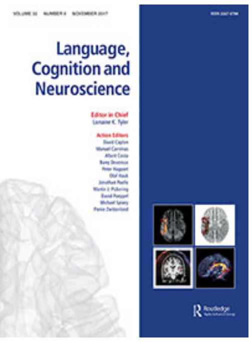

Language, Cognition and Neuroscience

ISSN: 2327-3798 (Print) 2327-3801 (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

The interaction between phonological information and pitch type at pre-attentive stage: an ERP study of lexical tones

Keke Yu, Yacong Zhou, Li Li, Jing’an Su, Ruiming Wang & Ping Li

To cite this article: Keke Yu, Yacong Zhou, Li Li, Jing’an Su, Ruiming Wang & Ping Li (2017) The interaction between phonological information and pitch type at pre-attentive stage: an ERP study of lexical tones, Language, Cognition and Neuroscience, 32:9, 1164-1175, DOI: 10.1080/23273798.2017.1310909

To link to this article: https://doi.org/10.1080/23273798.2017.1310909

LANGUAGE, COGNITION AND NEUROSCIENCE, 2017
VOL. 32, NO. 9, 1164–1175
https://doi.org/10.1080/23273798.2017.1310909

REGULAR ARTICLE

The interaction between phonological information and pitch type at pre-attentive stage: an ERP study of lexical tones

Keke Yu[a], Yacong Zhou[a], Li Li[b], Jing'an Su[a], Ruiming Wang[a] and Ping Li[c]

[a]Guangdong Provincial Key Laboratory of Mental Health and Cognitive Science, and Center for Studies of Psychological Application, School of Psychology, South China Normal University, Guangzhou, People's Republic of China; [b]The Key Laboratory of Chinese Learning and International Promotion, and College of International Culture, South China Normal University, Guangzhou, People's Republic of China; [c]Department of Psychology and Center for Brain, Behavior, and Cognition, Pennsylvania State University, University Park, PA, USA

ABSTRACT
Previous studies on the processing of lexical tones have typically confounded effects due to phonological information (different meanings of words signalled by syllables with different tonal categories) with effects due to specific acoustic information (pitch type: pitch height/pitch contour). The present study is designed to dissociate these two kinds of effects and further investigate the processing of lexical tones at pre-attentive stage by mismatch negativity (MMN). We chose level tones and contour tones in Cantonese to differentiate pitch height from pitch contour, and manipulated tonal category (within-category/across-category) to distinguish phonological information from acoustic information. The results showed clear interactions between tonal category and pitch type in MMN mean amplitude and peak latency, suggesting the interaction between phonological information and pitch type in the pre-attentive processing of lexical tones. These results are discussed in light of cognitive and neural mechanisms underlying auditory processing of lexical tones.

ARTICLE HISTORY
Received 21 May 2016
Accepted 8 March 2017

KEYWORDS
Lexical tone processing; phonological information; acoustic information; pitch type; Mismatch Negativity (MMN)

Introduction

Lexical tone is a crucial phonological feature for tonal languages, with which different categories of tones for the same syllable can express distinct meanings. For example, in Mandarin Chinese, a typical tonal language, "ma1" means "mother", "ma2" means "hemp", "ma3" means "horse" and "ma4" means "scold"(Yip, 2002). How tones are perceived and processed by native speakers of tonal languages has been extensively studied in the literature (e.g. Francis, Ciocca, & Ng, 2003; Gandour, 2006; Jongman, Wang, Moore, & Sereno, 2006; Li, Yang, & Hagoort, 2008). In recent years, there has also been a surge of interest in the neural mechanisms of tonal processing (e.g. Gandour et al., 2003; Gu, Zhang, Hu, & Zhao, 2013; Kaan, Wayland, Bao, & Barkley, 2007; Wang, Sereno, Jongman, & Hirsch, 2003; Wong, Parsons, Martinez, & Diehl, 2004).

Two approaches toward speech processing with regard to lexical tones have been adopted in the literature. The first approach has focused on the processing of acoustic information and phonological information contained in the tonal signal: specifically, the acoustic information that includes pitch features such as fundamental frequency (F0) and pitch contour variations, and the phonological information that expresses different lexical semantics based on specific tonal categories of syllables, for example, in Mandarin, syllable /ba/ with tone-4 (/ba4/) means father in English, but the same syllable in tone-2 (/ba2/) means pulling (Luo et al., 2006; Xi, Zhang, Shu, Zhang, & Li, 2010).

Many studies with this approach have investigated hemispheric lateralisation patterns for acoustic and phonological processing (e.g. Gandour et al., 2000; Xi et al., 2010; Zhang, Xi, Xu, et al., 2011). For example, Gandour et al.'s (2000; Gandour et al., 2004) and Gandour (2006) cross-language studies suggested that the hemispheric lateralisation of tonal processing depends on the kinds of information included in lexical tones: pure acoustic physical features (acoustic information) tend to be processed in the right hemisphere, while semantically distinctive phonological features (phonological information) tend to be processed in the left hemisphere. Xi et al. (2010), Zhang, Xi, Xu, et al. (2011), and Zhang, Xi, Wu, Shu, and Li (2011) further identified the independent and interactive roles of the acoustic versus the phonological information, and demonstrated that for native speakers of tonal language, the acoustic information was mainly processed in the right hemisphere (more specifically the right superior temporal gyrus or STG), while the phonological information was mainly processed in the left hemisphere (more specifically the left

CONTACT Ruiming Wang ✉ wangrm@scnu.edu.cn

middle temporal gyrus or MTG). These differences reflect different brain processes for analysing the same stimuli at different levels, which in turn are conducted in different hemispheres.

