. Both T2 and T4 reach a peak of acceleration late in the pitch contour; T2′ (dashed line), in contrast, reaches a peak early in the pitch contour.

(24.42 ms) of duration. Regardless of its location, the acceleration peak was constant across all three stimuli (0.28 Hz/ms). Duration was fixed at 250 ms across stimuli.

IRN was used to create these stimuli by applying procedures that generate dynamic curvilinear pitch patterns (Swaminathan et al., 2008). They were generated by applying polynomial equations (Supporting Information, Data S1). 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 1. By using IRN, we utilised pitch-specific stimuli that preserved dynamic variations in pitch of auditory stimuli that lacked a waveform periodicity, formant structure, temporal envelope and recognisable timbre characteristic of speech. Each stimulus condition consisted of two segments (crossfaded with 5 ms cos2 ramps): an

initial 500-ms noise segment followed by a 250-ms pitch segment, i.e., T2, T2′ and T4 (Fig. 1, top panel; Supporting Information, Media S1). The overall root-mean-square level of each segment was equated such that there was no discernible difference in intensity between initial and final 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 0.94/ s. Stimulus presentation order was randomised 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).

Cortical pitch response acquisition

Participants reclined comfortably in an electroacoustically shielded booth to facilitate recording of neurophysiological responses. They were instructed to relax and refrain from extraneous body movement to minimise myogenic artifacts. They were told to ignore the sounds they heard and were encouraged to sleep throughout the duration of the recording procedure. Almost all participants slept through the recording session and were awakened at the end of the session. The electroencephalogram (EEG) was acquired continuously (5000 Hz sampling rate; 0.3–2500 Hz analog band-pass) using the ASA-Lab EEG system (ANT Inc., The Netherlands) utilising a 32-channel amplifier (REFA8-32; TMS International BV) and WaveGuard (ANT Inc.) 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 slower cortical pitch components. Because the primary objective of this study was to characterise the cortical pitch components, the EEG acquisition electrode montage was limited to nine electrode locations: Fpz, AFz, Fz, F3, F4, Cz, T7, T8, M1 and 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 canthus 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 < 10 kΩ. For each stimulus, EEGs were acquired in blocks of 1000 sweeps. The experimental protocol took ~ 2 h to complete.

Extraction of the CPR

CPR responses were extracted off-line from the EEG files. To extract the cortical pitch response 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 minimise the sustained component. Sweeps containing electrical activity exceeding ± 50 μV 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. The re-referenced electrode site, Fz-linked T7/T8, was used to characterise the transient pitch response components. 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 and (ii) the inverted components recorded at the temporal electrode sites (T7 and T8). 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 1200 ms including the 100-ms pre-stimulus baseline.

Analysis of CPR

The evoked response to the entire three segment (noise-pitch-noise) stimulus is characterised 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 occurring after the onset of the pitch-eliciting segment of the stimulus and an offset component following the offset of the last noise segment in the stimulus. To characterise 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 response components (Na, Pb, Nb: time interval between pitch-eliciting stimulus onset and response peak of interest) and interpeak latency (Na–Pb, Pb–Nb: time interval between response peaks) were measured to enable us to identify the components associated with pitch onset, pitch acceleration, pitch direction and stimulus offset. Peak-to-peak amplitude of Na–Pb and Pb–Nb was measured to determine whether variations in amplitude indexed specific aspects of the pitch contour (i.e., pitch acceleration and/or direction). 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 laterality effects. To enhance visualisation of the laterality 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. The obligatory onset responses to the noise precursor were not analysed because they were invariant across the three stimuli.

Statistical analysis

Separate mixed-model 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 electrode site, and peak-to-peak amplitude derived from the T7/T8 and F3/F4 electrode sites. At the Fz electrode site, two-way ANOVAS were performed separately for each component on peak latency and peak-to-peak amplitude to assess language group (Chinese, English) and stimulus (T2, T2′, T4) effects. In the analysis of peak latency there were three components (Na, Pb, Nb) and in the analysis of peak-to-peak amplitude there were two components (Na–Pb, Pb–Nb). At the T7/T8 and F3/F4 electrode sites, three-way (group, stimulus, hemisphere) mixed-model ANOVAS were conducted separately on peak-to-peak amplitude of Na–Pb and Pb–Nb. Language group (Chinese, English) was treated as a between-subjects factor and subjects as a random factor nested within group; stimulus (T2, T2′, T4) and hemisphere (T7/T8, F3/F4) were treated as within-subject factors. A priori and *post hoc* multiple comparisons were corrected with a Bonferroni adjustment at $\alpha = 0.05$. Where appropriate, partial eta-squared (η_p^2) values were reported to indicate effect sizes.

Results

Response morphology of CPR components

Grand averaged cortical pitch response waveforms to the three stimuli are shown for the Chinese (red trace) and the English (blue

trace) groups in Fig. 2. CPR components (gray background) are clearly identifiable in both groups. The amplitude of the pitch-relevant components (Na, Pb, Nb) appears to be more robust for the Chinese group especially in response to T2 and T4. The larger amplitude for these two stimuli, which are exemplary representations of lexical tones, but not T2′, may be attributed to changes in location of peak acceleration and language-dependent sensitivity to specific acoustic attributes associated with pitch processing.

