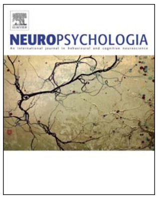
Neuropsychologia 109 (2018) 140–148

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Testing native language neural commitment at the brainstem level: A cross-linguistic investigation of the association between frequency-following response and speech perception

Luodi Yu[a,b], Yang Zhang[a,b,c,*]

[a] Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, MN 55455, USA
[b] School of Psychology, South China Normal University, Guangzhou 510631, China
[c] Center for Neurobehavioral Development, University of Minnesota, Minneapolis, MN 55455, USA

ARTICLE INFO

Keywords:
Native language neural commitment Theory
Frequency following response
Speech perception
Lexical tones
Garner paradigm

ABSTRACT

A current topic in auditory neurophysiology is how brainstem sensory coding contributes to higher-level perceptual, linguistic and cognitive skills. This cross-language study was designed to compare frequency following responses (FFRs) for lexical tones in tonal (Mandarin Chinese) and non-tonal (English) language users and test the correlational strength between FFRs and behavior as a function of language experience. The behavioral measures were obtained in the Garner paradigm to assess how lexical tones might interfere with vowel category and duration judgement. The FFR results replicated previous findings about between-group differences, showing enhanced pitch tracking responses in the Chinese subjects. The behavioral data from the two subject groups showed that lexical tone variation in the vowel stimuli significantly interfered with vowel identification with a greater effect in the Chinese group. Moreover, the FFRs for lexical tone contours were significantly correlated with the behavioral interference only in the Chinese group. This pattern of language-specific association between speech perception and brainstem-level neural phase-locking of linguistic pitch information provides evidence for a possible native language neural commitment at the subcortical level, highlighting the role of experience-dependent brainstem tuning in influencing subsequent linguistic processing in the adult brain.

1. Introduction

Language experience plays a crucial role in the development of the auditory-perceptual system for speech learning. Experience-dependent changes take place at both cortical and subcortical levels early in life (Dahmen and King, 2007; Johnson, 2001). One influential framework for understanding the underlying brain mechanisms for such changes is the native language neural commitment (NLNC) theory (Kuhl, 2010). The NLNC theory features four main claims supported by evidence from cross-language studies. (1) Early exposure and learning drive the formation and functional properties of the neural architecture dedicated to the detection of auditory and linguistic patterns in the native language, which can predict later language skills (Kuhl et al., 2005; Mamiya et al., 2016). As a result, specific cortical regions in native speakers show more sensitive and efficient processing for native speech contrasts in comparison with non-native speakers (Zhang et al., 2005). (2) NLNC is a self-reinforcing process with perceptual benefits and costs. While NLNC promotes the perception and acquisition of phonological structures that conform to the native language, it interferes with the detection of non-native sound units and linguistic structures (Iverson et al., 2003; Zhang et al., 2005). (3) Although NLNC has its sensitive period early in life, it can be continually shaped by experience throughout the life span. For instance, enriched speech input (e.g., formant exaggeration), which heightens cortical activation in the infant brain (Zhang et al., 2011), has been found to be effective in reshaping the neural circuitry in the adult brain to promote second language learning (Zhang et al., 2009). (4) NLNC relies on the perception-production link to facilitate speech communication, and this process takes place by integrating the multimodal information in a structured context of social interaction (Imada et al., 2006; Kuhl, 2007; Ferjan Ramirez and Kuhl, 2017; Stevens and Zhang, 2014).

While previous studies in line with the NLNC have largely focused on cortical processing of phonemic representations (e.g., /r/ vs /l/), some studies have found supporting evidence at the suprasegmental level. For instance, when compared with non-tonal language users, native speakers of tonal languages such as Chinese and Thai show enhanced cortical responses to perceptually challenging but linguistically relevant pitch contrasts (Chandrasekaran et al., 2007; Krishnan et al.,

[*] Corresponding author at: Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, MN 55455, USA.
E-mail address: zhanglab@umn.edu (Y. Zhang).

https://doi.org/10.1016/j.neuropsychologia.2017.12.022
Received 20 March 2017; Received in revised form 28 November 2017; Accepted 11 December 2017
Available online 12 December 2017

2014; Xi et al., 2010) and better discriminative ability for tones in the native speech context (Lee et al., 1996).

Although the NLNC theory provides insights into experience-driven perceptual reorganization at the cortical level (Zhang and Wang, 2007), it remains unknown whether NLNC could involve lower-level brain structures as well. For the current investigation, we are particularly interested in testing NLNC using the frequency following response (FFR) for pitch encoding at the brain stem level. The FFR is widely considered to be generated in the inferior colliculus of the midbrain. Evidence for this structural origin came from lesion studies on animals and humans, imaging studies using source localization, and distinctive signal characteristics between subcortical and cortical responses including amplitude (nanovolts vs. microvolts), latency (5–10 vs. > 50 ms), and phase-locking frequency (> 100 vs. < 100 Hz) (for a review, see Chandrasekaran and Kraus, 2010). Notably, cross-language studies have consistently reported effects of language experience on pitch encoding with stronger representation of lexical tones in favor of the native Chinese speakers (Krishnan et al., 2010b, 2010c, 2009a, b, 2005; Swaminathan et al., 2008a). Unlike the speech-specific NLNC effects at the cortical level (Bidelman et al., 2011b; Zhang et al., 2005), however, subcortical tuning effects for pitch encoding as a result of language/music learning has been shown to be domain-general. For example, compared with speakers of non-tonal languages (such as English), native Chinese speakers show enhanced FFRs for pitch encoding of both speech and non-speech stimuli (Krishnan et al., 2009a). Similarly, pitch strength computed from the FFR using an autocorrelation algorithm was found to be more robust in musicians than non-musicians for both speech (lexical tones) and non-speech (musical interval) stimuli (Bidelman et al., 2011a). Moreover, the degrees of FFR enhancement for lexical tone encoding in musicians had been found to be correlated with the amounts of musical training (Wong et al., 2007).

Further studies suggest that the seemingly domain-general (i.e., non-language-specific) FFR enhancement is constrained by the listener's prior experience. For example, the experience-dependent enhancement of FFR pitch tracking in tonal language speakers was found to occur only within frequency sections that contain characteristic dynamic contours of lexical tones (Krishnan et al., 2009a, b; Swaminathan et al., 2008a; Xu et al., 2006), and FFR enhancement in musicians was found only within the frequency range corresponding to the onset of a musical note (Bidelman et al., 2011a). A recent cross-language study by Intartaglia et al. (2016) reported even more subtle FFR differences in the native speakers of non-tonal languages. In that study, the FFRs in native English speakers showed better representation of fundamental frequency (F_0) than native French speakers, which may reflect the relative importance of pitch in a stress-timed language like English as against a syllable-timed language like French. Meanwhile, both subject groups showed better FFR representation of the first formant (F1) for syllables in their native language in comparison with the non-native language, which may reflect the importance of the F1 cue for vowel identification in both languages.

Despite the mounting evidence of experience-dependent FFR enhancement, few studies have found a direct correlation between behavioral expertise and the FFR measure. While years of musical training have been identified as a major contributor (Musacchia et al., 2007, 2008), the FFRs in musicians did not correlate with musical aptitude or basic pitch discrimination (Musacchia et al., 2007). Auditory training studies have reported improved brainstem representation of speech, but many studies did not find a direct association between the effects in the FFRs and those at the behavioral level (Chandrasekaran et al., 2012; Russo et al., 2005; Song et al., 2008). Bidelman et al. (2011b) showed that although the musicians and Chinese speakers both had stronger FFR encoding of musical pitch relative to the English-speaking nonmusicians, only the musicians displayed advantageous behavioral performance of pitch discrimination, which correlated with the FFR measure. The absence of perceptual benefit for basic pitch discrimination and the lack of significant correlation with FFR in the Chinese speakers suggested a possible dissociation between brainstem sensory coding and perception. Moreover, categorical speech perception in the neural responses has been found to emerge no earlier than 150 ms post-stimulus, and the FFRs occurring at an earlier time window represent continuous (as opposed to categorical) sensory code (Bidelman et al., 2013). These data suggest that the FFR reflects the physiological extraction of F_0 characteristics in the acoustic stimuli rather than endogenous cognitive processing for stimulus categorization (Bidelman, 2017). While some studies have demonstrated a positive relationship between FFR for pitch encoding and behavioral performance as a result of short-term perceptual training (Carcagno and Plack, 2011; Intartaglia et al., 2017), it remains unclear whether similar correlations can be found to index effects of long-term language experience at the brainstem level.

