

Cortical pitch response components show differential sensitivity to native and nonnative pitch contours



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

Department of Speech Language Hearing Sciences, Purdue University, USA

ARTICLE INFO

Article history:

Accepted 21 September 2014

Keywords:

Pitch
Iterated rippled noise
Cortical pitch response
Pitch acceleration
Experience-dependent plasticity
Functional asymmetry
Tone language
Lexical tone
Mandarin Chinese

ABSTRACT

The aim of this study is to evaluate how nonspeech pitch contours of varying shape influence latency and amplitude of cortical pitch-specific response (CPR) components differentially as a function of language experience. Stimuli included time-varying, high rising Mandarin Tone 2 (T2) and linear rising ramp (Linear), and steady-state (Flat). Both the latency and magnitude of CPR components were differentially modulated by (i) the overall trajectory of pitch contours (time-varying vs. steady-state), (ii) their pitch acceleration rates (changing vs. constant), and (iii) their linguistic status (lexical vs. non-lexical). T2 elicited larger amplitude than Linear in both language groups, but size of the effect was larger in Chinese than English. The magnitude of CPR components elicited by T2 were larger for Chinese than English at the right temporal electrode site. Using the CPR, we provide evidence in support of experience-dependent modulation of dynamic pitch contours at an early stage of sensory processing.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Pitch is an important information-bearing perceptual component of language and music. As such, it provides an excellent window for studying experience-dependent effects on both cortical and brainstem structures of a well-coordinated, hierarchical network. It is our view that a complete understanding of the neural organization of language (and music) can only be achieved by assuming that linguistic (musical) computations are implemented in the brain in real time at different levels of biological analysis (Poeppl & Embick, 2006). In the case of pitch, continuous physical signals are transformed into neural representations at different stages of processing modulated by experience-dependent sensitivity to relevant features. Recent empirical data show that neural representation of pitch is shaped by one's experience with language and music at the level of the auditory brainstem as well as the cerebral cortex (Besson, Chobert, & Marie, 2011; Gandour & Krishnan, 2014; Koelsch, 2012; Kraus & Banai, 2007; Krishnan, Gandour, & Bidelman, 2012; Kuhn, Elmer, Meyer, & Jancke, 2013; Meyer, 2008; Moreno & Bidelman, 2014; Munte, Altenmuller, & Jancke, 2002; Patel & Iversen, 2007; Tervaniemi

et al., 2009; Zatorre & Baum, 2012; Zatorre, Belin, & Penhune, 2002; Zatorre & Gandour, 2008). These empirical findings notwithstanding, we have yet to achieve a more precise characterization of neural representation of pitch-relevant attributes that are sensitive to one's language experience.

Pitch is a multidimensional perceptual attribute that relies on several acoustic dimensions, one of which is *contour* (i.e., changes in pitch direction between onset and offset). Indeed, F_0 height and *contour* are important, experience-dependent dimensions of pitch underlying the perception of lexical tone (Francis, Ciocca, Ma, & Fenn, 2008; Gandour, 1983; Huang & Johnson, 2011; Khouw & Ciocca, 2007). The extant literature aimed at cortical processing of pitch contours in the language domain is sparse. Using the mismatch negativity (MMN), Chinese listeners, relative to English listeners, were more sensitive to pitch contour than pitch height in response to Mandarin tones, indicating that MMN may serve as a neural index of the relative saliency of underlying dimensions of pitch that are differentially weighted by language experience (Chandrasekaran, Gandour, & Krishnan, 2007). In Cantonese, the magnitude and latency of MMN were sensitive to the size of pitch height change, while the latency of P3a (automatic attention shift induced by the detection of deviant features in the passive oddball paradigm) captured the presence of a change in pitch contour (Tsang, Jia, Huang, & Chen, 2011). Though contour and height are important dimensions that are implicated in early, cortical pitch processing, the MMN itself is *not* a pitch-specific

* Corresponding author at: Purdue University, Department of Speech Language Hearing Sciences, Lyles-Porter Hall, 710 Clinic Drive, West Lafayette, IN 47907, USA. Fax: +1 765 494 0771.

E-mail addresses: rkrish@purdue.edu (A. Krishnan), gandour@purdue.edu (J.T. Gandour), hs0@purdue.edu (C.H. Suresh).

response. It is comprised of both auditory and cognitive mechanisms of frequency change detection in auditory cortex (Maess, Jacobsen, Schroger, & Friederici, 2007). This parallel processing is consistent with the near-simultaneity of neurophysiological indicators of psycholinguistic information in the first 200–250 ms (Pulvermüller, Shtyrov, & Hauk, 2009). Thus, it is necessary to develop an early, preattentive cortical brain response that is exclusive to pitch in order to disentangle pitch from other neurophysiological indicators of the processing of lexical tone. Such a pitch-specific, neural metric will also provide us a window to examine possible interactions between pitch and higher-order linguistic and cognitive mechanisms at an early, sensory level of processing in the auditory cortex.

At the cortical level, magnetoencephalography (MEG) has been used previously to study the sensitivity to pitch-relevant periodicity by investigating the N100m component. However, a large proportion of the N100m is simply a response to the onset of sound energy, and not exclusive to pitch (Alku, Sivonen, Palomaki, & Tiitinen, 2001; Gutschalk, Patterson, Scherg, Uppenkamp, & Rupp, 2004; Lutkenhoner, Seither-Preisler, & Seither, 2006; Soeta & Nakagawa, 2008; Yrttiaho, Tiitinen, May, Leino, & Alku, 2008). In order to disentangle the pitch-specific response from the 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) matched in intensity and overall spectral profile (Krumbholz, Patterson, Seither-Preisler, Lammertmann, & Lutkenhoner, 2003). Interestingly, a transient pitch onset response (POR) was evoked from this noise-to-pitch transition only. The reverse stimulus transition from pitch-to-noise failed to produce a POR. It has been proposed that the human POR, as measured by MEG, reflects synchronized cortical neural activity specific to pitch (Chait, Poeppel, & Simon, 2006; Krumbholz et al., 2003; Ritter, Gunter, Dorsch, Specht, & Rupp, 2005; Seither-Preisler, Patterson, Krumbholz, Seither, & Lutkenhoner, 2006). POR latency and magnitude, for example, have been shown to depend on pitch salience. A more robust POR with shorter latency is observed for stimuli with stronger pitch salience as compared to those with weaker pitch salience. Source analyses (Gutschalk, Patterson, Rupp, Uppenkamp, & Scherg, 2002; Gutschalk et al., 2004; Krumbholz et al., 2003), corroborated by human depth electrode recordings (Griffiths et al., 2010; Schonwiesner & Zatorre, 2008), indicate that the POR is localized to the anterolateral portion of Heschl's gyrus, the putative site of pitch processing (Bendor & Wang, 2005; Griffiths, Buchel, Frackowiak, & Patterson, 1998; Johnsrude, Penhune, & Zatorre, 2000; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Penagos, Melcher, & Oxenham, 2004; Zatorre, 1988).

Using a similar two-segment stimulus paradigm, we demonstrated that the EEG-derived human cortical pitch response (CPR) elicited by IRN *steady-state* pitch stimuli increased in magnitude with increasing temporal regularity (waveform pattern that repeats regularly in time) of the stimulus (Krishnan, Bidelman, Smalt, Ananthakrishnan, & Gandour, 2012). This change in CPR response amplitude with increasing stimulus temporal regularity was strongly correlated with behavioral measures of change in pitch salience. No CPR was evoked to a “no-pitch” IRN stimulus. Thus, the CPR is specific to pitch and its salience rather than simply a neural response to IRN elicited by slow, spectrotemporal modulations unrelated to pitch (Barker, Plack, & Hall, 2012).

