ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Categorical Perception of Chinese Lexical Tones by Late Second Language Learners With High Proficiency: Behavioral and Electrophysiological Measures

Jie Xi,[a,b] Hongkai Xu,[c] Ying Zhu,[c] Linjun Zhang,[c] Hua Shu,[d] and Yang Zhang[e]

[a]CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China [b]Department of Psychology, University of Chinese Academy of Sciences, Beijing, China [c]Beijing Advanced Innovation Center for Language Resources and College of Advanced Chinese Training, Beijing Language and Culture University, China [d]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, China [e]Department of Speech-Language-Hearing Sciences, University of Minnesota, Twin Cities, Minneapolis

ARTICLE INFO

Article History:
Received April 27, 2020
Revision received June 30, 2021
Accepted August 11, 2021

Editor-in-Chief: Bharath Chandrasekaran
Editor: Chao-Yang Lee

https://doi.org/10.1044/2021_JSLHR-20-00210

ABSTRACT

Purpose: Although acquisition of Chinese lexical tones by second language (L2) learners has been intensively investigated, very few studies focused on categorical perception (CP) of lexical tones by highly proficient L2 learners. This study was designed to address this issue with behavioral and electrophysiological measures.

Method: Behavioral identification and auditory event-related potential (ERP) components for speech discrimination, including mismatch negativity (MMN), N2b, and P3b, were measured in 23 native Korean speakers who were highly proficient late L2 learners of Chinese. For the ERP measures, both passive and active listening tasks were administered to examine the automatic and attention-controlled discriminative responses to within- and across-category differences for carefully chosen stimuli from a lexical tone continuum.

Results: The behavioral task revealed native-like identification function of the tonal continuum. Correspondingly, the active oddball task demonstrated larger P3b amplitudes for the across-category than within-category deviants in the left recording site, indicating clear CP of lexical tones in the attentive condition. By contrast, similar MMN responses in the right recording site were elicited by both the across- and within-category deviants, indicating the absence of CP effect with automatic phonological processing of lexical tones at the pre-attentive stage even in L2 learners with high Chinese proficiency.

Conclusion: Although behavioral data showed clear evidence of categorical perception of lexical tones in proficient L2 learners, ERP measures from passive and active listening tasks demonstrated fine-grained sensitivity in terms of response polarity, latency, and laterality in revealing different aspects of auditory versus linguistic processing associated with speech decoding by means of largely implicit native language acquisition versus effortful explicit L2 learning.

Categorical perception (CP) of speech sounds refers to the ability of human listeners to perceive across-category differences as more salient than within-category differences, even though physical differences are the same along some relevant parameter in both of the cases (Fry et al., 1962; Liberman et al., 1957). CP is a phenomenon that has been extensively investigated for many years with most of the early studies focused on segmental features such as voice onset time and second-formant transition of consonants (Liberman et al., 1957, 1961; Lisker, 1975; also see Casserly & Pisoni, 2010, for a review). Recent studies on suprasegmental features have consistently indicated that fundamental frequency (F_0) contours of Chinese lexical

Correspondence to Linjun Zhang: zhanglinjun75@gmail.com and Yang Zhang: zhang470@umn.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

tones are also perceived categorically (Francis et al., 2003; Hallé et al., 2004; Peng et al., 2010; Xu et al., 2006).

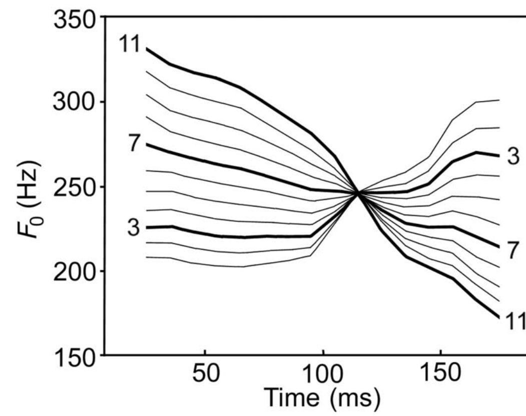
In addition to behavioral research, electrophysiological studies on CP of segmental features have revealed that three event-related potential (ERP) components, that is, the mismatch negativity (MMN), N2b, and P3b are clear indicators of CP at different processing stages. The MMN component, which is typically identified at fronto-central electrodes and peaks at approximately 100–250 ms from stimulus onset, reflects early pre-attentive detection of acoustic and phonological changes under passive listening conditions in which attention is not required (Kasai et al., 2001; Kazanina et al., 2006; Maiste et al., 1995; Nenonen et al., 2005). The N2b and P3b components, which are identified at centro-parietal and temporal–parietal electrodes and peak at about 200–300 ms and 350–450 ms from stimulus onset, respectively, reflect later attentive detection of the acoustic and phonological deviations under active listening paradigms in which attention is directed toward the target stimuli (Buchwald et al., 1994; Frenck-Mestre et al., 2005). Consistent with the CP research on segmental features, our own ERP studies (Xi, Zhang, et al., 2010; L. Zhang, Xi, et al., 2012; Y. Zhang, Zhang, et al., 2012), among others (Shen & Froud, 2019; Zheng et al., 2012), have consistently demonstrated that these three ERP components are reliable neural indices of CP of Chinese lexical tones in native speakers. In particular, across-category tonal deviants elicited stronger electrophysiological responses than within-category tonal deviants especially in the left recording sites in both passive (which elicits the MMN component) and active (which produces the N2b and P3b components) listening conditions, reflecting pre-attentive and attentive phonological processing of lexical tones, respectively.