If the FFR is tuned for linguistic purpose, we would expect that this subcortical response can at least partly account for native vs. non-native perceptual differences in a cross-language study. More specifically, if FFR enhancement reflects a form of subcortical NLNC that promotes the perception of linguistic structures (Kuhl, 2010; Zhang et al., 2005), we would expect to see a strong correlation between FFR pitch tracking and behavioral measures of lexical tone processing in tonal language users but not in non-tonal language users. Examination of this hypothesis will help elucidate the exact implication of “clearer” FFR signal in individuals with perceptual expertise for pitch processing (Krishnan et al., 2009a, b; Weiss and Bidelman, 2015), which can provide evidence for the involvement of both subcortical and cortical structures in the process of NLNC. As previous studies tend to focus on group-level differences in the FFRs (e.g., musicians vs. nonmusicians, tonal-language speakers vs. non-tonal language speakers), there are methodological limitations for establishing the direct association between the FFR and higher-order cognitive/linguistic behavior. For instance, the lack of brainstem-behavioral correlation might be due to ceiling-level (or floor-level) performance (Song et al., 2008). When implementing behavioral tasks to address the potential FFR-behavior relationship, it can be challenging to make a distinction between the mechanisms for language-specific processing and those for general auditory processing. Moreover, processing the pitch information can be confounded by other informational dimensions in the auditory stimuli. Selectively attending to task-relevant pitch processing may enhance the cortical activation for pitch encoding (Rao et al., 2010). Particularly, listeners with linguistic/musical expertise may recruit specialized cortical networks for linguistic or musical processing (Baumann et al., 2008; Zatorre and Gandour, 2008), which may override the subcortical sensory contribution to the behavioral performance (Bidelman, 2017).

One way to avoid behavioral ceiling/floor effect for a cross-language study is to require the listeners to attend to informational dimensions other than the lexical tones in the speech stimuli such that the perceptual interference from lexical tone knowledge on the listener's detection and perception of the target dimension can be measured. This interference paradigm was first introduced by Garner and Felfoldy (1970), and adapted by Repp and Lin (1990), among others, to test how lexical tone knowledge might interfere with classification of consonants and vowels in simple consonant-vowel (CV) syllables, or vice versa. In the Repp and Lin (1990) study, listeners were instructed to classify /da/ and /du/ with or without the trial-by-trial variation of lexical tones. The Chinese speakers' vowel classification speed was found to be more affected by the lexical tone variation than that of the English speakers. This pattern of language-specific integrity between tones and vowels in tonal language users is likely due to committed cortical mechanisms for lexical representations that bind the phonemes and tonemes at the syllable level (Tong et al., 2008; Ye and Connine, 1999). Thus the Garner interference paradigm allows testing the automaticity of higher-order phonological processing of lexical tones cross-linguistically without requiring the listeners to attend to the tonal aspect of the stimuli. For the current study of testing NLNC at the brainstem level, we would expect a strong correlation between Chinese listeners' FFR pitch tracking ability and higher-order automatic encoding of lexical tone as tapped by the interference effect of lexical tones on vowel classification.

Our cross-linguistic experimental design in the Garner paradigm could be further strengthened by having a control stimulus condition with an informational dimension that does not depend on the knowledge of lexical tones. For this purpose, we chose to test vowel duration estimation. Unlike vowel identity, vowel duration in the CV syllabic context does not serve a phonemic role in either Chinese or English. It has been shown that F_0 influences perceived vowel duration irrespective of language background such that both tonal and non-tonal language users perceive vowels with a dynamic F_0 as being longer than vowels with a flat F_0 (Lehiste, 1976; Pisoni, 1976; Yu, 2010). Perceived sound duration has also been shown to be positively correlated with pitch height regardless of tonal language experience (Brigner, 1988; Yu, 2010).

There were three research objectives in the current study. First, we aimed to replicate previous FFR findings on experience-dependent subcortical pitch encoding. That is, Chinese speakers would show enhanced FFR pitch tracking for lexical tones in comparison with English speakers. We chose to use Tone 2 (rising) and Tone 3 (dipping) in Mandarin Chinese for FFR recording as they had been shown to have more reliable results than Tone 1 and Tone 4 (Krishnan et al., 2004). To examine domain-general transfer of learning, we employed naturally spoken Chinese syllables as well as non-speech noise that carried the pitch contours of the target lexical tones.

Second, we aimed to test behaviorally the language-specific interference of lexical tones on vowel perception in contrast with the language-general interference effect of pitch contour/height on perceived sound duration. The use of the Garner interference paradigm allowed a systematic assessment of how well adult Chinese and English listeners detected vowel identity and duration with or without tonal variations in the stimuli. In the vowel classification task, we expected to see greater interference (or rather reduced efficiency) in the native speakers of Chinese due to their language-specific integration of tones in lexical representation and access. In the duration estimation task, we expected that both subject groups would rate Tone 2 as being longer than Tone 3 due to perceptual lengthening of sounds with rising pitch contour and higher pitch height.

Third, we aimed to test the effects of native language neural commitment at the subcortical level by conducting a correlation analysis between the FFRs and behavioral measures in the two subject groups. We hypothesized that if the FFR was tuned in service of higher-order linguistic function in accordance with the NLNC, we would see a significant correlation between the FFR pitch tracking ability and the language-specific interference effect in vowel classification task. Specifically, we expected that this correlation would be observed in the native speakers of Chinese but not in the native speakers of English. Alternatively, if no significant correlation was detected in either group or if significant correlation was found in both groups, the data would serve as counter-evidence for NLNC at the subcortical level.

2. Methods

2.1. Participants

Seventeen native Chinese speakers (10 females and 7 males; mean age = 26 years, standard deviation = 5) and nineteen native English speakers (10 females and 9 males; mean age = 23 years, standard deviation = 5) participated and completed the study. They had no speech-language-hearing disorders, and no medical history of neurological disorders, brain injury, or cognitive deficits. The English-speaking subjects had not studied Chinese or other tonal languages. All participants were self-reported non-musicians who had fewer than three years of formal musical training and did not currently play a musical instrument or participate in musical practice on a regular basis. All participants were screened for normal hearing with pure tones (≤ 20 dB HL) at octave frequencies between 250 Hz and 8000 Hz, and had normal or corrected-to-normal vision. Informed consent was obtained from each participant following a protocol approved by the institutional review board at the University of Minnesota.

2.2. Stimuli and FFR test

The participants' FFRs were measured with lexical tones embedded in speech and non-speech noise. For the speech condition, the digitally processed syllables /yi/ spoken with Tone 2 or Tone 3 were recorded from a native male speaker of Mandarin Chinese. The /yi/ syllable was chosen so that the results could be directly compared with previous studies on FFR pitch tracking of lexical tones (Jeng et al., 2016a, b; Krishnan et al., 2005, 2004; Wong et al., 2007). The F_0 ranges of the Tone 2 and Tone 3 were 95–140 Hz and 78–116 Hz, respectively. The F_0 mean and range are within typical variations in Mandarin Chinese (Keating and Kuo, 2012; Xu, 1997). For the non-speech condition, an iterated rippled noise (IRN) with 100 Hz frequency was created using 32 iteration steps, and then the pitch tier was replaced with the F_0 contours extracted from the /yi/ syllables. The use of such an IRN-based carrier is to create stimuli that match the pitch contour of natural speech while eliminating the prominent waveform periodicity and envelope that characterize speech sounds (Swaminathan et al., 2008a, b). All stimulus tokens were digitally processed to have a duration of 250 ms including a 5 ms fade-in/out time.

For the FFR recording session, each of the four sounds was presented 2200 times in separate blocks with jittered inter-stimulus interval (ISI) between 70 and 110 ms. Half of the trials were reversed in polarity to reduce artifacts such as cochlear microphonic (Skoe and Kraus, 2010). The stimuli were presented to both ears through Etymotic Research ER-1 ear inserts at about 65 dB SPL delivered by Tucker-Davis-Technology (TDT) System 3 hardware. The signal delivered by the ER-1 inserts was calibrated using a 1000 Hz calibration tone with a real-ear coupler based on root mean square (RMS) value (Jamieson et al., 2004). The participants were instructed to watch a self-chosen muted video with subtitles and ignore the sound stimuli. The recording was performed in an electrically and acoustically treated booth (ETS-Lindgren Acoustic Systems).