This initial finding prompted us to examine the sensitivity of the multiple transient components of the CPR to *time-varying* pitch stimuli: three, within-category variants of Mandarin Chinese Tone 2 (T2) (Krishnan, Gandour, Ananthakrishnan, & Vijayaraghavan, 2014a). Based on responses from Chinese listeners, the pitch onset component, Na, was invariant to changes in pitch acceleration. In contrast, Na–Pb and Pb–Nb showed a systematic decrease in

interpeak latency and decrease in amplitude with increasing pitch acceleration that followed the time course of the pitch contours. Pc–Nc marked unambiguously the stimulus offset. We hypothesized that a series of neural markers flag different temporal attributes of a dynamic pitch contour: onset of temporal regularity (Na); changes in temporal regularity between onset and offset (Na–Pb, Pb–Nb); and offset of temporal regularity (Pc–Nc). A right hemisphere (RH) preference was observed at temporal electrode sites only for the prototypical variant of T2. Taken together, CPR responses to dynamic pitch appear to provide a window on the emergence of hemispheric preferences at an early sensory level of processing, and moreover, the interaction between acoustic and linguistic properties of the stimulus.

In a companion study (Krishnan et al., 2014a), we employed the same three within-category variants of T2 to examine how language experience (Mandarin vs. English) shapes the processing of temporal attributes of pitch as reflected in the CPR components. The magnitude of Na–Pb and Pb–Nb and their correlation with pitch acceleration were stronger for Chinese than for English listeners. Discriminant function analysis revealed that the Na–Pb component was more than twice as important as Pb–Nb in grouping listeners by language affiliation. In addition, a stronger, stimulus-dependent RH preference was observed for the Chinese group at the temporal, but not frontal, electrode sites. These combined findings suggest that long-term language experience shapes early sensory level processing of pitch in the auditory cortex, and that the sensitivity of the CPR may vary depending on the relative linguistic importance of specific temporal attributes of dynamic pitch.

Up to the present, we have investigated dynamic (curvilinear; T2) and static (steady-state) pitch stimuli separately. Thus, the overall objective of the present study is to evaluate how pitch contours of varying shape may influence latency and amplitude of CPR components differentially as a function of language experience (Chinese, English). We chose three, nonspeech pitch stimuli: high rising Mandarin Tone 2 (T2); linear rising ramp (Linear); steady-state or constant (Flat). T2 and Linear exhibit dynamic, time-varying pitch; Flat, static, steady-state pitch. T2, however, is the only stimulus that is representative of a pitch contour that occurs in natural speech. These differences in pitch trajectory are of crucial importance to our experimental design because of the sensitivity of the CPR to specific temporal attributes of dynamic pitch (Krishnan, Gandour, Ananthakrishnan, & Vijayaraghavan, 2014b; Krishnan et al., 2012). The use of iterated rippled noise (IRN) enables us to create stimuli that preserve dynamic variations in pitch minus waveform periodicity, formant structure, temporal envelope, and recognizable timbre characteristic of speech (Swaminathan, Krishnan, & Gandour, 2008). By including a non-tone language group (English), we can evaluate whether or not any observed effects on pitch representation are language-dependent. By comparing T2 and Linear to Flat, we can assess the effect of dynamic vs. static pitch on CPR components. A direct comparison of curvilinear T2 to phonetically-similar Linear enables us to determine whether a pitch contour exemplary of a lexical tone modulates pitch encoding at an early sensory level processing in the auditory cortex. A positive language-dependent effect (Chinese > English) would point to an interaction between sensory and cognitive components in pitch processing. By evaluating CPR components at frontal and temporal electrode sites over both hemispheres, we are able to evaluate the presence/absence of language-dependent hemispheric preferences in the processing of dynamic vs. static pitch. Chinese listeners, relative to English, are hypothesized to exhibit a stronger rightward asymmetry at the temporal electrode sites. This experimental outcome would support the idea of experience-dependent modulation of pitch-specific mechanisms at an early sensory stage of processing in right auditory cortex.

2. Materials and methods

2.1. Participants

Twelve native speakers of Mandarin Chinese (6 male, 6 female) and English (7 male, 5 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: 22.1 ± 2.1 years; English: 21.6 ± 1.6), years of formal education (Chinese: 15.3 ± 1.8 years; English: 15.8 ± 1.3), and were strongly right handed (Chinese: $93 \pm 9.2\%$; English: $95.8 \pm 8.3\%$) 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 (11.3 ± 2.2 years). As determined by a music history questionnaire (Wong & Perrachione, 2007), all Chinese and English participants had less than two years of musical training (Chinese, 1.2 ± 1.3 years; English, 1 ± 1.2) on any combination of instruments. No participant had any training within the past five years. Each participant was paid and gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

2.2. Stimuli

Three iterated rippled noise (IRN) stimuli were constructed to investigate CPR responses to steady-state and time-varying pitch stimuli (Fig. 1, top panel). There were two time-varying pitch stimuli. One consisted of a curvilinear pitch contour modeled after productions of Mandarin Tone 2 (T2) on an isolated monosyllable (Howie, 1976; Moore & Jongman, 1997; Xu, 1997), with an average F0 of 111 Hz and a changing pitch acceleration rate. Its peak acceleration rate occurred at 177 ms. The other was a linear rising ramp (Linear), a crude approximation of T2 that is not observable in natural speech, with an average F0 of 117 Hz. Unlike T2, its acceleration rate was constant. The Linear stimulus, however, shared F0 onset/offset (103/131 Hz) and average F0 acceleration (0.112 Hz/ms) in common with T2 (Fig. 1, top and bottom panels, respectively). The third stimulus, Flat, exhibited a steady-state pitch of 103 Hz. Like the Linear stimulus, it does not occur in natural speech. The Flat stimulus shared only pitch onset in common with T2 and Linear. Duration was fixed at 250 ms across stimuli.

IRN was used to create these stimuli by applying procedures that generate static and dynamic (linear, curvilinear) pitch patterns (Swaminathan, Krishnan, Gandour, & Xu, 2008). T2 and Linear were generated by applying polynomial and linear equations, respectively; Flat was constant at 103 Hz (Appendix A.1, equations). 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 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 two segments (crossfaded with 5 ms cos2 ramps): an initial 500 ms noise segment followed by a 250 ms pitch segment, i.e., T2, Linear, and Flat (Fig. 1; Appendix A.2, audio files; Appendix B.1, Fig. S1). The overall RMS 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

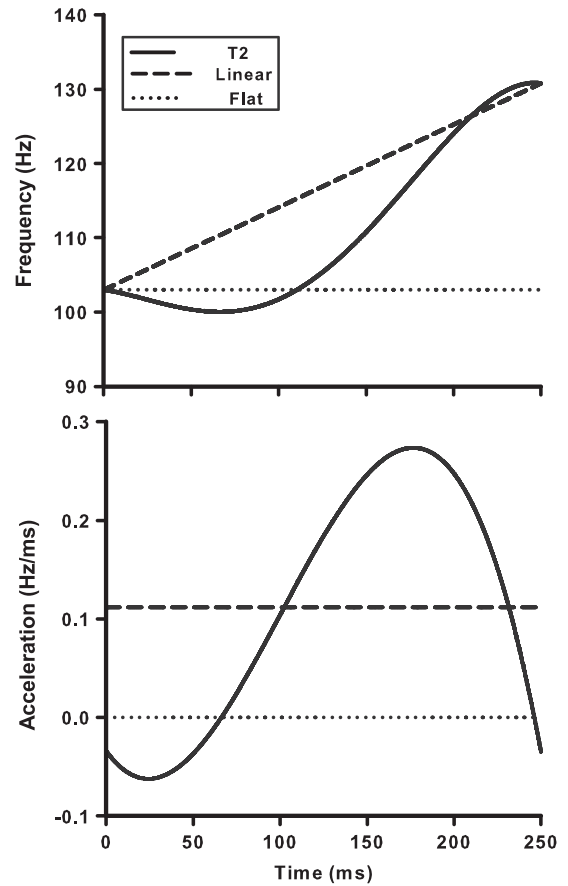


Fig. 1. IRN stimuli used to evoke cortical responses to pitch patterns that are differentiated phonetically by the shape of the contour. Voice fundamental frequency (F_0) contours (top panel) and corresponding acceleration trajectories (bottom panel) are displayed for all three stimuli. T2 (curvilinear), exemplary of Mandarin Tone 2, and Linear both represent time-varying rising pitch contours; Flat represents a steady-state or flat pitch. T2 is the only pitch pattern that occurs in natural speech and the only one to exhibit a changing acceleration rate.