CP of lexical tones is acquired early in life through the linguistic experience of native languages. While newborn infants have biologically innate sensitivities to variations in F_0 contours irrespective of native languages (Mampe et al., 2009; Wermke et al., 2017), perceptual reorganization for tones occurs within the first year of life and CP of lexical tones gradually develops as a result of perceptual refinement of tones (Harrison, 2000; Mattock et al., 2008; Tsao, 2017; Yeung et al., 2013). Behavioral studies showed that children aged 6 years old have already acquired adultlike competence of CP of Chinese lexical tones (Chen et al., 2017; Xi, Jiang, et al., 2010). Our previous ERP study further revealed that 10-year-old children exhibited significantly enhanced MMN responses to across-category tonal deviants compared with within-category deviants in the left recording site, indicating that children of this age have formed adultlike phonemic representations of Chinese lexical tones at the neurophysiological level (Y. Zhang, Zhang, et al., 2012). Partly because CP of lexical tones develops and matures early in life, it is particularly difficult for adult nontonal language speakers. For example, previous behavioral research (Hallé et al., 2004; Xu et al., 2006) and our ERP study (see the supplementary data of Xi, Zhang, et al., 2010) have demonstrated that perception of Chinese lexical tones by native French, English, and Korean speakers is psychophysically based.

However, it is still possible for adult second language (L2) learners to establish CP of lexical tone if appropriate training and exposure are given. For example, Zhao and Kuhl (2015) demonstrated that short-term perceptual training on Chinese lexical tones could result in a somewhat categorical-like perception pattern across a tonal continuum in native speakers of English. In two studies on L2 learners of Chinese in a naturalistic classroom learning environment, Shen and Froud (2016) and Yu et al. (2019) found that nontonal language speakers with intermediate Chinese proficiency (native English speakers in the former and native Indonesian, Russian, and Kirghiz speakers in the latter) showed CP patterns of lexical tones indistinguishable from those of native Chinese speakers. By contrast, their following ERP studies demonstrated that the L2 learners did not show CP effects at either pre-attentive or attentive processing stage. Specifically, across-category deviants did not elicit stronger MMN response under the passive oddball paradigm or N2b or P3b responses under the active oddball paradigm compared with within-category deviants. The discrepancies between behavioral and ERP results indicate that CP of Chinese lexical tone as determined by behavioral tasks alone does not predict the same neural processing as employed by native Chinese speakers.

Taken together, the existing literature suggests that despite native-like behavioral performance, L2 learners may demonstrate different neurophysiological responses and neural substrates in categorically perceiving nonnative speech sounds. The L2 learners in the previous studies, however, appear to have limited learning experience and Chinese proficiency. For example, in the study of Yu et al. (2019), all participants learned Chinese in the target language environment and passed the intermediate level of Chinese proficiency test, and in the study of Shen and Froud (2016), the learners took Chinese courses in a native language (L1) environment for about only three semesters. It remains unclear whether nontonal language speakers who have attained high Chinese proficiency as an L2 would show neural processing similar to those in native Chinese speakers during CP of Chinese lexical tones. This study aimed to fill this gap by adopting both passive and active oddball paradigms to examine MMN, N2b, and P3b responses elicited by across-category and within-category tonal deviants in proficient L2 learners of Chinese. On the basis of the well-known theory of L2 acquisition (Krashen, 1981), which draws a distinction between language acquisition and language learning, we predicted that proficient L2 learners would demonstrate native-like electrophysiological patterns for CP of lexical tones in the attentive condition (i.e., across-category deviants elicited stronger N2b and P3b than within-category deviants in the left recording sites), reflecting learners' conscious knowledge about lexical tones as a result of explicit language learning. Furthermore, we predicted that proficient L2 learners might not necessarily demonstrate native-like MMN response patterns in the non-attentive condition, which could index the learners' insufficient ability to perceive lexical tones subconsciously as their L2 learning experience does not mirror that of the first language acquisition. By contrast, across- and within-category deviants would elicit similar MMN responses in the right recording site, reflecting automatic detection of acoustic rather than phonological differences by L2 learners irrespective of Chinese proficiency.

Figure 1. Fundamental frequency contours of the tonal continuum from Tone 2 to Tone 4. Continua 3, 7, and 11 used in the event-related potential session are marked with thick lines.



Materials and Method

Participants

Twenty-three neurologically healthy volunteers (10 women, 13 men; aged 20–26 years) with normal hearing and minimal musical experience (less than 1 year of total musical training and no musical training within the past 5 years) participated in the study. All participants were native speakers of Korean and started learning Chinese after the age of 16 years. They were all senior students majoring in Chinese language and literature with experience of Chinese learning in Beijing for at least 3.5 years and achieved the highest level (HSK-6) of *Hanyu Shuiping Kaoshi* (a standardized Chinese proficiency test for nonnative learners). All of them were all right-handed according to a handedness questionnaire adapted from a modified Chinese version of the Edinburgh Handedness Inventory (Oldfield, 1971). Participants gave written consent before they took part in the experiment. The experiment was approved by the institutional review board of Beijing Language and Culture University. Participants were paid for their participation in both behavioral and ERP sessions. The ERP data from two male and three female participants were excluded from further analyses due to insufficient number of acceptable trials as a result of strong electroencephalogram (EEG) artifacts including excessive blinking.

Stimuli

The same stimuli as that used in our previous studies (Xi, Zhang, et al., 2010; L. Zhang, Xi, et al., 2012; Y. Zhang, Zhang, et al., 2012) were used in this study and were described only briefly here (see Figure 1). Specifically, a Chinese lexical tonal continuum from the high-rising tone (Tone 2) to the high-falling tone (Tone 4) was adopted in the behavioral session. Three tonal stimuli were chosen from the continuum for the ERP session, forming an across-category tonal pair (3 vs. 7) and a within-category tonal pair (7 vs. 11). Because the physical intervals between the across- and within-category pairs were equated, the key difference between the tonal pairs is that the former involves a change to a different phonological category, whereas the latter involves only an acoustic change. As in our previous studies (Xi, Zhang, et al., 2010; L. Zhang, Xi, et al., 2012; Y. Zhang, Zhang, et al., 2012), Stimulus 7 was used as the standard and Stimuli 3 and 11 as the across- and within-category deviants, respectively.