2.3. EEG data acquisition and processing

Continuous electroencephalogram (EEG) signal was recorded using a Biosemi ActiveTwo System at the Multi-Sensory Perception (MSP) Laboratory, Center for Applied & Translational Sensory Science of the University of Minnesota. The sampling rate for signal recording was 16,384 Hz. The recording electrode was placed at Cz and re-referenced to averaged mastoids for off-line analysis. A bandpass FIR (finite impulse response) filter of 30-1000 Hz was applied. Epochs were extracted with 40 ms pre-stimulus baseline and 265 ms after stimulus onset. Trials with instantaneous voltage exceeding ± 35 μV were removed. Accepted trials were then averaged for further analysis.

FFR was obtained using a sliding window approach as implemented in previous studies (Krishnan et al., 2005; Wong et al., 2007). Two measures of brainstem pitch tracking were computed for each tone, namely, autocorrelation and stimulus-to-response correlation. Autocorrelation was performed for each 40 ms bin of the averaged EEG signal for each sound. The autocorrelation procedure correlates the signal with a delayed copy of itself so that periodicity of the signal can be determined from estimations of the time-lag. The 40 ms bin was slid by 1 ms step size, resulting in a total of 225 overlapping bins encompassing FFR from stimulus onset to 265 ms post-stimulus with zero padding. The peak autocorrelation value and the corresponding time lag were computed for each bin, with the lag indicating F_0 of the bin. Peak autocorrelation values were averaged across bins indicating the pitch strength of the whole signal. The same F_0 extraction procedure was applied to the acoustic waveforms for each stimulus. The stimulus-to-response correlation expressed as the Pearson's correlation coefficient r between the stimulus and response F_0 was derived for each stimulus for each participant, indicating FFR pitch tracking accuracy.

2.4. Behavioral tests

The behavioral tests in the Garner paradigm included a vowel classification task and a duration estimation task. The stimuli in the vowel classification task were /da/ and /du/ syllables that carried lexical tone contours for Tone 2 and Tone 3. The /da/-/du/ syllables are common to both Chinese and English languages, and had been used in a number of cross-language studies adopting the Garner interference paradigm (Lee and Nusbaum, 1993; Repp and Lin, 1990; Tong et al., 2008). The /da/ and /du/ stimuli were recorded from the same Chinese speaker who produced the /yi/ syllables for the EEG test. They were digitally edited using a procedure similar to the preparation of the /yi/ sounds. To ensure identical pitch contours of the two vowel classes, F_0 contours of the /du/ syllables were replaced with those extracted from the /da/ syllables. The F_0 ranges of the Tone 2 and Tone 3 were 89–153 Hz and 76–99 Hz, respectively.

The stimuli in the duration estimation task were based on the speech and non-speech stimuli for Tone 2 and Tone 3 in the EEG recordings. The /yi/ syllables and non-speech noise tokens were obtained at 350 ms. Then the sounds were manipulated in duration with the PSOLA (pitch synchronous overlap and add) method with four variants for each stimulus: 150 ms, 250 ms, 350 ms, and 450 ms.

In the vowel classification task, reaction times were recorded and compared between conditions that included lexical tone variation and conditions that did not. The purpose was to investigate degrees of perceptual interference of one dimension on the other. To this end, two conditions (fixed vs. roving) were tested in the current experiment. In the fixed condition, the tone dimension was kept constant with separate stimulus blocks for Tone 2 (block 1) and Tone 3 (block 3). In the roving condition (block 2), a within-block lexical tone variation was introduced so that the tone dimension was changing between Tone 2 and Tone 3 on a trial-by-trial basis. Two blocks of the fixed condition with 40 trials in each and one block of the roving condition with 80 trials were presented with a 500 ms ISI. The stimuli were presented through a Sennheiser headphone at a comfortable level of 65 dB SPL. The participant was instructed to indicate whether the syllable in each trial as /da/ or /du/ by pressing the corresponding button on a keyboard as fast as possible. A visual prompt "'da' or 'du'" was presented in each trial. Three practice trials were given for familiarization with the task before the test session. The additional reaction time in the roving condition compared with the fixed condition was calculated as the amount of dimensional interference of lexical tone on vowel classification.

In the duration estimation task, subjective duration of sound was recorded using a bar-length adjustment procedure. This procedure was adapted from Schlauch et al. (2001). A similar magnitude estimation procedure was used in a recent study in our lab (Zhang et al., 2016b). Specifically, participants were presented with a sound and a bar on the computer screen for each trial, and they were instructed to adjust the length of the bar to match the duration of the sound. For both speech and non-speech conditions, each sound was presented with 5 repetitions, resulting in 40 trials for each stimulus condition. The ISI for the magnitude estimation was set to 1000 ms. Speech and non-speech stimuli were tested in separate blocks. Two practice trials were given to ensure participants' understanding of the bar-length adjustment procedure. The participant was instructed to arbitrarily assign a length for the sound in the first trial as a reference and keep using the same criteria throughout. On the 14-in. laptop computer display screen with a resolution of 1920*1080-pixel, the bar length ranged from 20 pixels to 1440 pixels with a step size for adjustment set at 10 pixels per button press. At the beginning of each trial, the bar length reset to its default of 720 pixels. Responses were recorded as bar length in pixels for each trial.

For all the tests, the sound tokens were RMS normalized in intensity. Sound creation and manipulation were carried out using Matlab R2014b (https://www.mathworks.com/), Praat (Boersma and Weenink, 2014), and SoundForge 9.0 (Sony Creative Software, USA). The tasks were programmed using a Matlab-based toolbox Psychtoolbox-3 (Kleiner et al., 2007) and presented using a Lenovo laptop. All the behavioral tests were conducted in an electrically and acoustically treated booth (ETS-Lindgren Acoustic Systems).

2.5. Statistical analysis

All data analyses were carried out in Matlab R2014b and R (https://www.r-project.org/). For the FFR measures, we performed a mixed 2*2*2 repeated measures ANOVA for pitch tracking strength as measured by autocorrelation and pitch tracking accuracy as measured by stimulus-to-response correlation. The main factors included subject group (Chinese vs. English) as a between-subject variable, stimulus type (speech vs. non-speech) and tone (Tone 2 vs. Tone 3) as within-subject variables. For the vowel classification task, we performed a 2*2 mixed ANOVA model for reaction time with subject group as between-subject variable and condition (fixed vs. roving) as within-subject variable. The reaction time differences between fixed and roving conditions were taken as indicators for the amount of tone interference effect on vowel perception or the degree of vowel-tone integration.

For the duration estimation task, we performed a 2*2*2 mixed model ANOVA for perceived duration. The main factors included subject group (Chinese vs. English) as a between-subject variable and stimulus type (speech vs. non-speech) and tone (Tone 2 vs. Tone 3) as within-subject variables. When significant interactions were observed, post-hoc *t*-tests or one-way ANOVA tests were conducted to examine simple effects.

To examine FFR-behavior relationship, we performed linear mixed-effects (LME) regression analyses. LME regression allows examination of how multiple variables can predict an outcome measure while considering the covariance structure among the repeated-measures predictor variables. This approach has begun to be widely adopted in human neuroscience research (Koerner and Zhang, 2017). In each model, we included FFR pitch strength or accuracy measures of Tone 2 and Tone 3 contours in speech and non-speech as fixed effects with subjects as a random effect. The tone interference effect measured by reaction time difference between the roving and fixed tone conditions was divided by 1000 to match the scale of FFR measures. The LME models were applied separately to the two subject groups.

3. Results

3.1. EEG measures

3.1.1. FFR Pitch strength measured by autocorrelation

As predicted, the repeated measures ANOVA revealed a significant group effect with the Chinese group showing stronger FFR pitch strength than the English group ($F(1,34) = 5.20$, $p < .05$, $\eta^2 = 0.13$) (Fig. 1). There was a significant tone effect ($F(1,34) = 4.18$, $p < .05$, $\eta^2 = 0.21$) and a significant stimulus type*tone interaction ($F(1,34) = 11.23$, $p < .01$, $\eta^2 = 0.24$). Further one-way ANOVA showed that Tone 3 elicited greater FFR pitch strength than Tone 2 in the speech condition ($F(1,35) = 12.11$, $p < .01$, $\eta^2 = 0.26$) but not in the non-speech condition ($F(1,35) = 0.95$, $p = .337$, $\eta^2 = 0.03$).