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).

2.3. Cortical pitch response acquisition

Participants reclined comfortably in an electro-acoustically shielded booth to facilitate recording of neurophysiologic responses. They were instructed to relax and refrain from extraneous body movement to minimize 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 EEG was acquired continuously (5000 Hz sampling rate; 0.3–2500 Hz analog band-pass) 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 slower 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 9 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.

2.4. Extraction of the cortical pitch response (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) to enhance the transient components and minimize the sustained component. Sweeps containing electrical activity exceeding ± 50 μ V were rejected automatically. Subsequently, averaging was performed on all 8 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 characterize 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.

2.5. Analysis of CPR

The CPR 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 response components occurring after the onset of the pitch-eliciting segment of 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 response components (Na, Pb, Nb: time interval between pitch-eliciting stimulus onset and response peak of interest) and interpeak latency (Na–Pb: time interval between response peaks) were measured to enable us to identify the components associated with pitch onset, pitch acceleration, 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 (e.g., pitch acceleration). 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 visualization 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. Since our primary focus is on pitch relevant components, the obligatory onset responses to the noise precursor, invariant across the three stimuli, were not analyzed.

2.6. Statistical analysis

Separate ANOVAs (SAS[®]; SAS Institute, Inc., Cary, NC, USA) were conducted on peak latency, interpeak latency, and peak-to-peak

amplitude of the CPR 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, separate one-way ANOVAs were performed on peak latency, interpeak latency, and peak-to-peak amplitude to assess language group effects at each combination of component and stimulus (T2, Linear, Flat). In the analysis of peak latency, there were three components (Na, Pb, Nb); interpeak latency, one component (Na–Pb); and peak-to-peak amplitude, two components (Na–Pb, Pb–Nb). At the T7/T8 and F3/F4 electrode sites, separate two-way (group \times hemisphere), mixed model ANOVAs were similarly conducted on peak-to-peak amplitude of Na–Pb and Pb–Nb at each combination of component and stimulus. Language group (Chinese, English) was treated as a between-subjects factor and subjects as a random factor nested within group. Group and hemisphere were treated as within-subject factors. By focusing on the frontal and temporal sites, we were able to determine whether pitch-related laterality effects on Na–Pb and Pb–Nb vary as a function of language experience. To make a direct comparison between T2 and Linear, we also performed a two-way ANOVA (group \times stimulus) on the peak-to-peak amplitude of Na–Pb and Pb–Nb at the Fz site, and a three-way ANOVA (group \times stimulus \times hemisphere) at the frontal (F3/F4) and temporal (T7/T8) sites. Within each ANOVA, a priori or post hoc multiple comparisons were corrected with a Bonferroni adjustment at $\alpha = 0.05$, and further adjusted across ANOVAs depending on the number of stimulus comparisons. In the case of separate ANOVAs conducted on three stimuli, for example, an alpha level of significance of .05 was adjusted to .0166. Where appropriate, partial eta-squared (η_p^2) values were reported to indicate effect sizes.

3. Results

3.1. 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) group in Fig. 2. CPR components are clearly identifiable for both groups. The amplitude of the pitch-relevant components (Na, Pb, Nb) generally appears to be more robust for the Chinese group for all three stimuli, especially in response to T2. The larger amplitude of T2 in the Chinese group may reflect an experience-dependent enhancement of components related to pitch. In contrast, the offset components (Pc, Nc) are more robust for the English group, particularly for the dynamic pitch stimuli (T2, Linear). For both groups, pitch-relevant components Na and Pb show longer peak latency for the Linear pitch contour compared to T2 and Flat. The offset components (Pc, Nc) show relatively longer latency for the English group across stimuli.

3.2. Latency and amplitude of CPR components derived from the Fz electrode site

3.2.1. Peak Latency

For both language groups, mean peak latencies of CPR components Na, Pb, and Nb increase systematically across stimuli in temporal order of occurrence (Fig. 3, top left). Regardless of stimulus, language groups were indistinguishable as reflected by the Na component (T2: $F_{1,22} = 0.38$, $p = 0.5416$; Linear: $F = 4.10$, $p = 0.0551$; Flat: $F = 0.32$, $p = 0.5799$), indicating that the pitch onset was homogeneous in terms of latency irrespective of language experience. In the case of Pb, the English group exhibited a longer latency than the Chinese group in response to T2 only ($\alpha_{\text{individual}} = 0.0166$; T2: $F_{1,22} = 12.31$, $p = 0.0020$, $\eta_p^2 = 0.36$; Linear: $F = 0.88$, $p = 0.3587$; Flat: $F = 1.39$, $p = 0.2510$). The language group effect means that the Chinese respond faster than nonnative

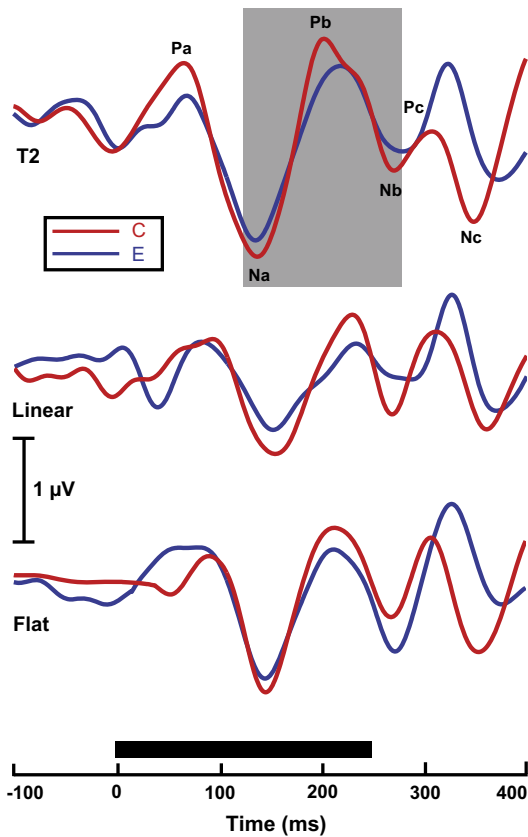


Fig. 2. Grand average waveforms (Chinese, red; English, blue) at the Fz electrode site per stimulus condition. Na, Pb, and Nb (highlighted in gray in the top panel) are the most robust pitch-relevant components. CPR waveforms elicited by the three stimuli (T2, Linear, Flat) show that amplitude of the pitch-relevant components (Na, Pb, Nb) appear to be more robust for the Chinese group, especially in response to T2. Offset components (Pc, Nc) are more robust for the English group, especially for the dynamic pitch stimuli (T2, Linear). Solid black horizontal bar indicates the duration of each stimulus.

speakers only when they are presented with a native pitch contour. As measured by Nb, the longer latency observed in the English group, relative to the Chinese, was elicited in response to Flat only (T2: $F_{1,22} = 1.55$, $p = 0.2268$; Linear: $F = 3.36$, $p = 0.0804$; Flat: $F = 11.98$, $p = 0.0022$).