Fz: latency of CPR components

For both language groups, mean Fz peak latencies of CPR components Na, Pb and Nb increase systematically across stimuli in temporal order of occurrence (Supporting Information, Results, Table S1, Fig. S1a–b; cf. T7/T8 and F3/F4, respectively, in Figs S2 and S3). As reflected by Na, the omnibus ANOVA yielded a group × stimulus interaction ($F_{2,52} = 5.89$, $P = 0.0049$, $\eta_p^2 = 0.185$). By group, T2′, the pitch stimulus exhibiting an early peak of acceleration, elicited longer peak latencies than either T2 (Chinese: $t_{52} = -3.21$, $P = 0.0069$; English: $t_{52} = -7.89$, $P < 0.0001$) or T4

(Chinese: $t_{52} = 5.18$, $P < 0.0001$; English: $t_{52} = 6.40$, $P < 0.0001$) in both groups. By stimulus, response peak latencies for T2′ ($t_{52} = -6.29$, $P < 0.0001$) and T4 ($t_{52} = -5.16$, $P < 0.0001$) were longer in the English than the Chinese group. As reflected by Pb, the ANOVA yielded a main effect of stimulus ($F_{2,52} = 100.07$, $P < 0.0001$, $\eta_p^2 = 0.794$). Both groups exhibited a longer latency evoked by T2′ than by either T2 or T4. This means that, regardless of language experience, listeners exhibited longer latency to a pitch contour with an early acceleration peak (T2′) than to those with a late peak (T2, T4). As indexed by Nb, the ANOVA yielded a main effect of group ($F_{1,26} = 10.42$, $P = 0.0034$, $\eta_p^2 = 0.286$). The longer latency in the English group, relative to the Chinese, was evoked across stimuli.

Fz: amplitude of CPR components

Figure 3 displays group and stimulus effects on peak-to-peak amplitude of Na–Pb and Pb–Nb in response to all three stimuli (T2, T2′, T4). For Na–Pb (top panel), an omnibus ANOVA revealed main effects of group ($F_{1,26} = 18.73$, $P = 0.0002$, $\eta_p^2 = 0.418$) and stimulus ($F_{2,52} = 15.99$, $P < 0.0001$, $\eta_p^2 = 0.381$). Pooled across stimuli, Chinese exhibited greater amplitude than English; pooled across groups, T2 and T4 were greater than T2′, T2 was greater than T4. In the absence of a group × stimulus interaction, these differences may be attributed to acoustic properties of the stimuli: auditory sensitivity to location of peak acceleration (T2, T4 > T2′) and pitch direction (T2 > T4). For Pb–Nb (bottom panel), an omnibus ANOVA
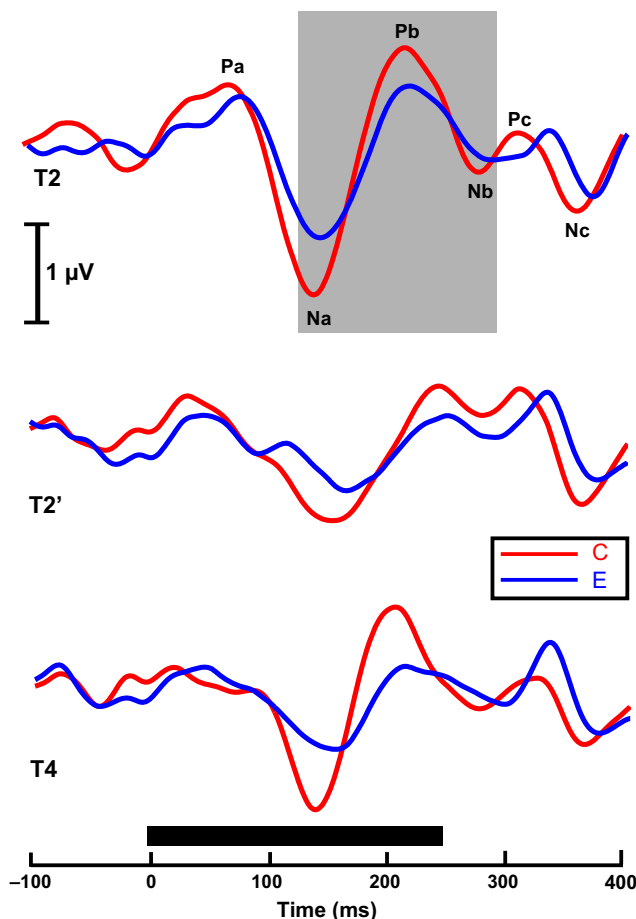