Experimental Procedures

Behavioral Session

An identification task was adopted. Participants were presented with 20 repetitions of each stimulus, and this equated to a total of 220 trials. Stimuli were presented in a random order, and participants were asked to identify the tones via forced choice between Tone 2 and Tone 4.

ERP Session

The within- and across-category deviants were presented pseudorandomly among standards with a probability of 10%, respectively, and any two adjacent deviants were separated by at least three standards for a total presentation of 1,000 stimuli (800 standards, 100 within-category deviants, and 100 across-category deviants) in a passive or active oddball block. The stimulus-onset asynchrony was 1,000 ms, including the 200-ms stimulus duration and 800-ms interstimulus interval. Participants were seated comfortably in an acoustically and electrically shielded chamber. In the passive condition, participants were instructed to ignore the presented sounds while watching a self-selected movie, which was presented in mute mode with subtitles. In the attended condition, they were instructed to pay attention to the stimuli presented and to press a button when they detected a different sound. Passive listening was conducted before active listening for each participant. Sound stimuli were presented binaurally through an insert Sony earphone (Sony Corporation). The right and left acoustic channels of the insert earphone were calibrated for equal and comfortable loudness (70 dB SPL) before the experiment. The ERP session lasted approximately 4 hr including preparation, data acquisition, and cleanup.

Electrophysiological Recording and Analysis

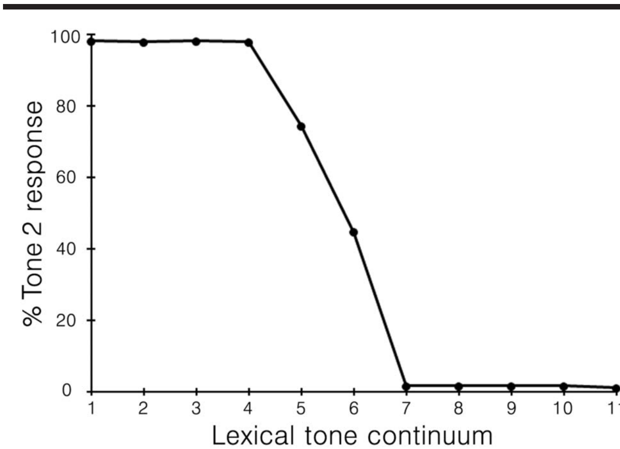
Continuous EEG was recorded using a HydroCel Geodesic Sensor Net (Electrical Geodesics Incorporated Company), consisting of 128 electrodes evenly distributed across the scalp and referenced against the vertex electrode. The Geodesic Sensor Net also includes electrodes next to, and below, the eyes for recording horizontal and vertical eye movements. The impedance of each electrode was maintained below 5 kΩ. Off-line signal processing was carried out using the Net Station (Electrical Geodesics Inc.), EEGLAB (Delorme & Makeig, 2004), and ERPLAB (Lopez-Calderon & Luck, 2014) software tools. The raw data were first digitally filtered using a 0.3- to 20-Hz bandpass filter and segmented for 1,000 ms starting 100 ms before the onset of stimuli. Data were then rereferenced to the average of all the electrodes and baseline corrected. Recorded trials with eye blinks or other activities beyond the range of −50 to +50 μV were rejected. To select electrodes for the MMN, N2b, and P3b amplitudes and latencies analysis, grand-averaged ERPs were made by averaging across subjects and stimulus conditions but separately for the passive and active blocks. The three components were respectively measured at frontal, central, and parietal electrode sites, where amplitudes of the corresponding components were maximal. Specifically, two recording sites were selected for statistical analyses of MMN: left frontal (recording site F3, Channels 19, 23, and 24) and right frontal (recording site F4, Channels 3, 4, and 124); two recording sites were selected for statistical analyses of N2b: left central (recording site C1, Channels 30, 36, and 37) and right central (recording site C2, Channels 87, 104, and 105); two recording sites were selected for statistical analyses of P3b: left parietal (recording site P1, Channels 66, 67, and 71) and right parietal (recording site P2, Channels 76, 77, and 84). Only the standards before the deviant were obtained for averaging and subtraction. Difference waves for MMN, N2b, and P3b were obtained by subtracting the averaged standard from the averaged deviant. The MMN peak latency for each subject was found within a 40-ms time window that was defined by the grand-average waveforms at Fz. The N2b peak latency for each participant was found within a 60-ms time window that was defined by the grand-average waveforms at Cz. The P3b peak latency for each participant was found within a 100-ms time window that was defined by the grand-average waveforms at Pz. The mean amplitude was calculated using a moving window technique: First, the negative or positive peak within a specific time window was found for each participant and then the values of a time window that extended ±40 ms surrounding the MMN peak, ±60 ms surrounding the N2b peak, and ±100 ms surrounding the P3b peak were averaged. Statistical analysis only included those participants with at least 80 accepted deviant trials in each block. The same time windows as those adopted in our previous studies on CP of Chinese lexical tone by native Chinese speakers (Xi, Zhang, et al., 2010; L. Zhang, Xi, et al., 2012) allowed us to compare the different response patterns between L1 listeners and proficient L2 listeners.