A direct comparison of peak latencies of T2 and Linear revealed a stimulus main effect for the Na component (Linear > T2: $F_{1,22} = 31.69$, $p < 0.0001$, $\eta_p^2 = 0.59$); for the Pb component, both stimulus and group main effects (Linear > T2: $F = 59.82$, $p < 0.0001$, $\eta_p^2 = 0.73$; English > Chinese: $F = 8.73$, $p = 0.0073$, $\eta_p^2 = 0.28$). The stimulus effect for Na and Pb indicates that linear rising pitch with a fixed rate of acceleration takes longer to process than a curvilinear pitch with a time-varying rate. The group effect means that peak latencies of the English group are longer than the Chinese regardless of stimulus. No effects reached significance for the Nb component, meaning that its peak latencies were invariant across language groups and pitch stimuli.

3.2.2. Interpeak latency

Interpeak latency analysis was limited to the Na–Pb interval because changes in peak latency across stimuli and between groups were observed in response to Na and Pb, but not Nb. The mean Na–Pb interval was shorter for the Chinese group compared to the English group in response to T2 and Linear (Fig. 3, bottom left; $\alpha_{\text{individual}} = 0.0166$; T2: $F_{1,22} = 17.85$, $p = 0.0003$, $\eta_p^2 = 0.45$; Linear: $F = 7.26$, $p = 0.0132$, $\eta_p^2 = 0.25$). This was primarily due to the relatively shorter latency for Pb compared to Na, suggesting

enhanced sensitivity of the Chinese group to rapidly-changing rising pitch contours. A direct comparison of interpeak latencies of T2 vs. Linear showed a group main effect for the Na–Pb component (Fig. 3, bottom left; English > Chinese: $F_{1,22} = 22.73$, $p < 0.0001$, $\eta_p^2 = 0.51$). The stimulus main effect was marginally significant (Linear > T2: $F = 4.26$, $p = 0.0510$). This result points to a language-dependent effect. T2 is native; Linear, albeit phonetically similar to T2, is nonnative. There were no significant language group effects elicited by the Flat stimulus, as measured by Na–Pb ($F = 0.52$, $p = 0.4787$).

3.2.3. Peak-to-peak amplitude

Language group effects on peak-to-peak amplitude of CPR components Na–Pb and Pb–Nb in response to the three pitch stimuli (T2, Linear, Flat) are displayed in Fig. 3. For Na–Pb, Chinese exhibited greater peak-to-peak amplitude than English in response to the native pitch contour only (Fig. 3, top right; $\alpha_{\text{individual}} = 0.0166$; T2: $F_{1,22} = 2.62$, $p = 0.0156$, $\eta_p^2 = 0.11$; Linear: $F = 1.53$, $p = 0.2285$; Flat: $F = 3.06$, $p = 0.0942$). No language group effects were observed for the Pb–Nb component (Fig. 3, bottom right; T2: $F_{1,22} = 2.39$, $p = 0.1362$; Linear: $F = 1.89$, $p = 0.1835$; Flat: $F = 3.06$, $p = 0.0942$).

A direct comparison of peak-to-peak amplitudes of T2 vs. Linear revealed both group and stimulus main effects for the Na–Pb component (Fig. 3, top right; Chinese > English: $F_{1,22} = 5.61$, $p = 0.0271$, $\eta_p^2 = 0.20$; T2 > Linear: $F_{1,22} = 19.22$, $p = 0.0002$, $\eta_p^2 = 0.47$). The absence of a group \times stimulus interaction indicates that Chinese listeners' superiority in processing of dynamic pitch extends even to linear rising ramps that do not occur in natural speech. For the Pb–Nb component, the stimulus main effect was significant (Fig. 3, bottom right; T2 > Linear: $F_{1,22} = 5.34$, $p = 0.0306$, $\eta_p^2 = 0.20$); the group effect, however, was only marginally significant (Chinese > English: $F_{1,22} = 3.97$, $p = 0.0588$). The group \times stimulus interaction was not significant for either component. The fact that T2 elicits greater amplitude than Linear for both components, regardless of language experience, points to the ecological relevance of dynamic, curvilinear pitch contours in natural speech.

3.3. Amplitude of CPR components derived from frontal (F3/F4) and temporal (T7/T8) electrode sites

3.3.1. T2, Linear, Flat

The 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 at frontal (F3/F4: Appendix B.2, Fig. S2) and temporal (T7/T8: Fig. 4) electrode sites. At the frontal sites, the waveforms reveal that regardless of language group, pitch-related components at frontal sites essentially overlap between F3 and F4 with no discernible difference in magnitude (left) and show essentially identical spectrotemporal plots (right). There is no evidence of a hemispheric preference in the frontal lobe. In contrast, the waveform data in Fig. 4 reveal that these same components are larger at the right (T8) than the left (T7) temporal electrode in response to T2 for the Chinese group only (left). The robust rightward preference for T2 is clearly evident in the spectrotemporal plots (right). Results of ANOVAs of peak-to-peak amplitudes of T2, Linear, and Flat separately at both frontal (F3/F4) and temporal (T7/T8) sites are displayed in Appendices B.3 (Fig. S3) and B.4 (Fig. S4), respectively.

3.3.2. T2 vs. Linear

At the frontal sites (F3/F4; Appendix B.5, Fig. S5), a direct comparison of peak-to-peak amplitudes of T2 vs. Linear yielded group and stimulus main effects for Na–Pb (Chinese > English, $F_{1,22} = 11.41$, $p = 0.0027$, $\eta_p^2 = 0.34$; T2 > Linear, $F = 10.47$, $p = 0.0038$, $\eta_p^2 = 0.32$). Similarly, for Pb–Nb, the stimulus main effect was significant (T2 > Linear, $F = 10.76$, $p = 0.0034$, $\eta_p^2 = 0.33$); the group main effect,

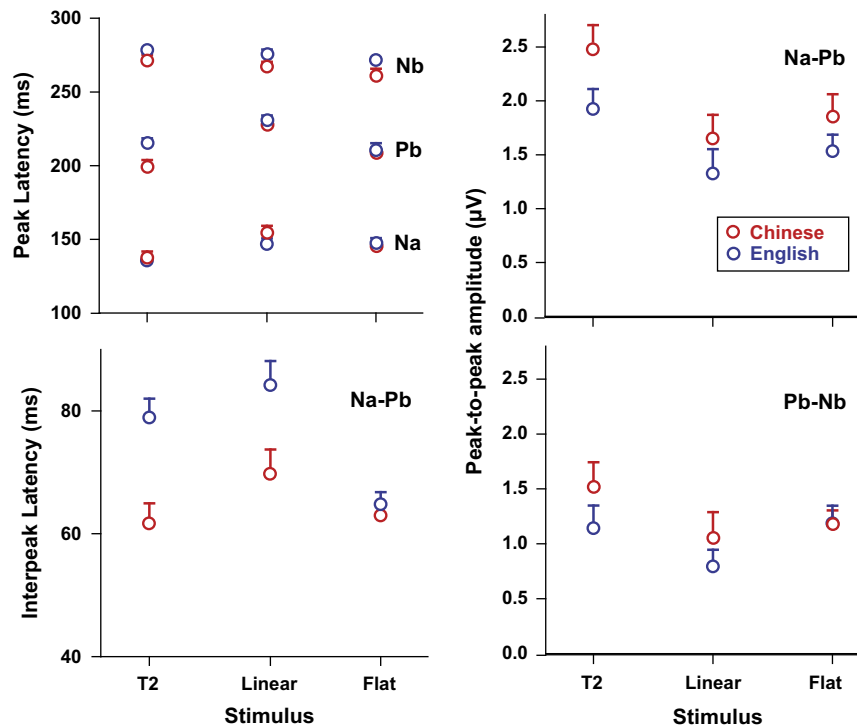


Fig. 3. Mean peak latency (top left) of CPR components (Na, Pb, Nb), and interpeak latency (bottom left) of Na–Pb extracted from Fz as a function of stimulus. Interpeak latency of the Na–Pb component is longer in English than Chinese listeners in response to dynamic pitch stimuli (T2 and Linear). No group effects are observed in response to Flat. Mean peak-to-peak amplitude of CPR components extracted from Fz as a function of stimulus (Na–Pb, top right; Pb–Nb, bottom right). Chinese listeners show greater peak-to-peak amplitude of the Na–Pb component than English in response to the native pitch contour only. No language group effects are observed in response to Flat regardless of component. Error bars = ± 1 SE.

however, was only marginally significant (Chinese > English, $F = 4.23$, $p = 0.0518$, $\eta_p^2 = 0.16$). Neither the hemisphere main effect nor two- and three-way interactions was significant. These data pooled across hemispheres indicate that Chinese amplitude at frontal sites is greater than English, especially for Na–Pb. The stimulus effect (T2 > Linear) suggests that the sensory-level CPR response components interact with higher-level language-related processes.