FIG. 2. Grand average waveforms of the Chinese and English groups at the Fz electrode site per stimulus condition (T2, T2′, T4). Na, Pb and Nb (highlighted in gray in the top panel) are the most robust pitch-relevant components. CPR waveforms elicited by the three stimuli show that amplitude of the pitch-relevant components (Na, Pb, Nb) appear to be more robust for the Chinese group especially in response to pitch stimuli with a late acceleration peak (T2, T4). Solid black horizontal bar indicates the duration of each stimulus.
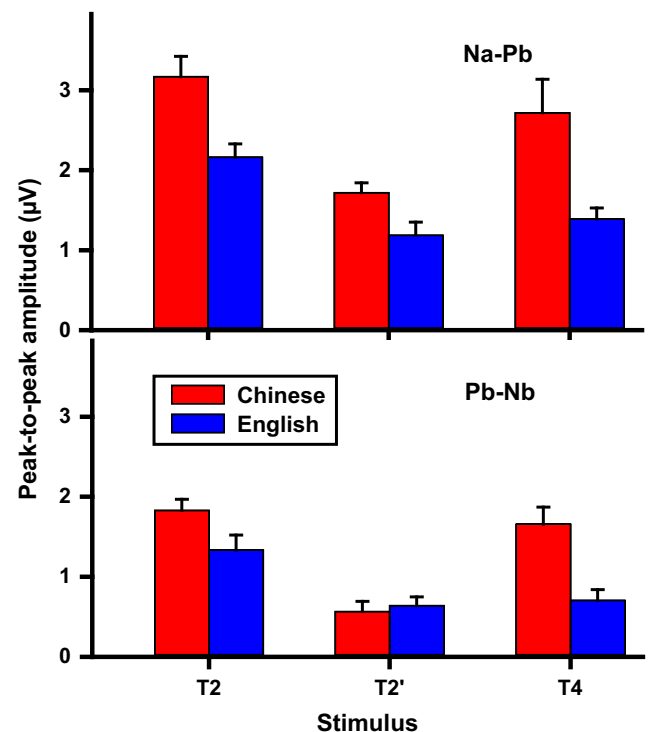
FIG. 3. Mean peak-to-peak amplitude of CPR components Na–Pb (top panel) and Pb–Nb (bottom panel) extracted from Fz electrode site as a function of stimulus. For Na–Pb, Chinese exhibit greater amplitude than English across stimuli; T2 and T4 are greater than T2′ across groups. For Pb–Nb, Chinese listeners show greater amplitude in response to those pitch contours with a late acceleration peak (T2, T4) as compared to T2′ with an early acceleration peak. English listeners, on the other hand, show greater amplitude of T2 relative to T4 in addition to T2′. Error bars are ±1 SE.

showed a significant group × stimulus interaction ($F_{2,52} = 6.32$, $P = 0.0035$, $\eta_p^2 = 0.195$). By group, *post hoc* comparisons indicated that Chinese exhibited greater amplitude in response to T2 ($t_{52} = 6.17$, $P < 0.0001$) and T4 ($t_{52} = -5.53$, $P < 0.0001$) as compared to T2′; in contrast, English showed greater amplitude to T2 than either T2′ ($t_{52} = 3.38$, $P = 0.0042$) or T4 ($t_{52} = -5.53$, $P < 0.0001$). This disparity in stimulus patterns between Chinese and English indicates that stimulus properties alone are insufficient to explain this language-dependent effect. We suggest that the dominant experience-dependent enhancement of the Chinese Pb–Nb component to native lexical tones T2 and T4 'masks' their differential sensitivity to location of peak acceleration and pitch direction, which is clearly observed for the English for whom there is no experience-dependent enhancement. By stimulus, Chinese amplitude was greater than English in response to T2 ($t_{52} = 2.25$, $P = 0.0288$) and T4 ($t_{52} = 4.31$, $P < 0.0001$), but not T2′. Taken together, these data suggest that early cortical stages of pitch processing are influenced by extrasensory, perceptually-relevant features of speech in one's native language.

T7/T8 & F3/F4: amplitude of CPR components

Grand average waveforms of the CPR components for each of the three stimuli per language group (left two columns) and their corresponding spectra (right two columns) are displayed in Fig. 4. CPR components in the Chinese group are greater in magnitude (left) and show a robust right hemisphere preference (right) for T2 and T4

with no discernible hemispheric asymmetry at the F3/F4 electrode sites (Supporting Information, Results, Fig. S4).

Group, stimulus and hemisphere effects on peak-to-peak amplitude of Na–Pb and Pb–Nb are displayed for the temporal sites (T7/T8) in Fig. 5 (Supporting Information, cf. frontal sites F3/F4 in Fig. S5). An omnibus three-way (group × stimulus × hemisphere) ANOVA on Na–Pb amplitude revealed a significant main effect of stimulus ($F_{2,52} = 27.32$, $P < 0.0001$, $\eta_p^2 = 0.735$) and a two-way interaction between group and hemisphere ($F_{1,26} = 15.87$, $P = 0.0005$, $\eta_p^2 = 0.379$). Pooled across group and hemisphere, stimuli varied in magnitude as a function of pitch direction (T2 > T4) and location of peak acceleration (T2, T4 > T2′). *Post hoc* comparisons at each level of group showed a RH advantage (T8 > T7) for the Chinese group only ($t_{26} = -5.76$, $P < 0.0001$). At each level of hemisphere, a group difference (Chinese > English) in Na–Pb amplitude was restricted to the RH ($t_{26} = 5.01$, $P < 0.0001$). The group × stimulus interaction failed to reach significance.