Results

Behavioral Session

The identification functions of the tonal continuum from Tone 2 to Tone 4 are shown in Figure 2. These functions were obtained by calculating the average percentage of Tone 2 and Tone 4 responses for each of the 11 tokens. The identification curve is similar with that obtained from native Chinese speakers (Xi, Zhang, et al., 2010; Yu et al., 2019). The identification data were then submitted to probit analysis (Finney, 1971), from which the values of slope were obtained. Probit analysis fits a cumulative

Figure 2. Identification function of the tonal continuum from Tone 2 to Tone 4.



normal curve to probability estimates as a function of stimulus level, and the slope serves as an index of identification consistency. Further t test on the slope values revealed no significant difference, $t(17) = -2.056$, $p = .055$, between the L2 listeners and native Chinese listeners (the slope value reported in Yu et al., 2019, was used here), indicating native-like CP of the tonal continuum by Korean proficient learners of Chinese. Visual inspection of the identification curve of each participant further confirmed that the selection of Stimuli 3, 7, and 11 truly conformed to the assigned categories.

Behavioral Data of the Active Oddball Task

The false alarm rate was very low (below 3%). The difference between the hit rate for the across-category deviants (M = 95%, SD = 5%) and that for the within-category deviants (M = 84%, SD = 7%) was statistically significant, $t(17) = 4.626$, $p < .001$, reflecting the effect of CP. However, there was no significant difference in the reaction time, $t(17) = -1.71$, $p = .105$, between the across-category (M = 699 ms, SD = 51 ms) and within-category deviants (M = 676 ms, SD = 62 ms).

ERP Session

The grand-average waveforms for the standards, within-category deviants, and across-category deviants in the passive and active oddball conditions are shown in Figure 3. Negative and positive peaks were observed in the deviant-minus-standard difference waves for both conditions (see Figure 4). Six separate two-way (deviant type: across-category vs. within-category × recording site: left vs. right) repeated-measures analyses of variance were conducted for mean amplitudes and peak latencies of MMN, N2b, and P3b components. For all analyses, degrees of freedom were adjusted according to the method of Greenhouse–Geisser when appropriate.

In the passive oddball condition, neither the main effect of deviant type, $F(1, 17) = 0.924$, $p = .350$, $\eta_p^2 = .052$, nor the main effect of recording site, $F(1, 17) = 0.822$, $p = .377$, $\eta_p^2 = .046$, reached significance for MMN peak latencies. The interaction effect was not significant either, $F(1, 17) = 0.532$, $p = .476$, $\eta_p^2 = .030$. For MMN mean amplitudes, the significant main effect of recording site showed that the amplitude elicited in the right recording site was larger than that elicited in the left recording site, $F(1, 17) = 9.997$, $p = .006$, $\eta_p^2 = .370$, but neither the main effect of deviant type, $F(1, 17) = 1.150$, $p = .299$, $\eta_p^2 = .063$, nor the interaction between deviant type and recording site, $F(1, 17) = 0.295$, $p = .594$, $\eta_p^2 = .017$, reached significance.

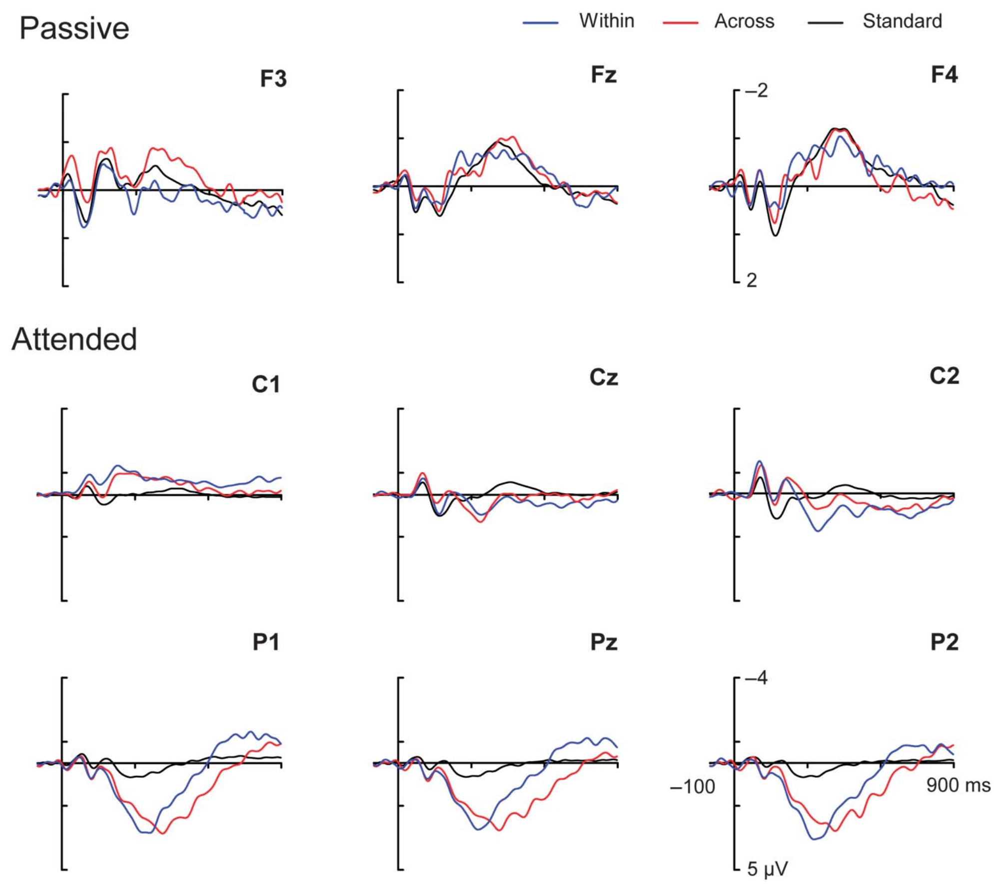
In the active oddball condition, the main effect of recording site reached significance for N2b peak latencies, $F(1, 17) = 15.205$, $p = .001$, $\eta_p^2 = .472$, with the response peaking earlier in the right than in the left recording site, but neither the main effect of deviant type, $F(1, 17) = 3.079$, $p = .097$, $\eta_p^2 = .153$, nor the interaction between recording site and deviant type, $F(1, 17) = 2.231$, $p = .154$, $\eta_p^2 = .116$, was significant. For the N2b mean amplitudes, the main effects of both deviant type, $F(1, 17) = 3.007$, $p = .101$, $\eta_p^2 = .150$, and recording site, $F(1, 17) = 0.391$, $p = .540$, $\eta_p^2 = .022$, and the interaction, $F(1, 17) = 3.459$, $p = .080$, $\eta_p^2 = .169$, failed to reach significance. For P3b peak latencies, only the main effect of deviant type was significant, $F(1, 17) = 13.119$, $p = .002$, $\eta_p^2 = .436$, indicating that across-category deviants are processed more slowly than within-category deviants, but neither the main effect of recording site, $F(1, 17) = 2.860$, $p = .109$, $\eta_p^2 = .144$, nor the interaction between deviant type and recording site, $F(1, 17) = 0.596$, $p = .451$, $\eta_p^2 = .034$, reached significance. For P3b mean amplitudes, neither the main effect of deviant type, $F(1, 17) = 3.182$, $p = .092$, $\eta_p^2 = .158$, nor recording site, $F(1, 17) = 0.509$, $p = .485$, $\eta_p^2 = .029$, reached significance. However, there was significant interaction between deviant type and recording site, $F(1, 17) = 5.528$, $p = .031$, $\eta_p^2 = .245$. Further simple effect analyses revealed that across-category deviants elicited greater amplitude than within-category deviants only in the left recording site, $F(1, 17) = 7.97$, $p = .012$, $\eta_p^2 = .319$, but not in the right recording site $F(1, 17) = 0.64$, $p = .434$, $\eta_p^2 = .036$.