In addition to hemispheric lateralisation, some studies have also examined the impacts of acoustic and phonological information on the extent versus the time course during tonal processing. The extent of processing reflects the degree of neural resources involved in cognitive processing, which is primarily measured by the mean amplitude of ERP components like mismatch negativity (MMN), while the time course reflects the time sequence of cognitive processing as it takes place, which could be indicated by the peak latency of ERP components (Duncan et al., 2009). For example, Yu, Wang, Li, and Li (2014) showed that the acoustic information of tones only impacted the extent of tonal processing, while the phonological information of tones affected both the extent and the time course of the processing.

Furthermore, the time course of acoustic and phonological processing has been further examined with regard to two stages of processing: the pre-attentive stage and the attentive stage. Luo et al. (2006) proposed that only acoustic information can be processed at pre-attentive stage, while phonological information is processed at the latter attentive stage. The pre-attentive stage is an early stage at which stimuli are processed automatically without the listener's conscious awareness or attention, and the attentive stage is a stage at which stimuli are processed with attention (Kubovy, Cohen, & Hollier, 1999; Neisser, 1967). This division of time course of processing has been formulated as two-stage model (Luo et al., 2006). However, previous work by Xi et al. (2010) and Yu et al. (2014) have suggested that both acoustic and phonological information can be processed at pre-attentive stage, counter to Luo et al.'s (2006) proposal.

The second approach in the study of tonal processing, unlike the focus on acoustic versus phonological information, has been mainly focused on the processing of physical features of lexical tones, especially pitch height and pitch contour. Pitch height is the relative height of F0, whereas pitch contour refers to the directional variation of F0. They are two basic physical dimensions of pitch (Chandrasekaran, Gandour, & Krishnan, 2007; Gandour, 1983), and are also two types of specific acoustic information. For example, Chandrasekaran, Krishnan, and Gandour (2007) found that the peak latency of MMN, an ERP component that reflects automatic detection of auditory stimuli (Näätänen & Alho, 1997; Näätänen, Gaillard, & Mäntysalo, 1978; see further discussion below), occurred earlier when elicited by large pitch contour differences than when elicited by small pitch contour differences, suggesting that pitch contour impacts the processing of lexical tones. Tsang, Jia, Huang, and Chen (2011) showed that pitch height impacted the size and the latency of MMN, while pitch contour influenced the latency of P3a, an ERP component that is also related to automatic auditory discrimination (Light, Swerdlow, & Braff, 2007; Polich, 2007). Their data revealed the different neurophysiological features during the processing of pitch height versus pitch contour. Finally, Wang, Wang, and Chen (2013) suggested that pitch height and pitch contour can also affect the hemispheric lateralisation of tonal processing. Specifically, pitch height tends to be lateralised to the right hemisphere, while pitch contour tends to be processed in the left hemisphere.

While each of the above two approaches is important in the study of tonal processing, there has been little work designed to understand the interaction between phonological information and acoustic physical features (pitch type) in lexical tone processing. The studies based on the first approach are generally focused on the phonological information conveyed by lexical tones and the holistic acoustic information, that is, the overall pitch features of lexical tones, but do not take into account specific acoustic features such as pitch type (e.g. Luo et al., 2006; Shuai & Gong, 2014). On the other hand, studies based on the second approach primarily explore the processing of pitch type (pitch height and pitch contour) in lexical tones, but do not generally consider the phonological or semantic distinctiveness of the tonal stimuli in the experiments (e.g. Jia, Tsang, Huang, & Chen, 2015; Tsang et al., 2011; Wang et al., 2013). Lack of simultaneous consideration of phonological information and specific acoustic information has led to the lack of understanding of how these variables jointly impact the processing of lexical tones. The current study is designed to address this research gap.

Acoustic information, along with phonological information, is an integral part of lexical tones. But researchers have tried several methods to distinguish them from each other. For example, Chandrasekaran, Krishnan, et al. (2007) compared the processing of Mandarin lexical tones by native and non-native speakers: native Chinese speakers can process both acoustic and phonological information in lexical tones while native English speakers can only process acoustic information. Jia et al. (2015) tested syllables vs. hums, and the hums were generated by eliminating the vowel and the consonant information from the syllables so that they shared the same acoustic information with syllables but contained no phonological information. Xi et al. (2010) adopted within-category and across-category tonal stimuli to differentiate these two types of information. Specifically, the within-category tonal stimuli differ from one another with respect to acoustic features, while the across-category tonal stimuli differ from each other in terms of both acoustic and phonological information. In the current study, we used the paradigm that involves contrasting within-category vs. across-category tonal stimuli as in Xi et al. (2010), to distinguish between the acoustic and the phonological information of tones. This method has been proven effective in a number of previous studies (e.g. Yu et al., 2014; Zhang et al., 2012). Unlike the studies that contrast syllables and hums (e.g. speech vs. non-speech information) or participant types (native vs. non-native speakers) on the processing of lexical tones, this method directly dissociates acoustic and phonological information in speech, and for native speakers.

As for pitch type, we used level tones and contour tones in Cantonese to differentiate pitch height and pitch contour. Cantonese is a southern dialect of Chinese and it is widely spoken in Guangdong Province, Hong Kong and Macao of China. Its tonal system consists of six tones, including three level tones: tone-1(55), tone-3(33), tone-6(22) and three contour tones: tone-2(25), tone-4(21), tone-5(23) (Bauer & Benedict, 1997; Chao, 1947).[1] As reflected by the features of tone pitch, level tones primarily differ in pitch height and have no variation in pitch contour, while contour tones mainly differ in pitch contour. Compared with the tonal system of Mandarin, which is composed of three contour tones (tone-2(35), tone-3(214) and tone-4(51)) and one level tone (tone-1(55)) (Li & Thompson, 1989), the tonal system of Cantonese can provide both across-category level and contour tones. This natural tonal variation in Cantonese with respect to both pitch height and pitch contour, when pitted against the within- vs. across- category in the stimuli, thus enables us to study the interaction as well as the joint contribution of the different types of information in the processing of lexical tones.