At the temporal sites (T7/T8; Fig. 5), we observe interactions between group, stimulus, and hemisphere. Results for the Na–Pb component revealed two significant interactions (group \times hemisphere: $F_{1,22} = 7.42$, $p = 0.0124$, $\eta_p^2 = 0.25$; group \times stimulus: $F_{1,43} = 5.90$, $p = 0.0194$, $\eta_p^2 = 0.12$). Regarding the group \times hemisphere interaction, simple effects by group showed a right-sided preference (T8 > T7) in the Chinese group only; by hemisphere, Na–Pb amplitude in the RH was greater in Chinese than English. As for the group \times stimulus interaction, simple effects by group showed that T2 evoked greater amplitude than Linear for Chinese only; by stimulus, Na–Pb amplitude elicited by T2, but not Linear, was greater in Chinese than English listeners. Results for the Pb–Nb component, on the other hand, revealed a significant three-way interaction (group \times hemisphere \times stimulus: $F_{1,41} = 7.97$, $p = 0.0073$, $\eta_p^2 = 0.16$). A priori comparisons by group-and-hemisphere indicated that T2 evoked greater amplitude than Linear at the right temporal site for Chinese. By hemisphere-and-stimulus, Pb–Nb amplitude elicited by T2 was greater in Chinese listeners relative to English. Taken together, these data provide evidence in support of a language-dependent (Chinese > English), right-sided advantage for early cortical pitch processing of native lexical tones (T2 > Linear) in the temporal lobe.

4. Discussion

The major findings of this cross-language study demonstrate that both the latency and magnitude of CPR components are differ-

entially modulated by (i) the overall trajectory of pitch contours (time-varying vs. steady-state), (ii) their pitch acceleration rates (changing vs. constant), and (iii) their linguistic status (lexical vs. non-lexical). Interpeak latency of Na–Pb shows that Chinese are faster than English in response to time-varying (T2, Linear) than steady-state (Flat) pitch. The shorter Na–Pb interpeak latency of the Chinese for time-varying pitch indicates enhanced sensitivity in processing dynamic pitch contours that share the same average rate of acceleration. Chinese show greater peak-to-peak amplitude than English in response to T2 only, as reflected in both Na–Pb and Pb–Nb. A direct comparison between T2 and Linear shows that even though T2, a time-varying pitch contour with changing rate of acceleration, elicits larger amplitude than Linear in both groups, the size of the effect is larger in Chinese than English. These amplitude data provide evidence of interaction with higher-order cognitive/linguistic processes beyond auditory cortex. Our findings further show a language-dependent, right-sided preference in the temporal lobe for processing CPR components. Hemispheric preferences reveal that at the T8 electrode site, amplitude of Na–Pb and Pb–Nb elicited by T2 is larger in Chinese than English; T2 evokes greater amplitude than Linear for Chinese only. By means of the CPR, we are therefore able to demonstrate that Chinese have an enhanced ability in processing dynamic pitch contours with changing rates of acceleration. No group or hemisphere effects are observed in response to stimuli with constant rates of acceleration (Linear, Flat).

4.1. Experience-dependent modulation of pitch as reflected by CPR components

Interpeak latencies are longer in English than Chinese for the Na–Pb component in response to the two dynamic pitch stimuli (T2, Linear). That is, Chinese responses are faster when presented with dynamic pitch contours that share the same

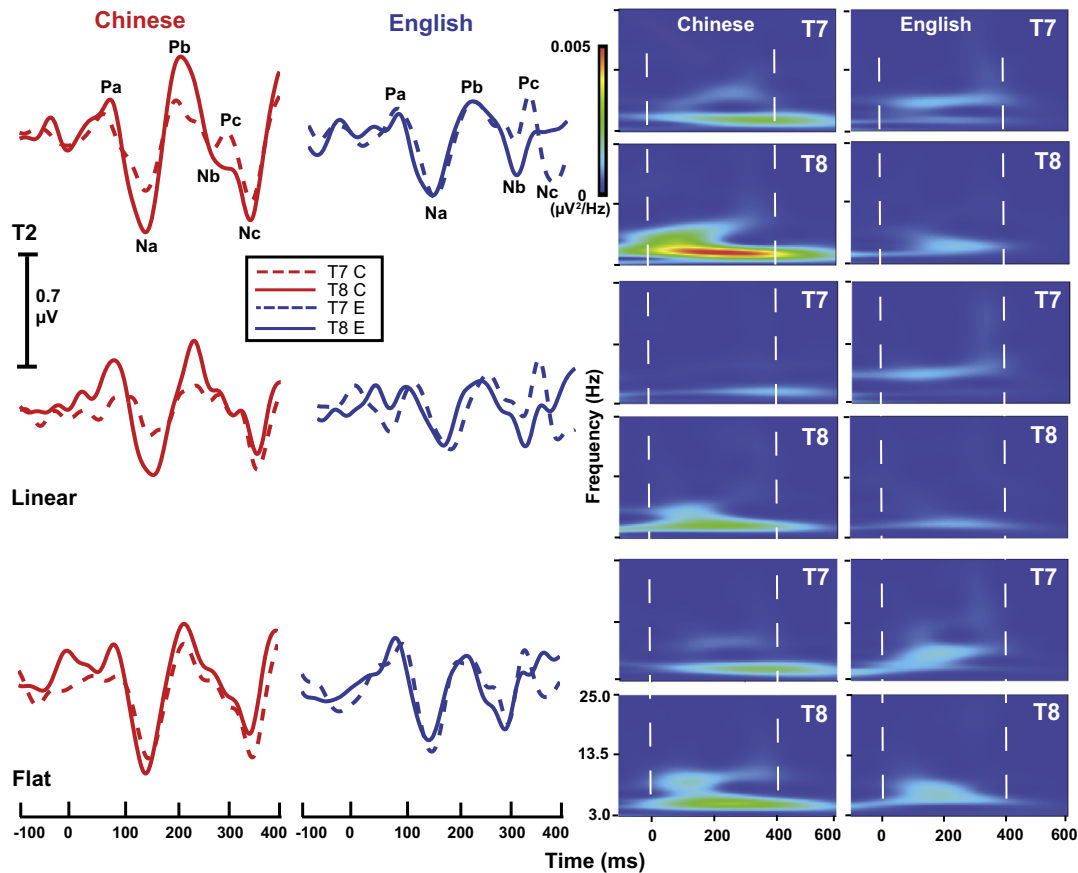


Fig. 4. Grand average waveforms (left) and their corresponding spectra (right) of the CPR components for the two language groups (Chinese, red; English, blue) recorded at electrode sites T7 (dashed) and T8 (solid) for each of the three stimuli (T2, Linear, Flat). CPR waveforms appear to show a right-sided preference (T8 > T7) for the Chinese group in response to dynamic pitch stimuli (T2, Linear). The robust rightward preference for T2 is clearly evident in the spectrotemporal plots. No hemisphere effects are observed in response to Flat for either language group. The zero on the x-axis denotes the time of onset of the pitch-eliciting segment of the three stimuli.

average acceleration rate (0.112 Hz/ms). This shorter Na–Pb interval for the Chinese is mainly due to the shorter peak latency of the Pb component relative to the invariant peak latency of the Na component. Thus, we can isolate Pb as the component that is sensitive to the rapidly-accelerating portion of the pitch contour. This shorter Na–Pb interval for T2 and Linear in the Chinese group may be indexing an increase in behaviorally-relevant sensitivity to rapid changes in pitch via faster integration of neural activity. Na–Pb amplitude is also greater in Chinese than English, but only when presented with a pitch contour representative of a lexical tone (T2). This experience-dependent effect converges with an earlier study in which Na–Pb amplitude is greater in Chinese than English for those curvilinear variants of T2 that more closely approximated its prototypical pitch contour (Krishnan et al., 2014a). In addition to shorter latency, the more robust amplitude for T2 suggests an experience-dependent response enhancement mediated by selective recruitment of neural elements with sharper tuning, greater temporal synchronization, and improved synaptic efficiency to optimally represent the rapidly changing portions of the pitch contour.