3.1.2. FFR Pitch tracking accuracy measured by stimulus-to-response correlation

Similar to the results for the autocorrelation measure, the repeated measures ANOVA test revealed a significant group effect ($F(1,34) = 7.04$, $p < .05$, $\eta^2 = 0.17$) (Fig. 2). The results also showed a significant stimulus type effect with the speech sounds eliciting greater FFR pitch tracking accuracy than non-speech sounds ($F(1,34) = 17.37$, $p < .001$,

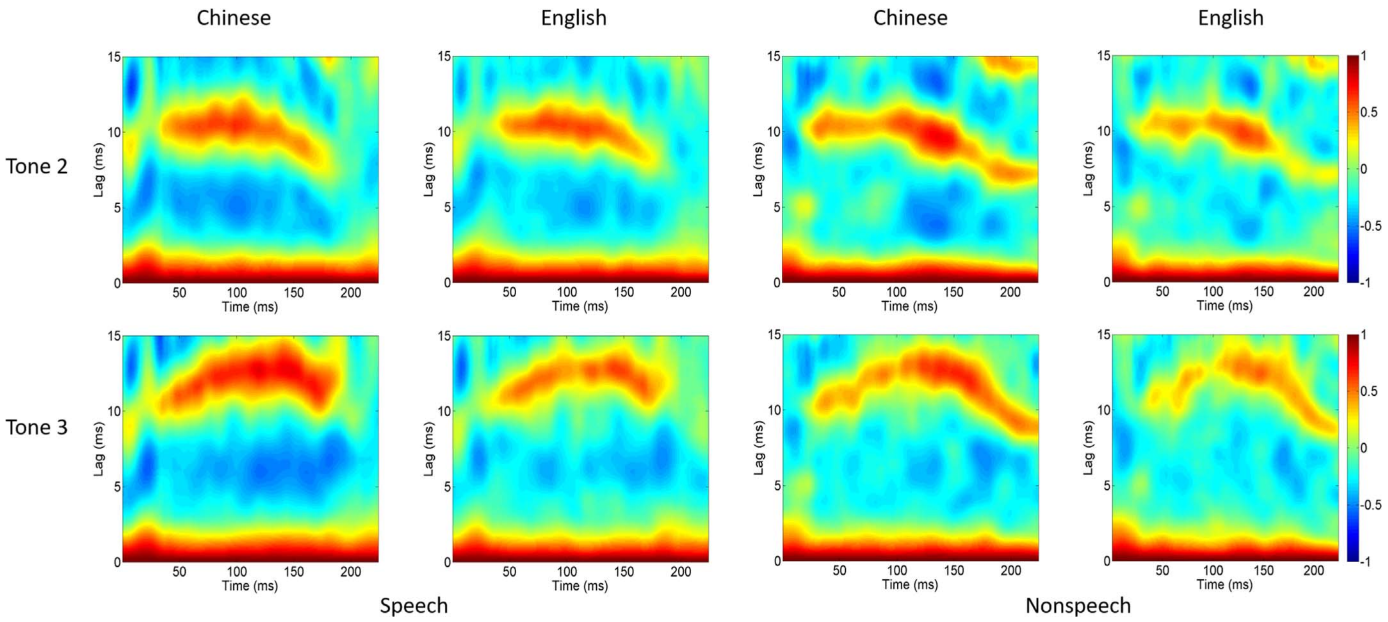


Fig. 1. Autocorrelograms of averaged FFR responses in the Chinese group and the English group elicited by Tone 2 (top) and Tone 3 (bottom) in speech (left) and non-speech (right). The red band curves follow the inverse of the tones (frequency = 1/lag), depicting pitch (F_0) tracking in the FFR.

$\eta^2 = 0.34$). There was also a significant tone effect as Tone 3 elicited greater FFR pitch tracking accuracy than Tone 2 ($F(1,34) = 20.27$, $p < .001$, $\eta^2 = 0.37$). No significant interaction was observed.

3.2. Behavioral measures

For the vowel classification task, longer reaction time in the roving condition compared to the fixed condition indicated lexical tone interference with vowel perception in both subject groups (Fig. 3). The ANOVA results revealed a significant group effect with the Chinese group showing longer reaction times across all stimuli than the English group ($F(1, 34) = 11.92$, $p < 0.01$, $\eta^2 = 0.26$). There were also a significant condition (fixed vs. roving) effect ($F(1, 34) = 41.19$, $p < .001$, $\eta^2 = 0.55$) and a significant group*condition interaction ($F(1,34) = 9.96$, $p < .01$, $\eta^2 = 0.23$). Further one-tailed paired-sample t-test showed that the roving lexical tones significantly slowed down vowel perception in both groups, with a greater effect in the Chinese group ($t(16) = -5.74$, $p < .001$, Cohen's $d = -2.87$) than in the English group ($t(18) = -3.03$, $p < .01$, Cohen's $d = -1.43$).

For the duration estimation (Fig. 4), the ANOVA test revealed a significant group effect, which indicated that the English group used longer arbitrary duration reference (i.e., longer bars to start with) than the Chinese group ($F(1,34) = 6.31$, $p < .05$, $\eta^2 = 0.16$). The results also showed a significant tone effect ($F(1,34) = 61.31$, $p < .001$, $\eta^2 = 0.64$) and a significant stimulus type*tone interaction ($F(1,34) = 5.65$, $p < .05$, $\eta^2 = 0.14$). Further one-way ANOVA for speech and non-speech conditions showed that Tone 2 was perceived as being longer than Tone 3, with a greater effect in the non-speech condition ($F(1,35) = 46.99$, $p < .001$, $\eta^2 = 0.57$) compared to in the speech condition ($F(1,35) = 33.31$, $p < .001$, $\eta^2 = 0.49$). No significant interaction involving subject group (Chinese vs. English) was found.

3.3. Relationship between behavioral and EEG measures

As predicted, language-specific tone interference effect was observed in the vowel classification task but not in the duration estimation task. LME regression analyses were performed in each subject group with the tone interference effect on vowel classification as outcome variable and FFR pitch strength and pitch tracking accuracy measures as fixed effects (Table 1). The LME models revealed that FFR pitch strength of non-speech Tone 3 was significantly associated with behavioral tone interference effect on vowel classification in the Chinese

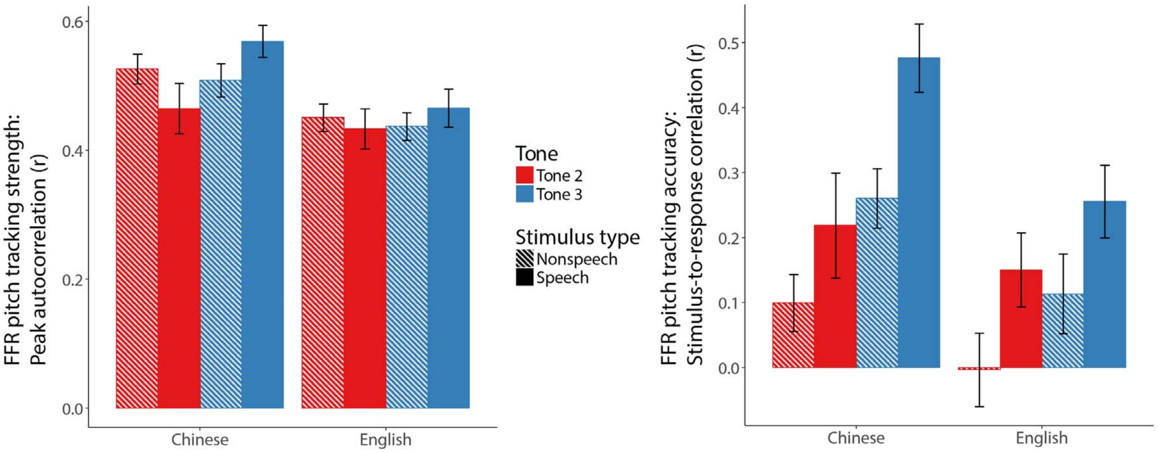


Fig. 2. Bar plots with error bars showing FFR pitch strength measured by autocorrelation coefficient (left), and FFR pitch tracking accuracy measured by stimulus-to-response correlation (right) as a function of tone (Tone 2 vs. Tone 3), stimulus type (speech vs. non-speech), and language group (Chinese vs. English).

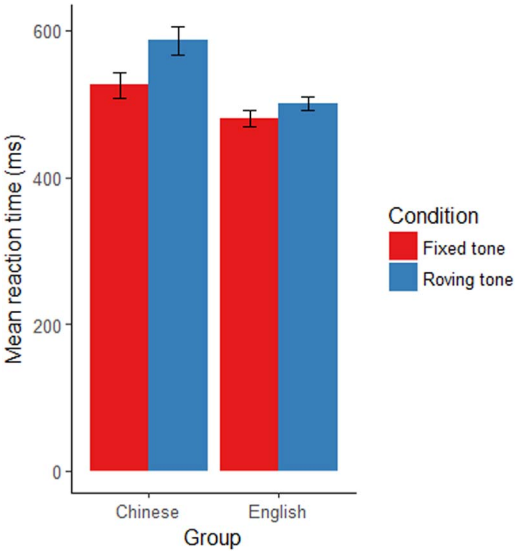


Fig. 3. Bar plot with error bars showing mean reaction times of vowel classification as a function of tone variation (Fixed vs. Roving) in the Chinese and English groups.