MMN is a classic ERP component that reflects the automatic processing of stimuli at pre-attentive stage (Näätänen et al., 1978; Näätänen & Alho, 1997). It is primarily elicited by the oddball paradigm, which consists of frequent standard stimuli (typically 70–90% of total stimuli) and infrequent deviant stimuli (typically 10–30% of total stimuli). MMN usually peaks at approximately 150–250 ms after the stimuli onset and is mainly distributed in the front-central area of the scalp (Näätänen, Paavilainen, Rinne, & Alho, 2007). This method has been widely used in the study of tonal processing (e.g. Chandrasekaran, Krishnan, et al., 2007; Luo et al., 2006; Xi et al., 2010). In the current study, we also utilised MMN to examine the processing of lexical tones at the pre-attentive stage. Specifically, both the mean amplitude and the peak latency of MMN would be examined to detect the extent and the time course of lexical tone processing. In our current experimental design, there were four types of deviant stimuli in our study: across-category pitch height deviant stimuli, within-category pitch height deviant stimuli, across-category pitch contour deviant stimuli and within-category pitch contour deviant stimuli.

We hypothesised that all four types of deviant stimuli would elicit MMNs, on the basis of findings from previous studies as reviewed above (e.g. Tsang et al., 2011; Yu et al., 2014). Findings of MMNs would confirm that both phonological information (reflected by across-/within-category stimuli) and pitch type (reflected by level/contour tones) could be processed at the pre-attentive stage. Furthermore, if phonological information or pitch type play significant roles on tonal processing, we could find differences in the mean amplitudes and/or the peak latencies of MMNs between stimuli with different tonal categories or different pitch types; if they do not impact tonal processing, the MMN mean amplitudes or peak latencies may be the same. More importantly, if phonological information interacts with pitch type during tonal processing, we could observe interactions between tonal category and pitch type in the MMN mean amplitudes and/or peak latencies; if they each affect tonal processing separately, no interactions between them would be obtained.

Method

Participant

Twenty-three undergraduate students from South China Normal University participated in the experiment (9 females, mean age = 20, range: 19–21 years). All the participants were native speakers of Cantonese, but they could also speak Mandarin Chinese, also a tonal language, which is the official language in Mainland China. For all the participants, they began to learn Cantonese from birth and learned Mandarin when they attended to kindergartens. Their self-reported Cantonese and Mandarin proficiency was both high and were not significantly different from each other (on a 7-point scale, Cantonese proficiency: 6.8, Mandarin proficiency: 6.7; $t(1, 22) = 1.37$, $p > .05$, $\eta^2 = 0.02$). However, their self-reported frequency of daily use in Cantonese was significantly higher than that in Mandarin (again on a 7-point scale, Cantonese: 5.6, Mandarin: 5; $t(1, 22) = 5.01$, $p < .05$, $\eta^2 = 0.24$), which showed that the participants used Cantonese more frequently in their daily lives.

All the participants had normal hearing and minimal musical experience, and were right-handed according to a modified Chinese version of the Edinburgh Handedness Inventory (Oldfield, 1971). The study was approved by the Ethics Review Board of South China Normal University. The participants all signed a consent form before they took part in the experiment, and received monetary compensation after the experiment.

Materials and design

The study was a within-subject design with two factors: tonal category (within-category/across-category), and pitch type (level tone/contour tone). The dependent variables were the mean amplitude and the peak latency of MMN elicited by the stimuli.

We chose Cantonese monosyllable /yi/ with four tones: /yi1/, /yi2/, /yi4/, /yi6/ in the study, in which /yi1/ and /yi6/ were monosyllables with level tones and /yi2/ and /yi4/ were monosyllables with contour tones. The stimuli were recorded from a female native Cantonese speaker, at a sampling rate of 44.1 kHz. Xi et al. (2010) generated a tonal continuum of 11 segments to derive within- and across-category tonal stimuli in Mandarin Chinese, and we followed their approach to generate the tonal stimuli for Cantonese used in this study. We used the original different categories of tones as across-category tonal stimuli (/yi1/ vs. /yi6/, /yi2/ vs. /yi4/) and used the Praat software (http://www.fon.hum.uva.nl/praat/) to generate appropriate within-category tonal stimuli. Specifically, we generated /yi2a/ as the within-category stimulus of /yi2/ (contour tone) by narrowing the F0 contour (as shown in Figure 1(b)). We also generated /yi6a/ as the within-category stimulus of /yi6/ (level tone) by decreasing the F0 height of the whole syllable (see Figure 1(c)). By such F0 variations, the generated /yi2a/ and /yi6a/ could be still recognised as tone-2 and tone-6, respectively, but were different in F0 from their original tones. In addition, they would not be misrecognised as tone-3 (/yi3/) or tone-5 (/yi5/), two tones which we did not use in the experiment. The more detailed F0 of the original and the generated tonal stimuli were shown in Table 1.