Discussion

In this study, behavioral identification and auditory ERPs (MMN, N2b, and P3b) were measured in order to investigate CP of Chinese lexical tones by proficient Korean L2 learners of Chinese. The identification task revealed native-like CP of the tonal continuum from Tone 2 to Tone 4. Correspondingly, the active oddball task demonstrated larger P3b amplitude for the across- than within-category deviants, indicating clear CP of lexical tones in the attentive condition. The CP effect, however, was not observed for MMN, indicating the absence of automatic CP of lexical tones even in L2 learners with high Chinese proficiency at the pre-attentive stage.

CP of Chinese lexical tones by L2 learners has been investigated by several behavioral studies. For example, Shen and Froud (2016) examined native English speakers who were studying Chinese in America and Yu et al. (2019) examined native tonal (Thai and Vietnamese) and nontonal language (Indonesian, Kyrgyz, and Russian) speakers who were studying Chinese in China. The L2 learners with intermediate Chinese proficiency in their studies showed native-like identification curve. Our result of the identification task, therefore, is consistent with

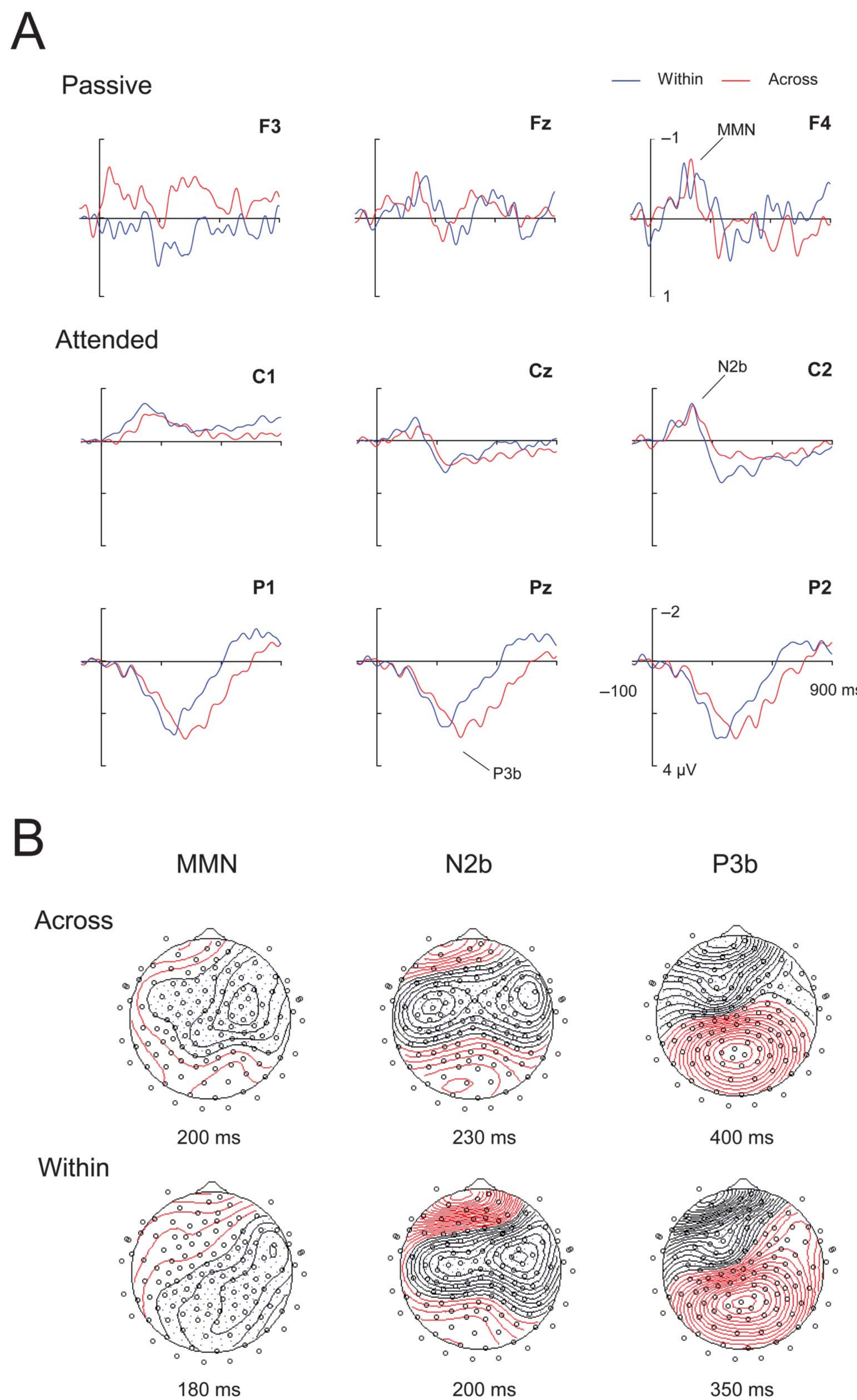
Figure 3. Grand-average waveforms elicited by the across- and within-category deviants and the standard at frontal, central, and posterior scalp recording sites.