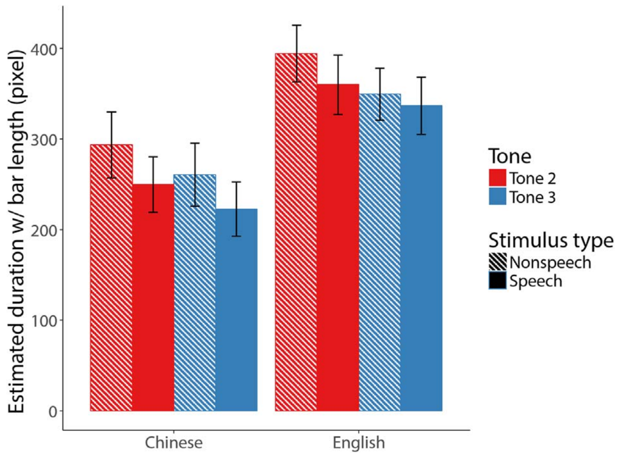


Fig. 4. Bar plot with error bars showing mean perceived duration as functions of tone (Tone 2 vs. Tone 3) and stimulus type (speech vs. non-speech), and language group (Chinese vs. English).

Table 1
F-statistics and regression coefficients (β) of fixed effects in the linear mixed-effects regression models for behavioral tone interference on vowel classification in the Chinese and English subject groups.

Variable	FFR stimulus condition		Chinese		English	
	Speechness	Tone	F	β	F	β
Intercept			45.89***	0.10	8.24*	− 0.01
FFR pitch strength	Speech	2	1.87	− 0.10	0.16	− 0.09
	Speech	3	1.74	0.17	0.99	0.05
	Nonspeech	2	0.14	0.11	0.04	0.01
	Nonspeech	3	**6.51***	**− 0.28**	0.97	0.10
Intercept			31.06***	0.07	8.76*	0.01
FFR Pitch tracking accuracy	Speech	2	0.21	− 0.02	0.34	0.01
	Speech	3	0.06	− 0.02	0.30	0.02
	Nonspeech	2	2.48	0.10	0.81	− 0.03
	Nonspeech	3	0.32	− 0.04	1.73	0.04

* $p < 0.05$.
*** $p < 0.001$.

group. The negative regression coefficient indicates that greater FFR pitch strength of non-speech Tone 3 was correlated with smaller tone interference on vowel perception in the Chinese group. However, no relationship between neural and behavioral response was observed in the English group.

4. Discussion

4.1. Neural coding of lexical pitch contour in FFR

Consistent with previous research, our cross-language data replicated FFR enhancement in pitch tracking strength and accuracy for both speech and non-speech stimuli in the tonal language speakers. This FFR enhancement is thought to reflect the fine-tuned auditory processing in Chinese speakers for stronger neural phase-locking of waveform periodicity and representation of F_0 for lexical tones (Krishnan et al., 2010a, b, c, 2009a, b, 2005; Swaminathan et al., 2008a). Notably, such cross-domain experience-dependent effect for pitch processing is also evident at cortical level (Krishnan et al., 2015, 2014; Xi et al., 2010). Although the exact mechanisms of experience-driven attunement in FFRs and cortical responses are unknown, it has been proposed that experience-dependent neural tuning relies on coordination among ascending, descending, and local pathways of auditory system (Krishnan and Gandour, 2017), with FFR representing a snapshot of such confluence of the hierarchical system on auditory processing (Kraus et al., 2017).

One novel finding of the current study is the language-specific tone effects (Tone 2 vs. Tone 3) with Tone 3 (dipping tone) eliciting stronger neural phase-locking or greater pitch strength of F_0 than Tone 2 (rising tone). The Chinese group had much stronger neural phase-locking with Tone 3 than with Tone 2 whereas the English group showed no discernable difference (Fig. 2). The pitch tracking accuracy measure mirrored the pitch strength measure that tone effect was more pronounced in the Chinese group than in the English group (Fig. 2). These group differences cannot be explained by acoustic differences such as pitch acceleration rate or turning point timing in the tonal stimuli as the two subject groups were given the same tests. According to Krishnan et al. (2010c), F_0 acceleration of contour pitch can influence subcortical encoding of the F_0. That is, FFR pitch strength decreases systematically with increasing F_0 acceleration rate. The pitch contour of our Tone 2 stimuli had an acceleration rate of approximately 0.56 Hz/ms starting from 150 ms, which is more exaggerated than previous studies with a pitch acceleration rate at 0.3 Hz/ms (Krishnan et al., 2010c; Xu, 1997). Our Tone 2 stimuli also had a later onset for the rise of F_0, which typically occurs within 30% of the duration (Shen et al., 1993). By contrast, the pitch contour characteristics (onset, range and turning point) for our Tone 3 stimuli were well within the reported range for Tone 3 in Mandarin Chinese (Jongman et al., 2006; Shen et al., 1993; Xu, 1997). Presumably, acoustic level differences would influence the FFR coding of the pitch contours for Tone 2 and Tone 3 similarly across the two subject groups. In order to account for the language-specific differences in the FFR strength for the two tones, we propose that the FFR may depend on the listener's ability to register subtle deviations from "prototypical" regularities in lexical tones at this early sensory processing stage. Our observation here is consistent with the notion that the FFR reflects one's auditory experience across the life span (White-Schwoch and Kraus, 2017), which is shown in the recent finding that brainstem-level auditory physiology is influenced by prior knowledge about a specific sound pattern (Skoe et al., 2015).

4.2. Language-specific phonological integrity of lexical tone and vowel dimensions

While both subject groups were less efficient in vowel classification with the presence of task-irrelevant trial-by-trial lexical tone variation, this lexical tone interference effect was significantly greater in the Chinese group. This finding suggests the integral relationship between segmental (vowel) and suprasegmental (tone) features at the syllable level in Chinese. The group effect reflects not merely language-independent acoustic interference from one dimension on the other, but language-specific phonological mediation. Our data are consistent with previous work on this point (Lee and Nusbaum, 1993; Repp and Lin, 1990; Tong et al., 2008). For a native Chinese speaker, a change of lexical tone might automatically trigger networks committed to the integrated phoneme/toneme for lexical processing (Tong et al., 2008; Ye and Connine, 1999). Thus the variations in tones would lead to greater interference in the vowel classification task in the Chinese group than the English group.

4.3. Language-general perceptual lengthening of rising tones

In addition to the language-specific effect in the vowel classification task, we observed a language-general tone effect on perceived sound duration. Both subject groups showed longer perceived duration of the rising tone than that of the dipping tone. As the duration cue is not used for phonemic contrasts in either Chinese or English, we expected similar trends of duration estimation in the two subject groups. It has been previously shown that sounds with dynamic pitch are heard as being longer than sounds with flat pitch, independent from listeners' language background (Cumming, 2011; Lehiste, 1976; Pisoni, 1976; Wang et al., 1976; Yu, 2010). Tone duration perception is also influenced by pitch height with a syllable perceived to be longer in a higher pitch level than in a lower pitch level (Brigner, 1988; Lake et al., 2014; Yu, 2010). In the current study, both the rising and dipping tones included dynamic pitch contours, and the rising tone had a higher average F_0 than the dipping tone, which may partially contribute to the perceived duration difference between the two tones. Furthermore, there is a well-established perceptual looming bias with sounds that have rising pitch contour or intensity perceived as longer and louder than sounds with falling pitch contour or intensity (Ghazanfar and Maier, 2009; Zhang et al., 2016b). Our data indicate that the perceptual lengthening of rising tones does not appear to be influenced by language experience.

4.4. Association between FFR and language-specific processing of lexical tone

The correlational analyses revealed a strong association between FFR and behavioral measure of lexical tone interference on vowel classification only in the Chinese group. Stronger FFR pitch tracking for Tone 3 in the non-speech stimuli was correlated with a reduction of lexical tone interference on vowel classification speed in the Chinese listeners. The correlational effect was only found for Tone 3 which showed greater between-group differences in the FFR measures than Tone 2.