In total, there were six types of stimuli in the experiment: /yi1/, /yi2/, /yi2a/, /yi4/, /yi6/, /yi6a/. In order to

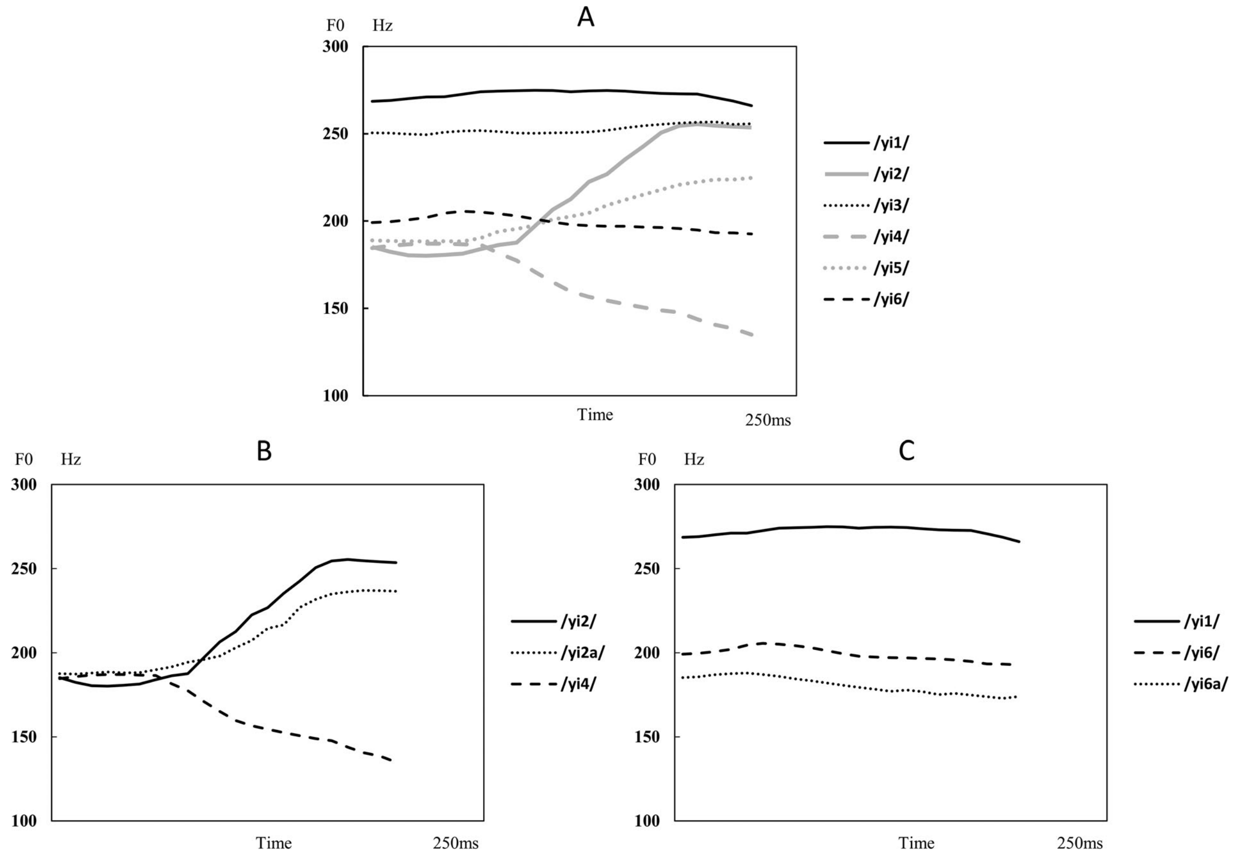

Figure 1. F0 of the original six tones embedded in Cantonese syllable /yi/ (/yi1/, /yi2/, /yi3/, /yi4/, /yi5/, /yi6/) and the two generated tones (/yi2a/ and/yi6a/). 1A: F0 of the original six tones. 1B: F0 of the original tone-2 (/yi2/), the generated tone-2 (/yi2a/) and the original tone-4 (/yi4/). 1C: F0 of the original tone-6 (/yi6/), the generated tone-6 (yi6a/) and the original tone-1 (yi1/).

Table 1. F0 details of the original and the generated tonal stimuli in the experiment.

	F0 onset (Hz)	F0 offset (Hz)	Mean F0 (Hz)	SD of F0 (Hz)
/yi1/	268.59	266.04	272.3045455	2.5256
/yi2/	185.16	253.63	214.2845455	30.90513
/yi2a/	187.48	236.5	208.1309091	20.14558
/yi4/	184.54	134.93	164.8677273	18.82242
/yi6/	199.15	192.66	198.9095455	3.949115
/yi6a/	185.23	173.9	180.5459091	5.180227

ensure that the stimuli were reliable, we conducted a rating experiment in which a separate group of 10 participants performed identification and discrimination tasks on the stimuli before the ERP experiment. In the identification task, the participants were asked to identify the exact tonal type (tone-1, tone-2, tone-4 or tone-6) of the presented stimuli by pressing their corresponding buttons on the keyboard (1: tone-1, 2: tone-2, 4: tone-4, 6: tone-6). The stimuli consisted of all the six stimuli and each stimulus was presented for 10 times randomly. In the discrimination task, the participants were asked to judge whether the tones of the presented pairs belong to a same tone (press "f" on the keyboard) or not (press "j"). There were four kinds of pairs in the experiments: /yi1/ vs. /yi6/, /yi6/ vs. /yi6a/, /yi2/ vs. /yi4/, /yi2/ vs. /yi2a/. The two stimuli in a pair were presented in both directions auditorially and each pair for 20 times (10 times for each direction). The stimulus pairs were presented randomly.

Results of the identification and the discrimination tasks were shown in Table 2. As seen from Table 2, /yi1/ was recognised as tone-1, both /yi2/ and /yi2a/ were recognised as tone-2, /yi4/ was recognised as tone-4 and both /yi6/ and /yi6a/ were recognised as tone-6. Moreover, /yi1/ vs. /yi6/ and /yi2/ vs. /yi4/ were correctly perceived as different tones, while the tones of /yi2/ vs. /yi2a/ and /yi6/ vs. /yi6a/ were considered the same respectively.