theirs, indicating that L2 learners are able to acquire the CP ability of Chinese lexical tones irrespective of whether their native languages are a tonal or nontonal language and whether Chinese is learned in a native or target language environment. In the active oddball task, the behavioral results showed that discrimination rates of both the within- and across-category deviants were far above chance level, indicating that the L2 learners are sensitive to both the acoustic and phonological variations when detecting changes in the series of tonal stimuli. Furthermore, across-category deviants were detected better than within-category deviants, reflecting the effect of CP.

More importantly, the ERP component, that is, P3b also demonstrated the CP effect of lexical tones. To the best of our knowledge, this is the first study providing electrophysiological evidence for CP of Chinese lexical tones by L2 learners. Previous studies that used P3b to examine acoustic and phonological processing in speech perception at attentive stages has showed that this component is more associated with phonological than with acoustic processing (Hisagi et al., 2010; Sussman et al., 2004). Studies on CP of consonants, vowels, and lexical tones have also showed that P3b is a clear indicator of CP (Dehaene-Lambertz, 1997; Frenck-Mestre et al., 2005; Peng et al., 2010; Sharma & Dorman, 1999; Xi, Zhang, et al., 2010; L. Zhang, Xi, et al., 2012). In this study, active detection of the oddball across-category deviants elicited larger P3b amplitudes than within-category deviants in the left recording site, indicating that phonological information of lexical tones is successfully processed by L2 learners. By contrast, no CP effect of lexical tones was observed for P3b in L2 learners with intermediate Chinese proficiency (Shen & Froud, 2019). It seems that CP of lexical tones, which is reflected by the identification function at the behavioral level and by the P3b component at the electrophysiological level, does not develop simultaneously. Specifically, native-like CP of lexical tones develops earlier at the behavioral level but later at the electrophysiological level. In a previous functional near-infrared spectroscopy study that examined CP of Japanese long/short vowel contrast by native Japanese speakers and L2 learners, the L2 listeners demonstrated behavioral performance indistinguishable from that of the L1 listeners, but the phonological-specific responses in the left auditory area were observed

Figure 4. MMN, N2b, and P3b responses for the across- and within-category deviants (A). Also shown here are the maps displaying the topographic distribution of the mean amplitudes in the MMN, N2b, and P3b analysis windows (B). MMN = mismatch negativity.



only in the L1 listeners (Minagawa-Kawai et al., 2005). The inconsistencies between behavioral performance and neural manifestation in L2 speech perception may be partly due to different amounts of noise in sensory decoding, attentional processing, and decision-making in the behavioral and neurophysiological measures associated with the test protocols. Functional neuroimaging studies revealed that native and L2 listeners may recruit different mechanisms for speech perception (Callan et al., 2003: Golestani & Zatorre, 2004; Zhang et al., 2009). For example, L2 listeners utilized articulatory–auditory- and articulatory–orosensory-based internal models but native listeners used auditory–phonetic representation (Callan et al., 2004), and L2 listeners perceptually weighted the acoustic dimensions for phonemic representations differently from native listeners (Zhang et al., 2009). More studies are needed to clarify the developmental trajectory and neural mechanisms underlying the native-like CP of L2 phonological contrasts such as Chinese lexical tones and Japanese long/short vowels.

It is interesting that within-category deviants elicited earlier peak latency of P3b than across-category deviants, which is consistent with the finding of our previous study on native Chinese listeners (L. Zhang, Xi, et al., 2012). Previous work has shown that P3b is sensitive to both acoustic and phonological changes in active perception of linguistic and nonlinguistic stimuli (Comerchero & Polich, 1999; Frenck-Mestre et al., 2005; Katayama & Polich, 1996; Sussman et al., 2004). As the pitch contours for lexical tone perception unfold on the temporal scale at the syllable level, a within-category change could be detected more quickly based on purely acoustic cues than an across-category contrast as the latter may involve not only acoustic processing of the actual pitch contour deviation but also additional phonological processing for both native and proficient L2 listeners. Specifically, phonological variations could be detected only when across-category deviants were recognized as a different category whereas acoustic variations alone would not implicate this process. It is notable that the reaction times did not differ between within- and across-category deviants, indicating that the subtle P3b latency differences may be attenuated by the decision-making process involving evaluation, decision, and response preparation.