Interestingly, the correlation was only found for the non-speech stimuli even though the speech and non-speech stimuli carried identical lexical tone contours. One might wonder why only the FFRs for the non-speech stimuli can predict behavioral speech perception outcome in the current study. This question boils down to which stimulus design (speech vs. non-speech) would be able to better capture the language-specific behavioral interference effect. The potential problem of using speech stimulus in measuring FFR tracking of linguistic pitch has long been noted. Changes in vocal pitch in speech sounds are over-learned multi-dimensional auditory patterns (including linguistic dimensions such as lexical tone and paralinguistic dimensions such as age, gender, voice quality, and emotion), which could introduce additional confounds in FFR coding of other informational dimensions in addition to lexical tone category. In contrast, the non-speech stimuli were novel to the listeners except for the linguistic pitch contour patterns. Krishnan and associates introduced the non-speech noise to eliminate the potential linguistic and paralinguistic confounds from the stimulus (Swaminathan et al., 2008a, b; Krishnan et al., 2009a, b). In the non-speech condition, the FFR encoding of linguistic pitch can be more precisely measured and better distinguished from the processing of other dimensions or the processing of syllable as a whole unit. The degraded temporal regularity of the non-speech stimulus compared with the natural speech sounds can also prevent saturation in neural responses to over-learned auditory patterns. Consequently, the novel non-speech stimuli carrying the linguistic pitch pattern might be advantageous in measuring pitch encoding that is linguistically tuned, which in turn correlated with higher-level linguistic skill.

Our data suggests that language-dependent subcortical neural phase-locking with the F_0 contour is associated with higher-order phonological processing for word recognition. According to the two-stage model of lexical tone processing (Luo et al., 2006), the early stage of processing primarily takes place within the right hemisphere for acoustic analysis, then migrates to the left hemisphere for lexical processing with language-specialized networks. The additional time required for selective processing of vowel categories in the Chinese listeners might reflect the inhibition of automatic processing of lexical tones and additional lexical processing of the speech syllables. One key claim of the NLNC is that neural commitment promotes neural efficiency in detecting linguistically meaningful structures in the native language (Zhang et al., 2005). The observed relationship between the stronger FFR pitch tracking and greater neural efficiency in detecting linguistic units in the Chinese listeners is thus in agreement with the hypothesis of a possible NLNC at the subcortical level.

Why would weaker FFR representation of prototypical lexical tone pitch contour be associated with additional lexical processing for vowel recognition in the Chinese listeners? We suspect that this language-specific correlation might reflect a compensatory mechanism of higher-level brain circuitries to counteract bottom-up signal loss of linguistically meaningful acoustic cues from subcortical sensory coding. In other words, the Chinese listeners with relatively "noisier" FFR pitch representation were also less efficient in vowel classification presumably with a heavier reliance on the higher-order processing based on their lexical tone knowledge. A similar phenomenon is seen in speech-in-noise perception comparing native vs. non-native listeners (Zhang et al., 2016a). In the presence of background noise, native listeners generally display advantage in speech comprehension due to accessibility of higher-order linguistic knowledge to compensate for sensory signal degradation, whereas non-native listeners are more susceptible to noise due to insufficient top-down linguistic influence (Bidelman and Dexter, 2015).

The absence of FFR-behavior correlation in the English group verified part of our original hypothesis for the lack of NLNC for lexical tone processing in non-tonal language users. In English, vowel identification primarily relies on the first and second formants instead of F_0 (Reetz and Jongman, 2011). Unlike the Chinese listeners, the non-tonal language users would not have the same level of neural commitment to the phonological integrity of phonemes and tonemes at the syllable level. The between-group differences here indicate that experience-dependent FFR enhancement in native speakers of a tonal language plays an important role in the robust representation and efficient processing of the linguistic pitch information, which can exert influences on the efficient and language-specific patterns of speech perception.

To our knowledge, this is the first study that used the Garner paradigm to establish a strong association between effects of language experience in FFRs and behavioral outcomes of speech perception in the tonal language users. Our findings coincide with a broad notion that the auditory equivalent of primary visual cortex (V1) for visual processing might not be the primary auditory cortex (A1), but the inferior colliculus (IC) which underlies some of the feature extraction functions (Chandrasekaran and Kraus, 2010; Nelken, 2004). As both cortical (e.g., Zhang et al., 2009) and subcortical functions (e.g., Intartaglia et al., 2017) demonstrate remarkable plasticity for speech learning in adulthood, developing auditory training protocols that target subcortical pitch tracking for linguistic processing may hold promise for second-language learners and individuals with language impairment.

4.5. Limitations and future directions

One limitation of the current design to test the NLNC at the subcortical level is the lack of supportive data from a developmental perspective. As we only tested adults to show the cross-language differences in FFR pitch tracking in relation to speech perception efficiency, we do not have any evidence to make claims about the early development of age-dependent milestones of NLNC at the subcortical level. One key concept with the NLNC is that committed neural architectures are formed in infancy to promote the detection of sound patterns in support of the acquisition of native language (Kuhl et al., 2005; Zhang et al., 2005). Existing literature indicates that the language-specific sensory reorganization happens during early stages of life through corticofugal modulation (Krishnan et al., 2015). However, our NLNC interpretation remains speculative without determining the early language-specific FFR development in parallel with early phonetic learning and language development at the cortical level (Cheour et al., 1998). In this regard, Jeng et al. (2011) conducted the first cross-language study of FFR in neonates and adults. While the adult data confirmed the cross-language FFR differences for lexical tones, no differences were observed between Chinese and American neonates. In a follow-up study, Chinese infants' FFR representation of lexical tone pitch contour showed improvements from birth to 3-month old (Jeng et al., 2016a), suggesting early signs of language learning effects and/or auditory maturation in pitch encoding at the brainstem level. Much more data are needed to verify the cross-language differences in infant FFRs and map out the developmental trajectory or age-dependent milestones of this early language-specific sensory reorganization at both subcortical and cortical levels.

There is also a technical concern regarding whether the FFR exclusively represents neural responses at the subcortical level. For decades, the putative generator of the FFR has been the inferior colliculus (Chandrasekaran and Kraus, 2010). However, a recent magnetoencephalography (MEG) study showed that several regions can contribute to the FFR signal, including subcortical structures and primary auditory cortex (Coffey et al., 2016). Thus, it is possible that the observed Chinese listeners' FFR advantage reflects to some extent cortical-level processing. Future research is needed to determine the neural generators of the FFR signal, the relative weights of subcortical and cortical contributions to the FFR, and its fine-grained relationship with behavioral perception.

5. Conclusions

The current adult cross-language study employed FFR and behavioral measures to test the effects of language experience at the subcortical level and the role of FFR in language-specific speech processing. At the neural level, we replicated the experience-dependent enhancement in the subcortical encoding of lexical tones in the Chinese listeners. At the behavioral level, vowel perception efficiency in the Chinese group was significantly impacted by task-irrelevant lexical tone variation, indicating the influence from the integration of tone and segmental information in native speakers of tonal languages. More importantly, vowel perception efficiency in the Chinese listeners was directly associated with their FFR pitch tracking ability, whereas such a brain-behavior correlation was absent in the English listeners. The results suggest that in addition to enhanced neural coding of the acoustic pitch information, FFR is shaped by language experience in service of higher-order linguistic function. This finding provides the initial evidence for the influences of native language neural commitment (NLNC) at the subcortical level.

Acknowledgements

This work was funded in part by the Graduate Research Partnership Program Fellowship (LY), the Grand Challenges Exploratory Research Grant (YZ) and Brain Imaging Research Project Award (YZ) at the University of Minnesota, and a grant from the Natural Science Foundation of China (YZ, NSFC 31728009). We thank Drs. Dorea Ruggles and Andrew Byrne at the Multisensory Perception Laboratory for their assistance in EEG data collection.

References

Baumann, S., Meyer, M., Jäncke, L., 2008. Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. J. Cogn. Neurosci. 20 (12), 2238–2249. http://dx.doi.org/10.1162/jocn.2008.20157.

Bidelman, G.M., 2017. Communicating in challenging environments: noise and reverberation. In: Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N. (Eds.), The Frequency-Following Response. Springer International Publishing, Cham, pp. 193–224.

Bidelman, G.M., Dexter, L., 2015. Bilinguals at the "cocktail party": dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. Brain Lang. 143, 32–41. http://dx.doi.org/10.1016/j.bandl.2015.02.002.

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011a. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J. Cogn. Neurosci. 23 (2), 425–434. http://dx.doi.org/10.1162/jocn.2009.21362.

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011b. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn. 77 (1), 1–10. http://dx.doi.org/10.1016/j.bandc.2011.07.006.

Bidelman, G.M., Moreno, S., Alain, C., 2013. Tracing the emergence of categorical speech perception in the human auditory system. NeuroImage 79, 201–212. http://dx.doi.org/10.1016/j.neuroimage.2013.04.093.