The rating experiment confirmed that the stimuli we chose met the requirement of the experimental design. Specifically, /yi1/ vs. /yi6/ were qualified as an across-category level tone contrast, /yi6/ vs. /yi6a/ as a within-category level tone contrast, /yi2/ vs. /yi4/ as an across-category contour tone contrast, and /yi2/ vs. /yi2a/ as a within-category contour tone contrast.

In addition, all the participants in formal ERP experiment reported that they were presented with Cantonese words when they were required to recall the auditory stimuli as much as possible after the experiment. It confirmed that the participants actually processed Cantonese lexical tones in the ERP experiment, and eliminated the potential influence of Mandarin lexical tone experience on the experiment.

Procedure

A classic passive oddball paradigm with standard stimuli mixed with deviant stimuli was used in the experiment (e.g. Näätänen, Pakarinen, Rinne, & Takegata, 2004). There was a total of four blocks and each block contained one type of standard stimuli and one type of deviant stimuli: (1) /yi2/ (standard) vs. /yi4/ (deviant); (2)/yi2/ (standard) vs. /yi2a/ (deviant); (3) /yi6/ (standard) vs. /yi1/ (deviant); /yi6/ (standard) vs. /yi6a/ (deviant). In all four blocks, the number of standard stimuli was 400 and the number of deviant stimuli was 100. The standard and the deviant stimuli in each block were presented pseudo-randomly, and there were at least three standard stimuli between any adjacent deviant stimuli. The presented sequence of the blocks was balanced across the participants. In addition, there were 15 additional standard stimuli presented to the participants at the beginning of each block to help the participants become familiar with the experiment. Each stimulus was presented for 250 ms. The stimulus-onset-asynchrony (SOA) between any two tonal stimuli was 850 ms. The interval between blocks was 2 min. The complete presentation of the tonal stimuli took about 35 min.

The participants were instructed to see a silent movie attentively and ignore the auditory stimuli in the experiment. They didn't need to respond to the auditory stimuli. In order to ensure that the participants focused on the movie, they had to answer five questions about the content of the movie after the experiment. The auditory stimuli were presented to the participants through E-prime 1.1 (Schneider, Eschman, & Zuccolotto, 2002) after the participants saw the movie for five minutes. The five minutes' movie before auditory stimuli was used to help participants acclimate to the experiment better. The whole experiment lasted for 40 minutes, including 5 minutes of movie alone and 35 minutes of movie plus auditory stimuli.

Electroencephalogram recording

Electroencephalogram (EEG) was recorded using a 64-channel (Ag–AgCl) NeuroScan system (NeuroScan, http://www.neuroscan.com/). Electrodes were positioned following the 10–20 system convention. The reference electrode was placed at the tip of the nose. Supra- and infra-orbitally from the left eye was recorded as the vertical electrooculogram (EOG), and the left versus right orbital rim was recorded as the horizontal EOG. The impedance of each electrode was kept below

Table 2. Accuracy of tone identification and discrimination in the rating tasks.

	/yi1/	/yi2/	/yi2a/	/yi4/	/yi6/	/yi6a/	
Accuracy of tone identification	85%	93%	91%	84%	78%	73%	
	/yi1/ vs. /yi6/		/yi6/ vs. /yi6a/		/yi2/ vs. /yi4/		/yi2/ vs. /yi2a/
Accuracy of tone discrimination ("different" response as the correct answer)	91.5%		1.5%		98.5%		10.5%
Accuracy of tone discrimination ("same" response as the correct answer)	8.5%		98.5%		1.5%		89.5%

5 kΩ. EEG and EOG signals were digitised online at 1000 Hz and band-pass filtered from 0.05 to 100 Hz.

Data analysis

Off-line signal processing was carried out using Scan 4.5 (NeuroScan, http://www.neuroscan.com/). The reference electrode was first converted to bilateral mastoid (M1 and M2). The interference of the horizontal and the vertical eye-movements were then eliminated. Data from two male participants were excluded from further analyses due to their excessive eye blinking. After that, the data were segmented for a 700 ms time window, including a 100-ms pre-stimulus baseline. Then, the baseline was corrected and the recorded trials with eye blinks or other activities beyond the range of −80–80 mV were rejected. In addition, the data from the whole-head recordings were off-line band-pass filtered (1–30 Hz) with a finite impulse response filter. At last, the ERPs elicited by standard stimuli and deviant stimuli were obtained by averaging the data from each participant. In addition, only those data with at least 80 accepted deviant trials in each deviant condition were adopted. MMNs were then derived by subtracting the ERPs evoked by the standard stimuli from those evoked by the deviant stimuli.

We selected 9 electrodes (F3, F4, FZ, FC3, FC4, FCZ, C3, C4 and CZ) to further analyse the MMNs, focusing on the distribution of MMN (front-central areas) and following the standard practice used in the MMN literature (e.g. Jiang, Yang, & Yang, 2014; Näätänen et al., 2007; Ren, Yang, & Li, 2009). Considering the grand-average waveforms of the present study and the general time window of MMN in previous studies (e.g. Kaan et al., 2007; Tsang et al., 2011; Wang et al., 2013), we chose 150–350 ms as the time window of MMNs. The peak latencies of MMNs evoked by different kinds of deviant stimuli were detected within the time window. The mean amplitudes were then calculated in a time window ranging from 20 ms before the detected peak of MMN in electrode FZ to 20 ms after that peak for each participant in each condition. For example, if the detected peak of MMN is 200 ms at electrode FZ in a certain condition, the mean amplitude of it would be calculated in the time window of 180–220 ms. The mean amplitudes and the mean peak latencies of MMNs at the nine chosen sites for different conditions were calculated for further statistical analyses.