In contrast, results from an omnibus ANOVA on Pb–Nb amplitude yielded significant two-way interaction effects of group × hemisphere ($F_{1,26} = 5.71$, $P = 0.0245$, $\eta_p^2 = 0.180$) and group × stimulus ($F_{2,52} = 3.83$, $P = 0.0280$, $\eta_p^2 = 0.128$). *Post hoc* comparisons at each level of group and hemisphere revealed the same pattern as Na–Pb amplitude. Only the Chinese group showed a rightward asymmetry ($t_{26} = -3.90$, $P = 0.0006$); the language group effect was limited to the RH ($t_{26} = 3.88$, $P = 0.0006$). *Post hoc* comparisons at each level of group and stimulus revealed that Chinese Pb–Nb amplitude was greater than English in response to T2 ($t_{52} = 2.58$,
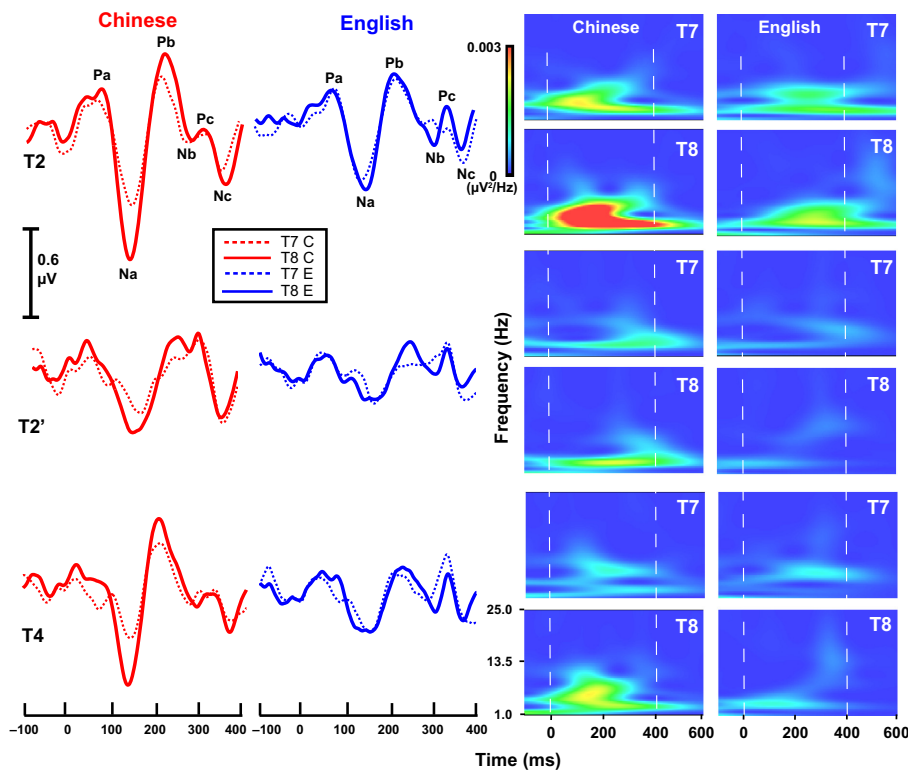