Boersma, P., Weenink, D., 2014. Praat: doing phonetics by computer (Version 5. 3. 79) [Computer program]. Retrieved from ⟨http://www.praat.org/⟩.

Brigner, W.L., 1988. Perceived duration as a function of pitch. Percept. Mot. Skills 67 (1), 301–302.

Carcagno, S., Plack, C.J., 2011. Subcortical plasticity following perceptual learning in a pitch discrimination task. J. Assoc. Res. Otolaryngol. 12 (1), 89–100. http://dx.doi.org/10.1007/s10162-010-0236-1.

Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. Psychophysiology 47 (2), 236–246. http://dx.doi.org/10.1111/j.1469-8986.2009.00928.x.

Chandrasekaran, B., Kraus, N., Wong, P.C., 2012. Human inferior colliculus activity relates to individual differences in spoken language learning. J. Neurophysiol. 107 (5), 1325–1336. http://dx.doi.org/10.1152/jn.00923.2011.

Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2007. Mismatch negativity to pitch contours is influenced by language experience. Brain Res. 1128 (1), 148–156. http://dx.doi.org/10.1016/j.brainres.2006.10.064.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Naatanen, R., 1998. Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. 1 (5), 351–353.

Coffey, E.B., Herholz, S.C., Chepesiuk, A.M., Baillet, S., Zatorre, R.J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. Nat. Commun. 7, 11070. http://dx.doi.org/10.1038/ncomms11070.

Cumming, R., 2011. The effect of dynamic fundamental frequency on the perception of duration. J. Phon. 39 (3), 375–387. http://dx.doi.org/10.1016/j.wocn.2011.01.004.

Dahmen, J.C., King, A.J., 2007. Learning to hear: plasticity of auditory cortical processing. Curr. Opin. Neurobiol. 17 (4), 456–464. http://dx.doi.org/10.1016/j.conb.2007.07.004.

Ferjan Ramirez, N., Kuhl, P., 2017. Bilingual baby: foreign language intervention in Madrid's infant education centers. Mind, Brain, Educ. 11 (3), 133–143. http://dx.doi.org/10.1111/mbe.12144.

Garner, W.R., Felfoldy, G.L., 1970. Integrality of stimulus dimensions in various types of information processing. Cogn. Psychol. 1 (3), 225–241. http://dx.doi.org/10.1016/0010-0285(70)90016-2.

Ghazanfar, A.A., Maier, J.X., 2009. Rhesus monkeys (Macaca mulatta) hear rising frequency sounds as looming. Behav. Neurosci. 123 (4), 822–827. http://dx.doi.org/10.1037/a0016391.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., Kuhl, P.K., 2006. Infant speech perception activates Broca's area: a developmental magnetoencephalography study. Neuroreport 17 (10), 957–962. http://dx.doi.org/10.1097/01.wnr.0000223387.51704.89.

Intartaglia, B., White-Schwoch, T., Kraus, N., Schon, D., 2017. Music training enhances the automatic neural processing of foreign speech sounds. Sci. Rep. 7 (1), 12631. http://dx.doi.org/10.1038/s41598-017-12575-1.

Intartaglia, B., White-Schwoch, T., Meunier, C., Roman, S., Kraus, N., Schon, D., 2016. Native language shapes automatic neural processing of speech. Neuropsychologia 89, 57–65. http://dx.doi.org/10.1016/j.neuropsychologia.2016.05.033.

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C., 2003. A perceptual interference account of acquisition difficulties for nonnative phonemes. Cognition 87 (1), B47–B57. http://dx.doi.org/10.1016/s0010-0277(02)00198-1.

Jamieson, D.G., Kranjc, G., Yu, K., Hodgetts, W.E., 2004. Speech intelligibility of young school-aged children in the presence of real-life classroom noise. J. Am. Acad. Audiol. 15 (7), 508–517.

Jeng, F.C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., Lin, C.D., 2011. Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. Ear Hear. 32 (6), 699–707. http://dx.doi.org/10.1097/AUD.0b013e31821cc0df.

Jeng, F.C., Lin, C.D., Chou, M.S., Hollister, G.R., Sabol, J.T., Mayhugh, G.N., Wang, C.Y., 2016a. Development of subcortical pitch representation in three-month-old Chinese infants. Percept. Mot. Skills 122 (1), 123–135. http://dx.doi.org/10.1177/0031512516631054.

Jeng, F.C., Lin, C.D., Wang, T.C., 2016b. Subcortical neural representation to Mandarin pitch contours in American and Chinese newborns. J. Acoust. Soc. Am. 139 (6), EL190. http://dx.doi.org/10.1121/1.4953998.

Johnson, M.H., 2001. Functional brain development in humans. Nat. Rev. Neurosci. 2 (7), 475–483. http://dx.doi.org/10.1038/35081509.

Jongman, A., Wang, Y., Moore, C.B., Sereno, J.A., 2006. Perception and Production of Mandarin Chinese Tones. 〈http://www.ku.edu/∼kuppl/sereno/Handbookjong%20in%20press.pdf〉.

Keating, P., Kuo, G., 2012. Comparison of speaking fundamental frequency in English and Mandarin. J. Acoust. Soc. Am. 132 (2), 1050–1060. http://dx.doi.org/10.1121/1.4730893.

Kleiner, M., Brainard, D., Pelli, D., 2007. What's new in Psychtoolbox-3? Perception 36 (14) (14-14).

Koerner, T.K., Zhang, Y., 2017. Application of linear mixed-effects models in human neuroscience research: a comparison with Pearson Correlation in two auditory electrophysiology studies. Brain Sci. 7 (3), 26. http://dx.doi.org/10.3390/brainsci7030026.

Kraus, N., Anderson, S., White-Schwoch, T., 2017. The frequency-following response: a window into human communication. In: Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N. (Eds.), The Frequency-Following Response: A Window into Human Communication. Springer International Publishing, Cham, pp. 1–15.

Krishnan, A., Gandour, J.T., 2017. Shaping brainstem representation of pitch-relevant information by language experience. In: Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N. (Eds.), The Frequency-Following Response: A Window into Human Communication. Springer International Publishing, Cham, pp. 45–73.

Krishnan, A., Gandour, J.T., Ananthakrishnan, S., Vijayaraghavan, V., 2015. Language experience enhances early cortical pitch-dependent responses. J. Neurolinguist. 33, 128–148. http://dx.doi.org/10.1016/j.jneuroling.2014.08.002.

Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010a. Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. Brain Res. 1313, 124–133. http://dx.doi.org/10.1016/j.brainres.2009.11.061.

Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010b. The effects of tone language experience on pitch processing in the brainstem. J. Neurolinguist. 23 (1), 81–95. http://dx.doi.org/10.1016/j.jneuroling.2009.09.001.

Krishnan, A., Gandour, J.T., Smalt, C.J., Bidelman, G.M., 2010c. Language-dependent pitch encoding advantage in the brainstem is not limited to acceleration rates that occur in natural speech. Brain Lang. 114 (3), 193–198. http://dx.doi.org/10.1016/j.bandl.2010.05.004.

Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009a. Experience-dependent neural representation of dynamic pitch in the brainstem. Neuroreport 20 (4), 408–413. http://dx.doi.org/10.1097/WNR.0b013e3283263000.

Krishnan, A., Swaminathan, J., Gandour, J., 2009b. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. J. Cogn. Neurosci. 12 (6), 1092–1105.

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. http://dx.doi.org/10.1016/j.bandl.2014.09.005.

Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res. Cogn. Brain Res. 25 (1), 161–168. http://dx.doi.org/10.1016/j.cogbrainres.2005.05.004.

Krishnan, A., Xu, Y., Gandour, J.T., Cariani, P.A., 2004. Human frequency-following response: representation of pitch contours in Chinese tones. Hear. Res. 189 (1–2), 1–12. http://dx.doi.org/10.1016/S0378-5955(03)00402-7.

Kuhl, P.K., 2007. Is speech learning 'gated' by the social brain? Dev. Sci. 10 (1), 110–120.

Kuhl, P.K., 2010. Brain mechanisms in early language acquisition. Neuron 67 (5), 713–727. http://dx.doi.org/10.1016/j.neuron.2010.08.038.

Kuhl, P.K., Conboy, B.T., Padden, D., Nelson, T., Pruitt, J., 2005. Early speech perception and later language development: implications for the "Critical Period". Lang. Learn. Dev. 1 (3–4), 237–264. http://dx.doi.org/10.1080/15475441.2005.9671948.