Results

The grand-average waveforms of standard stimuli and deviant stimuli at nine selected electrode locations were shown in Figures 2 and 3. By subtracting the waveforms elicited by the standard stimuli from the waveforms elicited by the deviant stimuli, we obtained the MMN waveforms as shown in Figure 4.

To ensure that all the deviant stimuli did elicit MMNs, four paired samples *t*-tests were conducted to compare the mean amplitudes of the deviant stimuli and their corresponding standard stimuli in the MMN time window at the selected nine electrodes locations. The results showed that the mean amplitudes between the deviant stimuli /yi1/ and /yi6a/ versus their standard stimuli /yi6/ were significantly different ($t(1, 20) = 2.46$, $p < .05$, $\eta^2 = 0.23$; $t(1, 20) = 2.53$, $p < .05$, $\eta^2 = 0.24$). The mean amplitude difference between the deviant stimuli /yi2a/ and /yi4/ versus their standard stimuli /yi2/ were also significant ($t(1, 20) = 2.26$, $p < .05$, $\eta^2 = 0.20$; $t(1, 20) = 7.67$, $p < .001$, $\eta^2 = 0.75$). Thus the four kinds of deviant stimuli elicited reliable MMNs.

We then conducted two repeated-measures 2 × 2 ANOVAs, with tonal category (within-category/across-category) and pitch type (pitch height/pitch contour) as independent variables, and the mean amplitudes and the peak latencies of MMNs as dependent variables. For all analyses, degrees of freedom were adjusted according to the Greenhouse–Geisser method when appropriate.

MMN mean amplitude

The MMN mean amplitudes at the selected locations of the nine electrodes were shown in Figure 5. The ANOVA results showed that the main effect of tonal category was significant ($F\,(1,20) = 19.48$, $p < .001$, $\eta^2 = 0.49$, within-category < across-category). The interaction between tonal category and pitch type was also significant ($F\,(1,20) = 28.18$, $p < .001$, $\eta^2 = 0.59$). Simple effect analysis suggested that the mean amplitude of across-category contour tones was significantly larger than

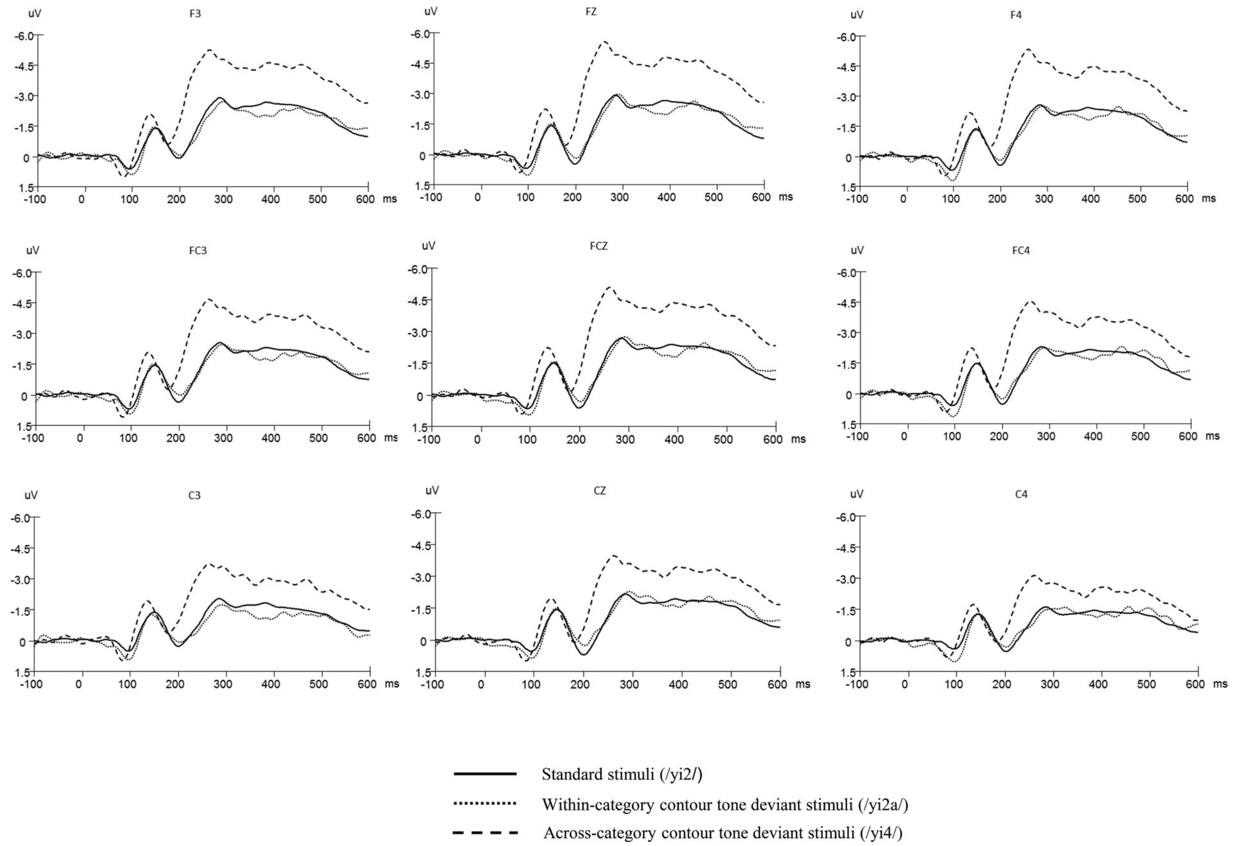

Figure 2. Grand-average waveforms evoked by standard stimuli (/yi2/) and deviant stimuli (/yi2a/ and /yi4/) at nine electrode locations (F3, F4, FZ, FC3, FC4, FCZ, C3, C4 and CZ).