FIG. 4. Grand average waveforms (left) and their corresponding spectra (right) of the CPR components for the two language groups (Chinese, English) recorded at electrode sites T7 (dashed) and T8 (solid) for each of the three stimuli (T2, T2′, T4). CPR waveforms appear to show a right-sided preference (T8 > T7) for the Chinese group especially in response to pitch stimuli with a late acceleration peak representative of native lexical tones (T2, T4). The robust rightward preference for T2 and T4 is clearly evident in the spectrotemporal plots. For either group, no hemisphere effects are observed in response to T2′, a pitch pattern that does not occur in any of the world's tonal inventories. The zero on the *x*-axis denotes the time of onset of the pitch-eliciting segment of the three stimuli. Na–Pb and Pb–Nb time windows are demarcated by two vertical white dashed lines.
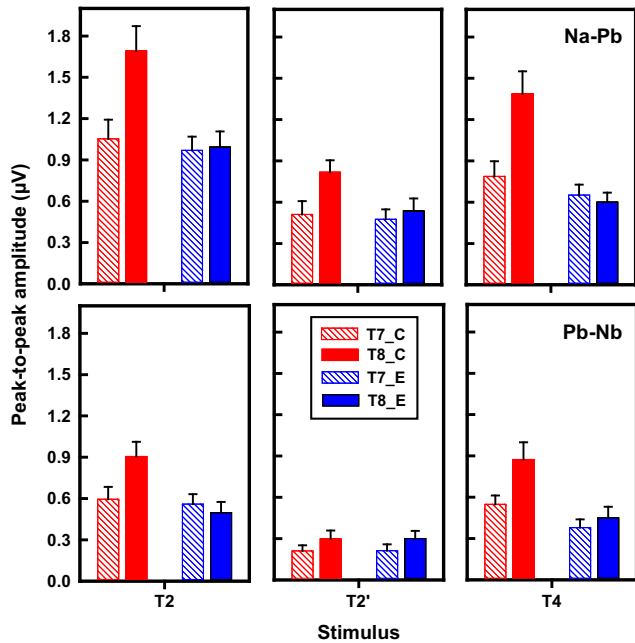
FIG. 5. Mean peak-to-peak amplitude of CPR components (Na–Pb, top row; Pb–Nb, bottom row) extracted from T7/T8 in the temporal lobe as a function of language group (Chinese, red; English, blue), stimulus (T2, T2′, T4), and hemisphere (left, diagonal; right, solid). Na–Pb amplitude shows a hemispheric asymmetry favoring the RH (T8 > T7) in response to T2 and T4 in the Chinese group only. These same two pitch stimuli also elicit a language group difference (Chinese > English) in the RH only. Both T2 and T4 are larger than T2′ only in the RH for the Chinese group. Pb–Nb amplitude similarly reveals a rightward asymmetry in the Chinese group exclusively as well as Chinese superiority relative to English in the RH. Error bars are ±1 SE. C, Chinese; E, English.

$P = 0.0127$) and T4 ($t_{52} = 3.45$, $P = 0.0011$), and that the amplitude of T2 ($t_{52} = 6.24$, $P < 0.0001$) and T4 ($t_{52} = -5.74$, $P < 0.0001$) was greater than T2′ for the Chinese group only. Important to note is that these stimulus patterns are identical to those obtained from Pb–Nb amplitude at the Fz electrode site. The stimulus patterns of English, on the other hand, differ from those of Chinese irrespective of electrode site: T7/T8 (T2 > T2′, $t_{52} = 3.41$, $P = 0.0037$) or F3/F4 (T2 > T2′, $t_{52} = 3.38$, $P = 0.0042$; T2 > T4, $t_{52} = 3.06$, $P = 0.0105$). In sum, we observe a divergence in stimulus patterns between Chinese and English. We also observe equivalent Nb–Pb responses of Chinese across temporal and frontal electrode sites. These combined findings lead us to suggest that extrasensory and sensory effects are differentially-weighted throughout the time course of the CPR.

Discussion

The major findings of this study demonstrate that pitch-relevant neural activity as reflected in the scalp-recorded CPR components show distinct changes that can be attributed to language-experience (Chinese vs. English); pitch patterns (native vs. nonnative); changes in acoustic attributes (rising vs. falling; and location of peak acceleration (early vs. late) of dynamic time-varying pitch contours. As reflected by the amplitude of the Pb–Nb component at both Fz and temporal (T7/T8) sites, Chinese, unlike English, show greater responses to pitch contours that occur in the Mandarin tonal space as compared to a nonnative pitch contour. As indexed by the amplitude of both components (Na–Pb, Pb–Nb) at temporal sites pooled

across stimuli, a rightward asymmetry occurs in the Chinese group only and, moreover, it is only over the right temporal site in which Chinese amplitude is greater than English. These findings suggest that basic neural mechanisms of pitch are sensitive to multiple attributes of dynamic pitch shared in common across languages at early stages of processing in the right auditory cortex, yet overlaid along the same pitch processing hierarchy are changes that reflect language-dependent modulation of those temporal attributes of pitch contours that provide perceptually salient cues to tonal recognition in one's native language.

Differential weighting of sensory and extrasensory effects in early cortical pitch processing

At a neurocomputational level, sensitivity may be manifested by response properties of neural elements including sharper tuning, greater temporal synchronisation and improved synaptic efficiency to enable optimal representation of behaviorally relevant dynamic pitch contours. Our experimental paradigm is free of task demands. Stimuli are reduced to the pitch parameter only. Thus, the observed electrophysiological responses are putatively specific to pitch. Our findings converge with an extant literature that attests to the crucial role of the RH in the processing of linguistic as well as nonlinguistic pitch (Zatorre et al., 2002; Friederici & Alter, 2004; Hyde et al., 2008; Meyer, 2008; Zatorre & Gandour, 2008). The preferential recruitment of pitch mechanisms in right auditory cortex by the Chinese group supports the notion that the RH is specifically involved in the analysis of suprasegmental parameters of speech (Friederici, 2011).