In the passive oddball task, however, the CP effect of lexical tones was not observed for the MMN component. Specifically, both the across- and within-category deviants elicited stronger MMN in the right than the left recording site, indicating that acoustic rather than phonological difference between the standard and across- or within-category deviants was successfully perceived by the L2 learners at the pre-attentive stage. Although MMN and P3b components are both sensitive to phonological changes in speech perception, they reflect different neural processes with MMN indexing early, automatic, and pre-attentive processing and P3b indexing later, active, and attentive processing (Dehaene-Lambertz, 1997; Frenck-Mestre et al., 2005; Sharma & Dorman, 1999; Xi, Zhang, et al., 2010; L. Zhang, Xi, et al., 2012; Y. Zhang, Zhang, et al., 2012; Zheng et al., 2012). According to the well-accepted theory of L2 acquisition, there are two independent systems of L2 performance, that is, the learned system and acquired system (Krashen, 1981). While the learned system is the product of formal instruction, which comprises conscious knowledge of the language, the acquired system is the product of a subconscious process similar to that children undergo when their first language skills are developed. For adult L2 learners, the learned system plays a primary role. Therefore, perception of phonological contrasts in L1 is automatic and efficient, whereas perception of most contrasts in nonnative languages by adult L2 learners turns out to be cognitively demanding and inaccurate (Gervain & Mehler, 2010; Mattock & Burnham, 2006; Werker & Tees, 1984). The distinction between the two systems might underlie the presence/absence of CP effect of lexical tones for the P3b/MMN components.

It is worth noting that no CP effect was observed for the N2b component, which is often elicited together with P3b in an active oddball task and act as an indicator of CP and phonological processing (Maiste et al., 1995; Sussman et al., 2004). In particular, our previous study showed strong CP effects of Chinese lexical tones reflected in both N2b and P3b in native listeners (L. Zhang, Xi, et al., 2012). If the highly proficient L2 learners were native-like in both behavioral and neurophysiological measures, we would expect the across-category deviants to elicit stronger N2b and P3b responses than the within-category deviants. These results supported our prediction on the P3b but did not show a stronger N2b response. Similar decoupling of N2b and P3b has been reported in some previous studies. Daffner et al. (2000) specifically suggested that the N2 and P3 are not equally sensitive to the same contextual manipulations, and, as such, the two components may reflect different aspects of processing rather than a unified wave complex. Frenck-Mestre et al. (2005) further argued that P3b is indicative of the categorization of stimuli into separable phonemic categories, but N2b is not. In this study, it could be the case that the adult L2 learners did not consciously detect the phonological difference between the standard and across-category deviants until the P3b time window. Given that the CP effect of lexical tones for N2b was observed in native Chinese listeners (L. Zhang, Xi, et al., 2012), we speculate that age of acquisition could be an important factor to look into, which calls for future examinations to verify whether native-like responses in both N2b and P3b can be observed in CP of lexical tones by bilingual speakers or L2 individuals with high Chinese proficiency and earlier acquisition onset (e.g., those who started learning Chinese before puberty).

Conclusions

In summary, this behavioral and ERP study provides the first investigation into CP of Chinese lexical tones by highly proficient L2 learners. The native-like identification function of the tonal continuum and dissociate P3b responses elicited by across- and within-category deviants in the left recording site indicates successful phonological processing of Chinese lexical tones by proficient L2 learners at the attentive stage. By contrast, the similar MMN responses in the right recording site elicited by across- and within-category deviants indicate automatic acoustic rather than phonological processing at the pre-attentive stage.

Acknowledgments

This research was supported by grants from the Social Science Fund of Beijing (17YYA004) and the Discipline Team Support Program of Beijing Language and Culture University (JC201901) to Linjun Zhang.

References

Buchwald, J. S., Guthrie, D., Schwafel, J., Erwin, R. J., & Vanlancker, D. (1994). Influence of language structure on brain-behavior development. *Brain and Language, 46*(4), 607–619. https://doi.org/10.1006/brln.1994.1033

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage, 22*(3), 1182–1194. https://doi.org/10.1016/j.neuroimage.2004.03.006

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., & Akahane-Yamada, R. (2003). Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. *NeuroImage, 19*(1), 113–124. https://doi.org/10.1016/S1053-8119(03)00020-X

Casserly, E. D., & Pisoni, D. B. (2010). Speech perception and production. *Cognitive Science, 1*(5), 629–647. https://doi.org/10.1002/wcs.63

Chen, F., Peng, G., Yan, N., & Wang, L. (2017). The development of categorical perception of Mandarin tones in four- to seven-year-old children. *Journal of Child Language, 44*(6), 1413–1434. https://doi.org/10.1017/S0305000916000581

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology, 110*(1), 24–30. https://doi.org/10.1016/S0168-5597(98)00033-1

Daffner, K. R., Mesulam, M. M., Scinto, L. F., Calvo, V., Faust, R., & Holcomb, P. J. (2000). An electrophysiological index of stimulus unfamiliarity. *Psychophysiology, 37*(6), 737–747. https://doi.org/10.1111/1469-8986.3760737

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *NeuroReport, 8*(4), 919–924. https://doi.org/10.1097/00001756-199703030-00021

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134,* 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Finney, D. J. (1971). *Probit analysis: A statistical treatment of the sigmoid response curve* (3rd ed.). Cambridge University Press.