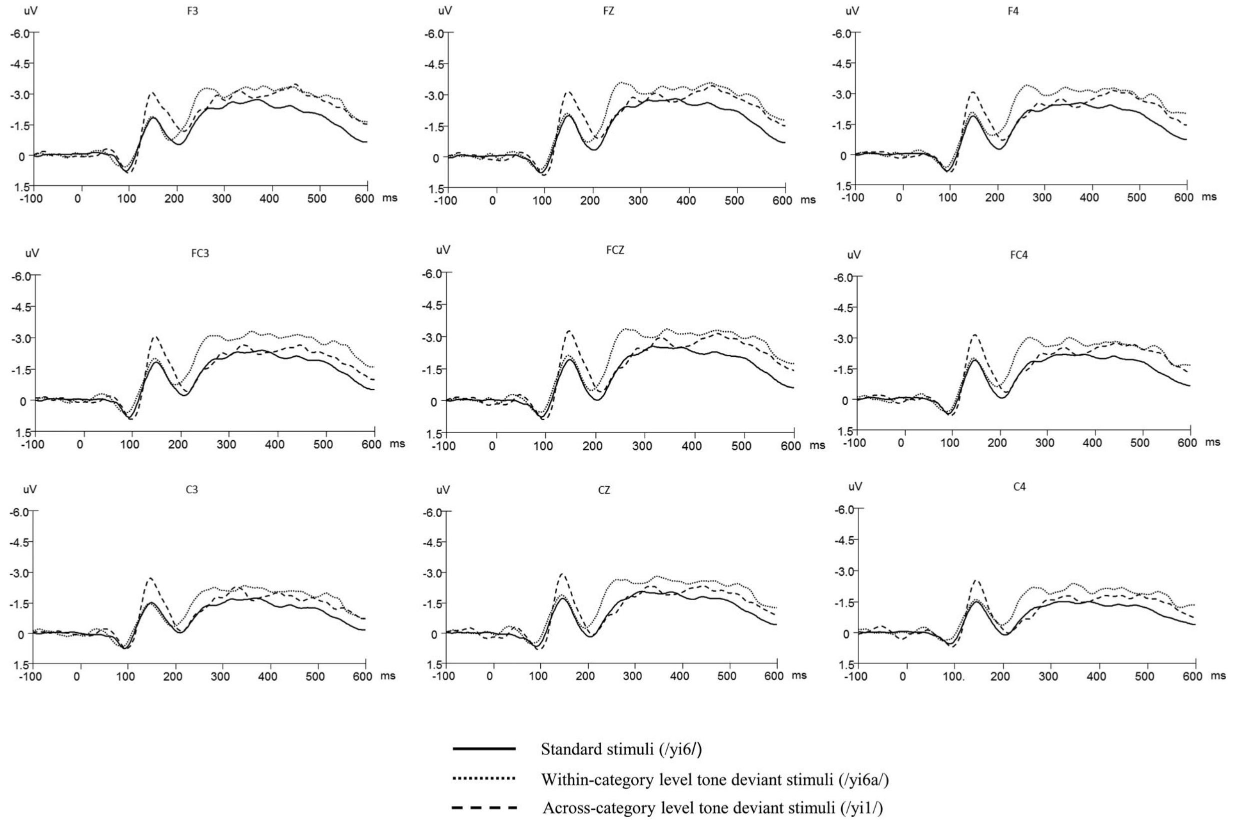

Figure 3. Grand-average waveforms evoked by standard stimuli (/yi6/) and deviant stimuli (/yi6a/ and /yi1/) at nine electrode locations (F3, F4, FZ, FC3, FC4, FCZ, C3, C4 and CZ).