The question arises whether the effects of language experience on hemispheric laterality at early cortical stages of processing are driven by extrasensory influences as well as purely acoustic properties of the stimuli. 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 analysed 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 memory stores (e.g., short-term memory). It contains analysed sensory codes including information about pitch height, time-varying pitch direction and acceleration, and timing of pitch onset and offset. In this study, two of the stimuli are exemplary of pitch contours associated with lexical tones (T2, T4); one is not (T2′). As indexed by Na–Pb amplitude over the temporal sites, a rightward asymmetry is limited to the Chinese group. It is over the right temporal site that Chinese Na–Pb amplitude is larger than English. But the group × stimulus interaction is not significant. Therefore, any stimulus effects must be attributed primarily to acoustic properties of the stimuli: location of the acceleration peak (T2, T4 late > T2′ early) and pitch direction (T2 rising > T4 falling). Chinese enhancement of amplitude relative to English can be explained by invoking sensory influences alone or by claiming that sensory influences predominate over extrasensory. However, as indexed by Pb–Nb amplitude over the same temporal sites, the group × stimulus interaction was significant. Chinese Pb–Nb amplitude was greater than English in response to native pitch contours only (T2, T4). Chinese amplitude of T2 and T4 was greater than T2′, but this was not so for English. We argue that 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. If purely sensory, then we cannot account for why we do not observe the same stimulus pattern in the English group.

We expect extrasensory influences to target especially those pitch attributes that are perceptually salient in a particular language. It is not accidental that extrasensory effects emerge in Pb–Nb instead of Na–Pb. In a behavioral experiment using excised segments from F_0 contours of Mandarin tones (Whalen & Xu, 1992), tonal recognition is 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 F_0. More recently, analysis of brainstem responses in both speech and nonspeech contexts reveals that pitch representations are stronger in Chinese than English in the later, rapidly changing, portions of Tone 2 and Tone 4 (Krishnan *et al.*, 2009a,b). Though we are unable to match up portions of F_0 contours with CPR components in the current experimental design, we speculate that Pb–Nb is targeting those same perceptually relevant portions of T2 and T4.

As indexed by either Na–Pb or Nb–Pb amplitude over the frontal electrode sites (F3/F4), no hemispheric preferences are evoked by any of the three stimuli irrespective of language group. Though Chinese amplitude is greater than English, stimulus patterns are similar to those over the temporal sites (T2, T4 > T2′; T2 > T4). In the absence of a significant group × stimulus interaction, these stimulus effects can be explained simply on physical properties of the stimuli (location of acceleration peak and pitch direction). This disparity between the temporal and frontal sites is consistent with extant literature that identifies the right auditory cortex as playing a critical role in early stages of pitch processing. Similar to the RH preference for processing linguistic pitch as reflected by the CPR, an RH preference has also been reported for processing the more salient consonantal musical stimuli in musicians using the pitch-onset response (Bidelman & Grall, 2014). These findings taken together suggest that the RH is preferentially recruited for optimal representation of pitch attributes that are perceptually relevant regardless of the domain in which they are presented.

Effects of acoustic properties of stimuli in early cortical pitch processing

Our findings on Fz peak latency of CPR components point to effects on the temporal integration window. Regardless of group, T2′ elicited a longer latency than T2 or T4 as indexed by components Na and Pb. A longer temporal integration window for T2′ probably reflects decreased temporal sensitivity and/or neural desynchronisation to a rapidly rising portion of the pitch contour that occurs early as compared to T2 and T4 in which the rapidly gliding portion occurs much later. In the case of Na, English listeners show a longer latency than Chinese in response to T2′ and T4. These findings together suggest that both the location of acceleration peak and pitch direction may differentially affect the duration of the temporal integration window.

Our findings on Fz peak-to-peak amplitude of CPR components reveal that Chinese exhibit greater sensitivity to both pitch direction and location of acceleration peak. As indexed by Na–Pb, the rising pitch contour (T2) evokes larger amplitude than the falling (T4) across language groups, yet the two stimuli share in common a late acceleration peak. This differential sensitivity to pitch direction is supported at multiple levels of the auditory system by various experimental techniques (human psychophysical, multidimensional scaling, electrophysiological, cochlear microphonics, 8th nerve compound action potentials, and responses of the ventral cochlear nucleus; see Supporting Information, Discussion, for

details). T2′ also has a rising pitch contour yet it does not exhibit larger amplitude than T4. Its absence is probably due to biomechanical constraints on the velocity of laryngeal movements in tone production (Ohala, 1978; Xu & Sun, 2002; D. Erickson, unpublished data). It cannot be due to differences in acceleration rate per se because T2 and T2′ share an identical acceleration trajectory. It must therefore be attributed to its early location near the onset of T2′, which probably reflects differences in neural synchronisation when a rapid rise in pitch occurs earlier as compared to later in the pitch contour. Differences in pitch direction notwithstanding, the two stimuli with a late acceleration peak (T2, T4) have larger amplitude than T2′.