A direct comparison of T2 vs. Linear provides evidence in support of language-universal as well as language-dependent effects on modulation of latency and amplitude of CPR components. English latencies are longer than Chinese for Na–Pb in response to both T2 and Linear. We infer that changing acceleration rates, as compared to constant, require longer temporal integration windows for pitch processing regardless of language experience. Overlaid is the effect of language experience. Na–Pb amplitude is greater in Chinese than English, and T2 amplitude is greater than Linear.

Again, we observe that CPR components may capture both experience-dependent effects as well as those that are independent of one's pitch experience.

4.2. Hemispheric preferences in pitch processing vary depending on acoustic and linguistic properties of the stimulus

A strong RH preference is observed at the temporal electrodes (T8/T7) in stark contrast to its absence at the frontal electrodes (F3/F4). It is important to note that our experimental protocol is free of task demands; stimuli are reduced to the pitch parameter only; electrophysiological responses are putatively, pitch-specific; and that hemispheric preference is derived from peak-to-peak amplitude responses extracted from two CPR components (Na–Pb, Pb–Nb). We infer that the temporal preference to the RH reflects selective recruitment of pitch-specific mechanisms in right auditory cortex that are influenced by language experience. This finding converges with an extant literature that attests to the greater role of the RH in the processing of pitch, presumably taking advantage of the finer pitch resolution afforded by the RH (Friederici & Alter, 2004; Hyde, Peretz, & Zatorre, 2008; Meyer, 2008; Poeppel, Idsardi, & van Wassenhove, 2008; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Zatorre & Baum, 2012; Zatorre & Gandour, 2008; Zatorre et al., 2002).

The amplitude of Na–Pb and Pb–Nb is larger in Chinese than English when elicited by T2, but not by Linear. In terms of overall F0 trajectory, both are dynamic. T2 has a changing acceleration rate; Linear, a constant rate. The Linear pitch contour does not occur in the Mandarin tonal space. Indeed, constant rates of pitch

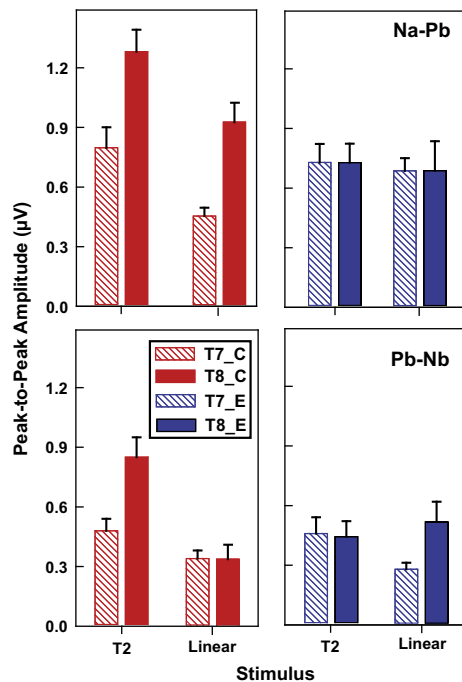


Fig. 5. Mean peak-to-peak amplitude of CPR components (Na–Pb, top row; Pb–Nb, bottom row) extracted from T7/T8 in the temporal lobe as a function of stimulus (T2, Linear) and hemisphere. The amplitude of Na–Pb and Pb–Nb elicited by T2 is larger at the right temporal site (T8 > T7) in the Chinese group only, as well as larger for Chinese relative to English listeners. Moreover, Na–Pb and Pb–Nb amplitude elicited by T2 is greater than Linear at the right temporal site for Chinese only. Error bars = ± 1 SE.

acceleration do not occur in any language of the world. Though T2 and Linear share *average* rate of acceleration, the lack of a group difference, in addition to absence of a RH preference, highlights rightward specialization for processing time-varying, *changing* rates of pitch acceleration that are ecologically representative of linguistic pitch. Previous work on cortical and subcortical responses to linear pitch stimuli similarly fail to show experience-dependent enhancement of pitch-relevant neural activity (Chandrasekaran, Krishnan, et al., 2007, MMN; Xu, Krishnan, & Gandour, 2006, FFR). Steady state or flat F0 patterns are of no linguistic relevance in the speech of any of the world's languages, tonal or otherwise. Consistent with our findings, MEG recordings fail to observe any hemispheric differences with regard to either latency or amplitude of the pitch-relevant cortical components elicited by stimuli with flat pitch (Gutschalk et al., 2004; Krumbholz et al., 2003; Seither-Preisler et al., 2006).

T2 also evokes greater amplitude than Linear at the right temporal site for the Chinese group only. How do we account for the selectivity to T2? We considered two possible explanations: (i) T2 is the only stimulus with a curvilinear pitch contour, i.e., one that is characterized by a changing acceleration rate typical of natural speech; (ii) T2 is the only one with a pitch contour representative of a lexical tone, i.e., one that signals a linguistic function. In a direct comparison of Fz-derived Na–Pb and Pb–Nb amplitude for T2 and Linear, we observe that T2 elicits greater amplitude than Linear for both components, regardless of language experience. What this means is that a curvilinear pitch contour may be a necessary but not a sufficient condition to explain the Chinese advantage for T2 at the right temporal site. This view is supported by recent findings showing RH preference only for T2 and not to other curvilinear approximations of T2 (Krishnan et al., 2014a,b). The second explanation assumes an experience-dependent functional asymmetry that involves interaction between sensory and

higher-order linguistic processes in the auditory cortex. In this study, we cannot tease them apart unambiguously due to the absence of a pitch stimulus that is curvilinear but does not occur in the Mandarin tonal space. An inverted curvilinear T2 stimulus, e.g., a mirror image of T2, would meet those specifications. In previous work at the level of the brainstem (Krishnan, Gandour, Bidelman, & Swaminathan, 2009), we found no group differences in response to the mirror image of T2 as well as two other linear approximations of T2. We therefore predict at the cortical level that language-dependent modulation of pitch extends optimally to curvilinear pitch contours that are representative of citation forms of lexical tones in the Mandarin tonal space. The emergence of an experience-dependent RH preference at this early stage of sensory processing likely reflects a selective recruitment of pitch processes that shows greater precision for optimal representation of behaviorally-relevant pitch attributes.

Indeed, our view is that a complete account of pitch processing must allow for interactions between sensory and cognitive/linguistic contributions that interact within the same time interval, as well as at different time intervals at different cortical levels of the brain. In this study, the time interval occurs at an early, pre-attentive stage of pitch processing in the auditory cortex. The language-dependent effect at the right temporal site suggests that CPR components exhibit heightened sensitivity to pitch contours that are exemplary of lexical tones.