Francis, A. L., Ciocca, V., & Ng, B. K. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65*(7), 1029–1044. https://doi.org/10.3758/BF03194832

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2005). Perceiving nonnative vowels. *Journal of Speech, Language, and Hearing Research, 48*(6), 1496–1510. https://doi.org/10.1044/1092-4388(2005/104)

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*(4), 171–189. https://doi.org/10.1177/002383096200500401

Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61*(1), 191–218. https://doi.org/10.1146/annurev.psych.093008.100408

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: Reallocation of neural substrates. *NeuroImage, 21*(2), 494–506. https://doi.org/10.1016/j.neuroimage.2003.09.071

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 291–453. https://doi.org/10.1016/S0095-4470(03)00016-0

Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua, 110*(8), 581–616. https://doi.org/10.1016/S0024-3841(00)00003-6

Hisagi, M., Shafer, V. L., Strange, W., & Sussman, E. S. (2010). Perception of a Japanese vowel length contrast by Japanese and American English listeners: Behavioral and electrophysiological measures. *Brain Research, 1360,* 89–105. https://doi.org/10.1016/j.brainres.2010.08.092

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., Itoh, K., Koshida, I., Yumoto, M., Iramina, K., Kato, N., & Ueno, S. (2001). Brain lateralization for mismatch response to across- and within-category change of vowels. *NeuroReport, 12*(11), 2467–2471. https://doi.org/10.1097/00001756-200108080-00036

Katayama, J., & Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *International Journal of Psychophysiology, 23*(1–2), 33–40. https://doi.org/10.1016/0167-8760(96)00030-X

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103*(30), 11381–11386. https://doi.org/10.1073/pnas.0604821103

Krashen, S. (1981). *Second language acquisition and second language learning*. Pergamon Press.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phonemic boundaries. *Journal of Experimental Psychology, 54*(5), 358–368. https://doi.org/10.1037/h0044417

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61*(5), 379–388. https://doi.org/10.1037/h0049038

Lisker, L. (1975). Is it VOT or a first-formant transition detector? *The Journal of the Acoustical Society of America, 57*(6), 1547–1551. https://doi.org/10.1121/1.380602

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8,* 213. https://doi.org/10.3389/fnhum.2014.00213

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing, 16*(1), 68–89. https://doi.org/10.1097/00003446-199502000-00006

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994–1997. https://doi.org/10.1016/j.cub.2009.09.064

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241–265. https://doi.org/10.1207/s15327078in1003_3

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*(3), 1367–1381. https://doi.org/10.1016/j.cognition.2007.07.002

Minagawa-Kawai, Y., Mori, K., & Sato, Y. (2005). Different brain strategies underlie the categorical perception of foreign and native phonemes. *Journal of Cognitive Neuroscience, 17*(9), 1376–1385. https://doi.org/10.1162/0898929054985482

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language, 92*(1), 26–32. https://doi.org/10.1016/j.bandl.2004.05.005

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Sharma, A., & Dorman, F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *The Journal of the Acoustical Society of America, 106*(2), 1078–1083. https://doi.org/10.1121/1.428048

Shen, G., & Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *The Journal of the Acoustical Society of America, 140*(6), 4396–4403. https://doi.org/10.1121/1.4971765

Shen, G., & Froud, K. (2019). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: An ERP study. *Bilingualism: Language and Cognition, 22*(2), 253–265. https://doi.org/10.1017/S136672891800038X

Sussman, E., Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P., & Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research, 190*(1–2), 128–140. https://doi.org/10.1016/S0378-5955(04)00016-4

Tsao, F. M. (2017). Perceptual improvement of lexical tones in infants: Effects of tone language experience. *Frontiers in Psychology, 8,* 558. https://doi.org/10.3389/fpsyg.2017.00558

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wermke, K., Ruan, Y., Feng, Y., Dobnig, D., Stephan, S., Wermke, P., Ma, L., Chang, H., Liu, Y., Hesse, V., & Shu, H. (2017). Fundamental frequency variation in crying of Mandarin and German neonates. *Journal of Voice, 31*(2), 255.e25–255.e30. https://doi.org/10.1016/j.jvoice.2016.06.009

Xi, J., Jiang, W., Zhang, L. J., & Shu, H. (2010). Categorical perception of VOT and lexical tones in Chinese and the developmental course. *Acta Psychologica Sinica, 41*(7), 572–579. https://doi.org/10.3724/SP.J.1041.2009.00572

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. https://doi.org/10.1016/j.neuroscience.2010.06.077

Xu, Y., Gandour, J., & Francis, A. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063–1074. https://doi.org/10.1121/1.2213572

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*(2), 123–139. https://doi.org/10.1016/j.jml.2012.09.004

Yu, K., Li, L., Chen, Y., Zhou, Y., Wang, R., Zhang, Y., & Li, P. (2019). Effects of native language experience on Mandarin lexical tone processing in proficient second language learners. *Psychophysiology, 56*(11), e13448. https://doi.org/10.1111/psyp.13448

Zhang, L., Xi, J., Wu, H., Shu, H., & Li, P. (2012). Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. *NeuroReport, 23*(1), 35–39. https://doi.org/10.1097/WNR.0b013e32834e4842

Zhang, Y., Kuhl, P. K., Imada, T., Iverson, P., Pruitt, J., Stevens, E., Kawakatsu, M., Tohkura, Y., & Nemoto, I. (2009). Neural signatures of phonetic learning in adulthood: A magnetoencephalography study. *NeuroImage, 46*(1), 226–240. https://doi.org/10.1016/j.neuroimage.2009.01.028

Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., & Li, P. (2012). Universality of categorical perception deficit in developmental dyslexia: An investigation of Mandarin Chinese tones. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 53*(8), 874–882. https://doi.org/10.1111/j.1469-7610.2012.02528.x

Zhao, T. C., & Kuhl, P. K. (2015). Effect of musical experience on learning lexical tone categories. *The Journal of the Acoustical Society of America, 137*(3), 1452–1463. https://doi.org/10.1121/1.4913457

Zheng, H., Minett, J., Peng, G., & Wang, W. S. (2012). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes, 27*(2), 184–209. https://doi.org/10.1080/01690965.2010.520493