As indexed by Pb–Nb, we found an interaction (group × stimulus) that reveals differential weighting of sensory and extrasensory components depending upon one's language experience. In the Chinese group, the two lexical tones (T2, T4) have greater amplitude than T2′. In contrast, T2 amplitude is greater than T2′ and T4 in the English group. The English pattern may be attributed to differences in auditory sensitivity to pitch direction (T2 > T4) and location of peak acceleration (T2 > T2′). Although the Chinese pattern similarly shows auditory sensitivity to location of peak acceleration, this specific pitch attribute also segregates native lexical tones (T2, T4) from a nonnative pitch stimulus (T2′). If strictly sensory, one would expect parallel stimulus patterns irrespective of language experience. On the contrary, we observed that Chinese amplitude was greater than English in response to just those pitch contours representative of Mandarin tones. The lack of a language group effect in response to T2′ is in agreement with previous studies which similarly failed to show experience-dependent enhancement of pitch-relevant neural activity for nonnative pitch contours at the level of the cerebral cortex (Chandrasekaran *et al.*, 2009; T2′; Krishnan *et al.*, 2014; flat and linear rising ramp) and at the level of the auditory brainstem (Krishnan *et al.*, 2009a, linear rising and trilinear rising ramps, T2′).

These findings can be accounted for by invoking the influence of extrasensory effects on pitch processing that are associated with perceptually relevant features of Mandarin lexical tones. Even though T2 (late acceleration peak) elicits larger amplitude than T2′ (early acceleration peak) in both groups, the amplitude of T2 is still larger in the Chinese group. This finding suggests that the fundamental neural mechanism is the same for Chinese and English listeners alike, but Chinese are more sensitive to pitch attributes that are behaviorally relevant for pitch processing because of their long-term experience with a tonal language. Because enhanced sensitivity to time-varying dimensions of pitch (e.g. acceleration) is already present in neural activity at the level of the brainstem (Krishnan & Gandour, 2009; Krishnan *et al.*, 2012; reviews) it seems plausible that cortical pitch mechanisms may be reflecting, at least in part, this enhanced pitch input from the brainstem (Bidelman *et al.*, 2014).

Predictive coding may underlie experience-dependent processing of pitch in the auditory cortex

Growing evidence suggests that pitch-relevant information is available in primary and non-primary areas of 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 & Samson, 1991; Johnsrude *et al.*, 2000), and magnetoencephalography (Gutschalk *et al.*, 2002, 2004; Krumbholz *et al.*, 2003). Lateral Heschl's gyrus (HG) also appears to be important for computations relevant

to extraction of pitch of complex sounds (Zatorre & 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 & Ballard, 1999; Kumar *et al.*, 2011; Kumar & 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 minimise predictive error and to optimise 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 greater 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 T2 and T4 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.

Pitch processing in the auditory cortex is also 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 (Balaguer-Ballester *et al.*, 2009). This expanded model represents a unified, physiologically plausible, theoretical framework that includes both cortical and subcortical components in the hierarchical processing of pitch.

Conclusions

CPR components provide a series of robust neurobiological markers that reveal 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.

Supporting Information

Additional supporting information can be found in the online version of this article:

Data S1. Equations to generate the IRN stimuli (*T2, T2′, T4*).

Table S1. Mean peak latencies of CPR components Na, Pb, and Nb derived from Fz electrode site.

Fig. S1. (a, b) Mean peak latencies of CPR components Na, Pb, and Nb derived from Fz electrode site.

Fig. S2. Mean peak latencies of CPR components Na, Pb, and Nb derived from T7/T8 electrode sites.

Fig. S3. Mean peak latencies of CPR components Na, Pb, and Nb derived from F3/F4 electrode sites.

Fig. S4. Grand average waveforms and their corresponding spectra of CPR components extracted from F3/F4.

Fig. S5. Mean peak-to-peak amplitude of CPR components extracted from F3/F4.

Media S1. Audio of IRN stimuli (*irnpitchT2.mp3*; *irnpitchT2′.mp3*; *irnpitchT4.mp3*) and stimulus condition (*irnnoisetopitchT2.mp3*).

Acknowledgements

Abbreviations

CPR, cortical pitch-specific response; EEG, electroencephalogram or electroencephalography; HG, Heschl's gyrus; IRN, iterated rippled noise; MMN, mismatch negativity; RH, right hemisphere; T2, Mandarin Tone 2; T2′, flipped variant of Mandarin Tone 2; T4, Mandarin Tone 4.

References

Alho, K., Grimm, S., Mateo-Leon, S., Costa-Faidella, J. & Escera, C. (2012) Early processing of pitch in the human auditory system. *Eur. J. Neurosci.*, **36**, 2972–2978.

Balaguer-Ballester, E., Clark, N.R., Coath, M., Krumbholz, K. & Denham, S.L. (2009) Understanding pitch perception as a hierarchical process with top-down modulation. *PLoS Comput. Biol.*, **5**, e1000301.

Bidelman, G.M. & Grall, J. (2014) Functional organization for musical consonance and tonal pitch hierarchy in human auditory cortex. *NeuroImage*, **101**, 204–214.

Bidelman, G.M., Weiss, M.W., 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.