4.3. The notion of 'contour' for real-time pitch processing in the language domain

The definition of 'contour' has been framed previously within the context of perception and production. In both music and speech, a contour is defined by the direction of pitch instead of specific relationships between pitches (Zatorre & Baum, 2012). In music, there are movements up and down in pitch over the course of a melody. In speech, there is a continuous, nonlinear, gliding movement within the pitch range of a syllable or larger unit of connected discourse. Though these definitions are acceptable for describing behavior, they have nothing to say about how surface changes in direction are generated within the context of real-time, pitch processing in the brain. By virtue of the CPR, we are now able to observe neurobiological correlates of pitch-specific, neural generators that modulate those changes in pitch for syllable-based, lexical tones. In search of a neurobiological definition, we define contour as *changes in rate of acceleration* between pitch onset and offset. From this perspective, it's not the overall shape that counts, but rather the rate of acceleration that changes continuously throughout the time course of a pitch contour (cf. speech production, Prom-on, Xu, & Thipakorn, 2009; Xu, 2001, 2006). In this study, a direct comparison between T2 and Linear shows that even though T2, a dynamic pitch contour with changing rate of acceleration, elicits larger amplitude than Linear in both groups, the size of the effect is larger in Chinese than English. 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. Interestingly, these experience-dependent effects in cortical pitch processing are compatible with evidence on pitch encoding at the level of the brainstem (Krishnan & Gandour, 2009; Krishnan et al., 2012, reviews).

It has been aptly demonstrated that the units of linguistic computation and the units of neurological computation are incommensurable (Poepfel & Embick, 2006). In other words, there is no direct mapping from the fundamental primitives for representation and processing at a given analytic level of linguistics to those at a given analytical level of neurobiology. In the extant literature on lexical

tone, only one set of phonological features has been proposed that grants ontological status to features of *dynamic* pitch contours (Wang, 1967, [contour, rising, falling, convex]). While Wang's features closely correspond with speech perception, they cannot be reduced to fundamental neurobiological units. The CPR fills that void in our knowledge base. At a neurobiological level, the transient components of the CPR represent the output(s) of pitch-specific neural generators that appear to index the neural processing of the temporal attributes of a pitch contour, e.g., pitch onset, pitch acceleration, duration, and sound offset (Krishnan et al., 2014b). Thus, the CPR provides a tool to examine the representation of different temporal attributes of pitch contours and to determine how they are shaped by experience.

4.4. Neural mechanism(s) for early sensory level pitch processing in the auditory cortex

It is generally agreed that lateral Heschl's gyrus is the putative source for the pitch onset component (Na). Generator sources for the remaining pitch-relevant components (Pb, Nb) are unknown and cannot be determined from this study. We speculate that these later components (Na–Pb, Pb–Nb) reflect neural activity from spatially distinct generators that represent later stages of sensory processing, relative to Na, along a pitch processing hierarchy. Whether pitch-relevant information extracted by these neural generators is based on a spectral and/or temporal code is unclear. At subcortical levels up to the midbrain, physiologic and computational modeling data support the possibility of either a purely temporal mechanism or a hybrid mechanism using both spectral and temporal information (Cariani & Delgutte, 1996a, 1996b; Cedolin & Delgutte, 2005; Plack, 2005). Neurons in the primary auditory cortex exhibit temporal and spectral response properties which could enable these pitch-encoding schemes (Lu, Liang, & Wang, 2001; Steinschneider, Reser, Fishman, Schroeder, & Arezzo, 1998). Whether they form a network with pitch-selective neurons to carry out this process warrants further investigation.

It has been suggested that the cortical pitch response represents the integration of pitch information across frequency channels and/or the calculation of pitch value and pitch strength in Heschl's gyrus (Gutschalk et al., 2004). Our findings show experience-dependent sensitivity to acceleration rates in dynamic pitch contours. This differential sensitivity points to a neural mechanism capable of encoding the rapidly-changing portion of the pitch contour. Such mechanism(s) must be able to recruit neurons with narrow tuning properties and good neural synchrony to be able to represent rapid changes in pitch.

4.5. Conclusions

The differential sensitivity of the CPR components to pitch contours reveal both a language universal (acoustic) and an overlaid, language-dependent (linguistic) attribute of pitch processing at the early sensory level processing in the auditory cortex. This latter attribute preferentially recruits the right hemisphere to take advantage of its higher precision of pitch processing necessary to represent the perceptually relevant, rapidly-changing portions of native pitch contours. 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, hierarchical pitch processing in the auditory cortex, and reflects an interaction with higher-order, cognitive/linguistic processes beyond auditory cortex. The components of the CPR provide a series of robust neurobiological markers that index processing of temporal attributes of dynamic pitch contours that are differentially sensitive and shaped by language experience.

Acknowledgments

Research supported by NIH 5R01DC008549 (A.K.). Thanks to Longjie Cheng for her assistance with statistical analysis (Department of Statistics); Jilian Wendel and Kate Geisen for their help with data acquisition; and Venkatakrisnan Vijayaraghavan with computer programming. Reprint requests should be addressed to Ananthanarayan Krishnan, Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, IN 47907-2038, USA, or via email: rkrish@purdue.edu.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bandl.2014.09.005>.

References

- Alku, P., Sivonen, P., Palomaki, K., & Tiitinen, H. (2001). The periodic structure of vowel sounds is reflected in human electromagnetic brain responses. *Neuroscience Letters*, 298(1), 25–28.
- Barker, D., Plack, C. J., & Hall, D. A. (2012). Reexamining the evidence for a pitch-sensitive region: a human fMRI study using iterated ripple noise. *Cerebral Cortex*, 22(4), 745–753. <http://dx.doi.org/10.1093/cercor/bhr065>.
- Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054), 1161–1165. [nature03867](http://dx.doi.org/10.1038/nature03867) [pii] 10.1038/nature03867.
- Besson, M., Chobert, J., & Marie, C. (2011). Language and music in the musician brain. *Language and Linguistics Compass*, 5(9), 617–634. <http://dx.doi.org/10.1111/j.1749-818x.2011.00302>.
- Cariani, P. A., & Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, 76(3), 1698–1716.
- Cariani, P. A., & Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *Journal of Neurophysiology*, 76(3), 1717–1734.
- Cedolin, L., & Delgutte, B. (2005). Pitch of complex tones: Rate-place and interspike interval representations in the auditory nerve. *Journal of Neurophysiology*, 94(1), 347–362. [01114.2004](http://dx.doi.org/10.1152/jn.01114.2004) [pii] 10.1152/jn.01114.2004.
- Chait, M., Poeppel, D., & Simon, J. Z. (2006). Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cerebral cortex*, 16(6), 835–848. New York, N.Y.: 1991. [doi:bjh027](http://dx.doi.org/10.1093/cercor/bhj027)[pii]10.1093/cercor/bhj027.
- Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, 25(3–4), 195–210.
- Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *NeuroReport*, 18(18), 1963–1967. <http://dx.doi.org/10.1097/WNR.0b013e3282f213c5> 00001756-200712030-00017 [pii].
- Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36(2), 268–294. <http://dx.doi.org/10.1016/j.wocn.2007.06.005>.
- Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89(2), 267–276. [http://dx.doi.org/10.1016/S0093-934X\(03\)00351-1](http://dx.doi.org/10.1016/S0093-934X(03)00351-1) S0093934X03003511 [pii].
- Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149–175.
- Gandour, J. T., & Krishnan, A. (2014). Neural bases of lexical tone. In H. Winkler & P. Padakannaya (Eds.), *Handbook of South and Southeast Asian psycholinguistics* (pp. 339–349). Cambridge, UK: Cambridge University Press.
- Griffiths, T. D., Buchel, C., Frackowiak, R. S., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, 1(5), 422–427.
- Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., et al. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology*, 20(12), 1128–1132. <http://dx.doi.org/10.1016/j.cub.2010.04.044>.
- 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(1), 207–216. <http://dx.doi.org/10.1006/nimg.2001.0949>.
- Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *Neuroimage*, 22(2), 755–766. <http://dx.doi.org/10.1016/j.neuroimage.2004.01.025> S1053811904000680 [pii].
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. New York: Cambridge University Press.
- Huang, T., & Johnson, K. (2011). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, 67, 243–267.

- Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, 46(2), 632–639. S0028-3932(07)00323-5 [pii] 10.1016/j.neuropsychologia.2007.09.004.
- Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, 123(Pt 1), 155–163.
- Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, 35(1), 104–117. <http://dx.doi.org/10.1016/j.wocn.2005.10.003>.
- Koelsch, S. (2012). *Brain & music*. Chichester, UK: Wiley-Blackwell.
- Kraus, N., & Banai, K. (2007). Auditory-processing malleability: Focus on language and music. *Current Directions in Psychological Science*, 16(2), 105–110.
- Krishnan, A., Bidelman, G. M., Smalt, C. J., Ananthakrishnan, S., & Gandour, J. T. (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, 50(12), 2849–2859. <http://dx.doi.org/10.1016/j.neuropsychologia.2012.08.013>.
- Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, 110(3), 135–148. S0093-934X(09)00042-X [pii] 10.1016/j.bandl.2009.03.005.
- Krishnan, A., Gandour, J. T., Ananthakrishnan, S., & Vijayaraghavan, V. (2014a). Language experience enhances early cortical pitch-dependent responses. *Journal of Neurolinguistics*. <http://dx.doi.org/10.1016/j.jneuroling.2014.08.002>.
- Krishnan, A., Gandour, J. T., Ananthakrishnan, S., & Vijayaraghavan, V. (2014b). Cortical pitch response components index stimulus onset/offset and dynamic features of pitch contours. *Neuropsychologia*, 59, 1–12. <http://dx.doi.org/10.1016/j.neuropsychologia.2014.04.006>.
- Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2012). Experience-dependent plasticity in pitch encoding: From brainstem to auditory cortex. *NeuroReport*, 23(8), 498–502. <http://dx.doi.org/10.1097/WNR.0b013e328353764d>.
- Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience-dependent neural representation of dynamic pitch in the brainstem. *NeuroReport*, 20(4), 408–413. <http://dx.doi.org/10.1097/WNR.0b013e3283263000>.
- Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13(7), 765–772.
- Kuhnis, J., Elmer, S., Meyer, M., & Jancke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia*, 51(8), 1608–1618. <http://dx.doi.org/10.1016/j.neuropsychologia.2013.04.007>.
- Lu, T., Liang, L., & Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nature Neuroscience*, 4(11), 1131–1138. <http://dx.doi.org/10.1038/nn737>.
- 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(3), 927–937. <http://dx.doi.org/10.1016/j.neuroimage.2005.10.034>.
- 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(2), 561–571. doi: S1053-8119(07)00462-4 [pii] 10.1016/j.neuroimage.2007.05.040.
- Meyer, M. (2008). Functions of the left and right posterior temporal lobes during segmental and suprasegmental speech perception. *Zeitschrift für Neuropsychologie*, 19(2), 101–115.
- Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102(3), 1864–1877.
- Moreno, S., & Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research*, 308, 84–97. <http://dx.doi.org/10.1016/j.heares.2013.09.012>.
- Munte, T. F., Altenmüller, E., & Jancke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews: Neuroscience*, 3(6), 473–478. <http://dx.doi.org/10.1038/nrn843>.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Patel, A. D., & Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, 11(9), 369–372.
- 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(4), 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. *Journal of Neuroscience*, 24(30), 6810–6815. <http://dx.doi.org/10.1523/JNEUROSCI.0383-04.2004> [pii].
- Plack, C. J. (2005). *Pitch: Neural coding and Pitch perception*. New York: Springer.
- Poeppel, D., & Embick, D. (2006). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 103–118). Mahwah, NJ: Lawrence Erlbaum.
- Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 363(1493), 1071–1086. doi: TM425571U1117682 [pii] 10.1098/rstb.2007.2160.
- Prom-on, S., Xu, Y., & Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America*, 125(1), 405–424. <http://dx.doi.org/10.1121/1.3037222>.
- Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language*, 110(2), 81–94. <http://dx.doi.org/10.1016/j.bandl.2008.12.001>.
- Ritter, S., Gunter Dosch, H., Specht, H. J., & Rupp, A. (2005). Neuromagnetic responses reflect the temporal pitch change of regular interval sounds. *NeuroImage*, 27(3), 533–543. <http://dx.doi.org/10.1016/j.neuroimage.2005.05.003>.
- Schonwiesner, M., & Zatorre, R. J. (2008). Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Experimental Brain Research*, 187(1), 97–105. <http://dx.doi.org/10.1007/s00221-008-1286-z>.
- Seither-Preisler, A., Patterson, R., Krumbholz, K., Seither, S., & Lutkenhoner, B. (2006). Evidence of pitch processing in the N100m component of the auditory evoked field. *Hearing Research*, 213(1–2), 88–98. <http://dx.doi.org/10.1016/j.heares.2006.01.003>.
- Soeta, Y., & Nakagawa, S. (2008). The effects of pitch and pitch strength on an auditory-evoked N1m. *NeuroReport*, 19(7), 783–787. <http://dx.doi.org/10.1097/WNR.0b013e328282fe2085>.
- Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., & Arezzo, J. C. (1998). Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. *Journal of the Acoustical Society of America*, 104(5), 2935–2955.
- Swaminathan, J., Krishnan, A., & Gandour, J. T. (2008). Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. *NeuroReport*, 19(11), 1163–1167. 10.1097/WNR.0b013e3283088d31 00001756-200807160-00017 [pii].
- 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 Transactions on Biomedical Engineering*, 55(1), 281–287. <http://dx.doi.org/10.1109/TBME.2007.896592>.
- Tervaniemi, M., Kruck, S., De Baene, W., Schroger, E., Alter, K., & Friederici, A. D. (2009). Top-down modulation of auditory processing: Effects of sound context, musical expertise and attentional focus. *The European Journal of Neuroscience*, 30(8), 1636–1642. <http://dx.doi.org/10.1111/j.1460-9568.2009.06955.x>.
- Tsang, Y. K., Jia, S., Huang, J., & Chen, H. C. (2011). ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height. *Neuroscience Letters*, 487(3), 268–272. S0304-3940(10)01374-1 [pii] 10.1016/j.neulet.2010.10.035.
- Wang, W. S.-Y. (1967). Phonological features of tone. *International Journal of American Linguistics*, 33(2), 93–105.
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, 156, 249–268. S0079-6123(06)56013-3 [pii] 10.1016/S0079-6123(06)56013-3.
- Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(4), 565–585.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61–83.
- Xu, Y. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33, 319–337.
- Xu, Y. (2006). Tone in connected discourse. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (Vol. 12, 2nd ed., pp. 742–750). Oxford, UK: Elsevier.
- Xu, Y., Krishnan, A., & Gandour, J. T. (2006). Specificity of experience-dependent pitch representation in the brainstem. *NeuroReport*, 17(15), 1601–1605. <http://dx.doi.org/10.1097/01.wnr.0000236865.31705.3a 00001756-200610230-00008> [pii].
- Yrttiäho, S., Tiitinen, H., May, P. J., Leino, S., & Alku, P. (2008). Cortical sensitivity to periodicity of speech sounds. *Journal of the Acoustical Society of America*, 123(4), 2191–2199. <http://dx.doi.org/10.1121/1.2888489>.
- Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *Journal of the Acoustical Society of America*, 84(2), 566–572.
- Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 363(1493), 1087–1104. 2009, doi: J412P80575385013 [pii] 10.1098/rstb.2007.2161 2161 Reprinted in B.C.J. Moore, L.K. Tyler, & W. Marslen-Wilson (Eds.), *The perception of speech: From sound to meaning* (pp. 275–304). Oxford University Press.
- Zatorre, R. J., & Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biology*, 10(7), e1001372. <http://dx.doi.org/10.1371/journal.pbio.1001372>.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, 6(1), 37–46.