Lake, J.I., LaBar, K.S., Meck, W.H., 2014. Hear it playing low and slow: how pitch level differentially influences time perception. Acta Psychol. (Amst.) 149, 169–177. http://dx.doi.org/10.1016/j.actpsy.2014.03.010.

Lee, L., Nusbaum, H.C., 1993. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. Percept. Psychophys. 53 (2), 157–165. http://dx.doi.org/10.3758/bf03211726.

Lee, Y.S., Vakoch, D.A., Wurm, L.H., 1996. Tone perception in Cantonese and Mandarin: a cross-linguistic comparison. J. Psycholinguist. Res. 25 (5), 527–542.

Lehiste, I., 1976. Influence of fundamental frequency pattern on the perception of duration. J. Phon. 4 (2), 113–117.

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 (51), 19558–19563. http://dx.doi.org/10.1073/pnas.0607065104.

Mamiya, P.C., Richards, T.L., Coe, B.P., Eichler, E.E., Kuhl, P.K., 2016. Brain white matter structure and COMT gene are linked to second-language learning in adults. Proc. Natl. Acad. Sci. USA 113 (26), 7249–7254. http://dx.doi.org/10.1073/pnas.1606602113.

Musacchia, G., Sams, M., Skoe, E., Kraus, N., 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc. Natl. Acad. Sci. USA 104 (40), 15894–15898. http://dx.doi.org/10.1073/pnas.0701498104.

Musacchia, G., Strait, D., Kraus, N., 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear. Res. 241 (1–2), 34–42. http://dx.doi.org/10.1016/j.heares.2008.04.013.

Nelken, I., 2004. Processing of complex stimuli and natural scenes in the auditory cortex. Curr. Opin. Neurobiol. 14 (4), 474–480. http://dx.doi.org/10.1016/j.conb.2004.06.005.

Pisoni, D.B., 1976. Fundamental frequency and perceived vowel duration (S39-S39). J. Acoust. Soc. Am. 59 (S1). http://dx.doi.org/10.1121/1.2002669.

Rao, A., Zhang, Y., Miller, S., 2010. Selective listening of concurrent auditory stimuli: an Event-Related Potential study. Hear. Res. 268, 123–132. http://dx.doi.org/10.1016/j.heares.2010.05.013.

Reetz, H., Jongman, A., 2011. Phonetics: Transcription, Production, Acoustics, and Perception 34 John Wiley & Sons.

Repp, B.H., Lin, H.B., 1990. Integration of segmental and tonal information in speech-perception - a cross-linguistic study. J. Phon. 18 (4), 481–495.

Russo, N.M., Nicol, T.G., Zecker, S.G., Hayes, E.A., Kraus, N., 2005. Auditory training improves neural timing in the human brainstem. Behav. Brain Res. 156 (1), 95–103. http://dx.doi.org/10.1016/j.bbr.2004.05.012.

Schlauch, R.S., Ries, D.T., DiGiovanni, J.J., 2001. Duration discrimination and subjective duration for ramped and damped sounds. J. Acoust. Soc. Am. 109 (6), 2880–2887. http://dx.doi.org/10.1121/1.1372913.

Shen, X.S., Lin, M., Yan, J., 1993. F0 turning point as an F0 cue to tonal contrast: a case study of Mandarin tones 2 and 3. J. Acoust. Soc. Am. 93 (4), 2241–2243. http://dx.doi.org/10.1121/1.406688.

Skoe, E., Kraus, N., 2010. Auditory brain stem response to complex sounds: a tutorial. Ear Hear. 31 (3), 302–324. http://dx.doi.org/10.1097/AUD.0b013e3181cdb272.

Skoe, E., Krizman, J., Spitzer, E., Kraus, N., 2015. Prior experience biases subcortical sensitivity to sound patterns. J. Cogn. Neurosci. 27 (1), 124–140. http://dx.doi.org/10.1162/jocn_a_00691.

Song, J.H., Skoe, E., Wong, P.C., Kraus, N., 2008. Plasticity in the adult human auditory brainstem following short-term linguistic training. J. Cogn. Neurosci. 20 (10), 1892–1902. http://dx.doi.org/10.1162/jocn.2008.20131.

Stevens, J., Zhang, Y., 2014. Brain mechanisms for processing co-speech gesture: a cross-language study of spatial demonstratives. J. Neurolinguist. 30, 27–47. http://dx.doi.org/10.1016/j.jneuroling.2014.03.003.

Swaminathan, J., Krishnan, A., Gandour, J.T., 2008a. Pitch encoding in speech and nonspeech contexts in the human auditory brainstem. NeuroReport 19 (11), 1163–1167. http://dx.doi.org/10.1097/WNR.0b013e3283088d31.

Swaminathan, J., Krishnan, A., Gandour, J.T., Xu, Y., 2008b. Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem. IEEE Trans. Biomed. Eng. 55 (1), 281–287. http://dx.doi.org/10.1109/TBME.2007.896592.

Tong, Y.X., Francis, A.L., Gandour, J.T., 2008. Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. Lang. Cogn. Process. 23 (5), 689–708. http://dx.doi.org/10.1080/01690960701728261.

Wang, W.S.Y., Lehiste, I., Chuang, C.K., Darnovsky, N., 1976. Perception of vowel duration (S92-S92). J. Acoust. Soc. Am. 60 (S1). http://dx.doi.org/10.1121/1.2003607.

Weiss, M.W., Bidelman, G.M., 2015. Listening to the brainstem: musicianship enhances intelligibility of subcortical representations for speech. J. Neurosci. 35 (4), 1687–1691. http://dx.doi.org/10.1523/JNEUROSCI.3680-14.2015.

White-Schwoch, T., Kraus, N., 2017. The Janus face of auditory learning: how life in sound shapes everyday communication. In: Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N. (Eds.), The Frequency-Following Response: A Window into Human Communication. Springer International Publishing, Cham, pp. 121–158.

Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10 (4), 420–422. http://dx.doi.org/10.1038/nn1872.

Xi, J., Zhang, L., Shu, H., Zhang, Y., Li, P., 2010. Categorical perception of lexical tones in Chinese revealed by mismatch negativity. Neuroscience 170 (1), 223–231. http://dx.doi.org/10.1016/j.neuroscience.2010.06.077.

Xu, Y., 1997. Contextual tonal variations in Mandarin. J. Phon. 25 (1), 61–83. http://dx.doi.org/10.1006/jpho.1996.0034.

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.

Ye, Y., Connine, C.M., 1999. Processing spoken Chinese: the role of tone information. Lang. Cogn. Process. 14 (5–6), 609–630. http://dx.doi.org/10.1080/016909699386202.

Yu, A.C., 2010. Tonal effects on perceived vowel duration. Lab. Phonol. 10, 151–168.

Zatorre, R.J., Gandour, J.T., 2008. Neural specializations for speech and pitch: moving beyond the dichotomies. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363 (1493), 1087–1104. http://dx.doi.org/10.1098/rstb.2007.2161.

Zhang, L., Li, Y., Wu, H., Li, X., Shu, H., Zhang, Y., Li, P., 2016a. Effects of semantic context and fundamental frequency contours on Mandarin speech recognition by second language learners. Front. Psychol. 7, 908. http://dx.doi.org/10.3389/fpsyg.2016.00908.

Zhang, Y., Cheng, B., Koerner, T.K., Schlauch, R.S., Tanaka, K., Kawakatsu, M., Imada, T., 2016b. Perceptual temporal asymmetry associated with distinct ON and OFF responses to time-varying sounds with rising versus falling intensity: a magnetoencephalography study. Brain Sci. 6 (3), 27. http://dx.doi.org/10.3390/brainsci6030027.

Zhang, Y., Koerner, T., Miller, S., Grice-Patil, Z., Svec, A., Akbari, D., Carney, E., 2011. Neural coding of formant-exaggerated speech in the infant brain. Dev. Sci. 14 (3), 566–581. http://dx.doi.org/10.1111/j.1467-7687.2010.01004.x.

Zhang, Y., Kuhl, P.K., Imada, T., Iverson, P., Pruitt, J., Stevens, E.B., Nemoto, I., 2009. Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. NeuroImage 46 (1), 226–240. http://dx.doi.org/10.1016/j.neuroimage.2009.01.028.

Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M., Tohkura, Y., 2005. Effects of language experience: neural commitment to language-specific auditory patterns. NeuroImage 26 (3), 703–720. http://dx.doi.org/10.1016/j.neuroimage.2005.02.040.

Zhang, Y., Wang, Y., 2007. Neural plasticity in speech acquisition and learning. Biling.-Lang. Cogn. 10 (2), 147–160. http://dx.doi.org/10.1017/S1366728907002908.