Chandrasekaran, B., Krishnan, A. & Gandour, J.T. (2009) Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear Hearing*, **30**, 552–558.

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. Human.*, **13**, 204–215.

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

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

Griffiths, T.D., Kumar, S., Sedley, W., Nourski, K.V., Kawasaki, H., Oya, H., Patterson, R.D., Brugge, J.F. & Howard, M.A. (2010) Direct recordings of pitch responses from human auditory cortex. *Curr. Biol.*, **20**, 1128–1132.

Gutschalk, A., Patterson, R.D., 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, R.D., Scherg, M., Uppenkamp, S. & Rupp, A. (2004) Temporal dynamics of pitch in human auditory cortex. *NeuroImage*, **22**, 755–766.

Hall, D.A., Johnsrude, I.S., Haggard, M.P., Palmer, A.R., Akeroyd, M.A. & Summerfield, A.Q. (2002) Spectral and temporal processing in human auditory cortex. *Cereb. Cortex*, **12**, 140–149.

Howie, J.M. (1976) *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge University Press, New York.

Hyde, K.L., Peretz, I. & Zatorre, R.J. (2008) Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, **46**, 632–639.

Itoh, K., Okumiya-Kanke, Y., Nakayama, Y., Kwee, I.L. & Nakada, T. (2012) Effects of musical training on the early auditory cortical representation of pitch transitions as indexed by change-N1. *Eur. J. Neurosci.*, **36**, 3580–3592.

Johnsrude, I.S., Penhune, V.B. & Zatorre, R.J. (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, **123**, 155–163.

Koelsch, S. (2012) *Brain & Music*. Wiley-Blackwell, Chichester, UK.

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

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

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

Krishnan, A., Gandour, J.T. & Bidelman, G.M. (2012) Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. *NeuroReport*, **23**, 498–502.

Krishnan, A., Gandour, J.T. & Suresh, C.H. (2014) Cortical pitch response components show differential sensitivity to native and nonnative pitch contours. *Brain Lang.*, **138**, 51–60.

Krishnan, A., Gandour, J.T., Ananthakrishnan, S. & Vijayaraghavan, V. (2015) Language experience enhances early cortical pitch-dependent responses. *J. Neurolinguist.*, **33**, 128–148.

Krumbholz, K., Patterson, R.D., 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, K.V., Kawasaki, H., Oya, H., Patterson, R.D., Howard, M.A. 3rd, Friston, K.J. & Griffiths, T.D. (2011) Predictive coding and pitch processing in the auditory cortex. *J. Cognitive Neurosci.*, **23**, 3084–3094.

Luo, H., Ni, J.T., Li, Z.H., Li, X.O., Zhang, D.R., Zeng, F.G. & Chen, L. (2006) Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proc. Natl. Acad. Sci. USA*, **103**, 19558–19563.

Lutkenhoner, B., Seither-Preisler, A. & Seither, S. (2006) Piano tones evoke stronger magnetic fields than pure tones or noise, both in musicians and non-musicians. *NeuroImage*, **30**, 927–937.

Maess, B., Jacobsen, T., Schroger, E. & Friederici, A.D. (2007) Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *NeuroImage*, **37**, 561–571.

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

Moore, C.B. & Jongman, A. (1997) Speaker normalization in the perception of Mandarin Chinese tones. *J. Acoust. Soc. Am.*, **102**, 1864–1877.

Ohala, J. (1978) The production of tone. In Fromkin, V. (Ed.), *Tone: A Linguistic Survey*. Academic Press, New York, pp. 15–39.

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*. Oxford University Press, New York, NY.

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S. & Griffiths, T.D. (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron*, **36**, 767–776.

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

Pulvermuller, F., Shtyrov, Y. & Hauk, O. (2009) Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain Lang.*, **110**, 81–94.

Puschmann, S., Uppenkamp, S., Kollmeier, B. & Thiel, C.M. (2010) Dichotic pitch activates pitch processing centre in Heschl's gyrus. *NeuroImage*, **49**, 1641–1649.

Rao, R.P. & Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.*, **2**, 79–87.

Schonwiesner, M., Rubsamen, R. & von Cramon, D.Y. (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, J.T. & Xu, Y. (2008) Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. *IEEE T. Bio-med. Eng.*, **55**, 281–287.

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

Wong, P.C. & Perrachione, T.K. (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, J.T. & Francis, A.L. (2006) Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J. Acoust. Soc. Am.*, **120**, 1063–1074.

Yrttiaho, S., Tiitinen, H., May, P.J., Leino, S. & Alku, P. (2008) Cortical sensitivity to periodicity of speech sounds. *J. Acoust. Soc. Am.*, **123**, 2191–2199.

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

Zatorre, R.J. & Belin, P. (2001) Spectral and temporal processing in human auditory cortex. *Cereb. Cortex*, **11**, 946–953.

Zatorre, R.J. & Gandour, J.T. (2008) Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos. T. Roy. Soc. B.*, **363**, 1087–1104.

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