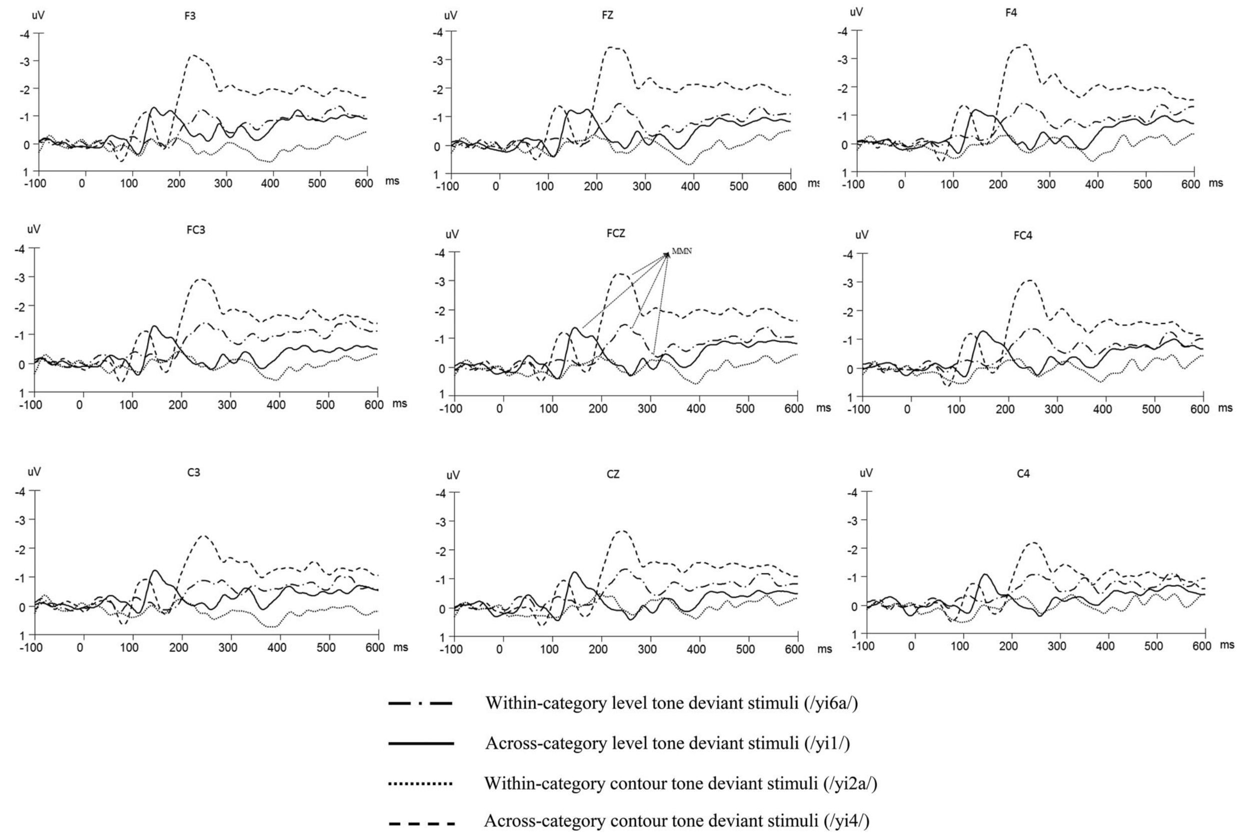

Figure 4. MMNs elicited by different kinds of deviant stimuli at nine electrode locations (F3, F4, FZ, FC3, FC4, FCZ, C3, C4 and CZ).

that of within-category contour tones (F (1,20) = 62.29, $p < .001$, $\eta^2 = 0.76$), but no significant difference between across-category and within-category was found in the level tones (F (1,20) = 0.36, $p > .05$, $\eta^2 = 0.02$). In addition, the mean amplitude of within-category level tones was larger than that of within-category contour tones (F (1,20) = 9.88, $p < .01$, $\eta^2 = 0.33$), whereas the mean amplitude of across-category level tones was smaller than that of across-category contour tones (F (1,20) = 17.15, $p < .001$, $\eta^2 = 0.46$). There was no significant main effect of pitch type (F (1,20) = 1.05, $p > .05$, $\eta^2 = 0.05$).

MMN peak latency

Figure 6 showed the MMN peak latency at the selected locations of the nine electrodes. The main effects of both tonal category and pitch type were significant (F (1,20) = 269.68, $p < .001$, $\eta^2 = 0.93$, within-category > across-category; F (1,20) = 187.48, $p < .001$, $\eta^2 = 0.90$, level tone < contour tone). Moreover, the interaction of tonal category and pitch type was also significant (F (1,20) = 16.63, $p < .01$, $\eta^2 = 0.45$). Simple effect analysis of the interaction showed that the peak latency by pitch type was significantly different between level tones (earlier) and contour tones (later), regardless of whether the tones were within-category or across-category (F (1,20) = 79.65, $p < .001$, $\eta^2 = 0.80$; F (1,20) = 160.97, $p < .001$, $\eta^2 = 0.89$), whereas the peak latency by tonal category was significantly different between across-category (earlier) and within-category (later), regardless of whether the tones were level and contour tones (F (1,20) = 210.64, $p < .001$, $\eta^2 = 0.91$; F (1,20) = 98.27, $p < .001$, $\eta^2 = 0.83$).

More importantly for the interaction, we further conducted a one-way ANOVA with conditional differences (i.e. difference between across- and within-category level tones; difference between across- and within-category contour tones; difference between across-category level and contour tones; difference between within-category level and contour tones) as independent variable and the MMN peak latency as dependent variable. The result of ANOVA showed a main effect of conditional differences (F (1,3) = 8.917, $p < .001$, $\eta^2 = 0.251$), and the post hoc tests with Bonferroni correction showed that the difference between across- and within-category tones was significantly larger for the level tones than for the contour tones, and the difference between level tones and contour tones was significantly larger for the across-category conditions than for the within-category conditions ($ps < .01$) (as shown in Figure 7).

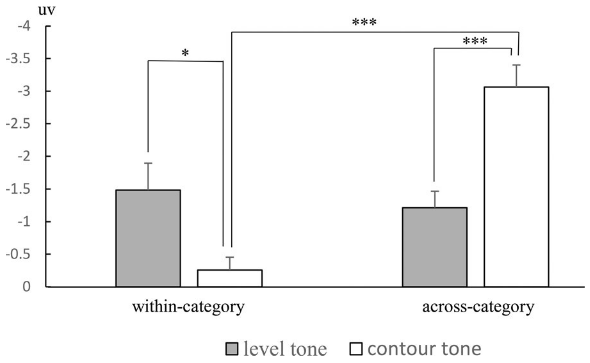

Figure 5. Mean amplitude of MMNs of different kinds of deviant stimuli in nine electrodes (F3, F4, FZ, FC3, FC4, FCZ, C3, C4 and CZ). Error bars represent one standard error, * $p < .05$, *** $p < .001$.

Discussion

The present study investigated the auditory processing of lexical tones at pre-attentive stage via the measures of both the mean amplitude and the peak latency of MMN, when both phonological information and pitch type are considered. Our results provided the first systematic electrophysiological evidence for the interaction between phonological information and pitch type on both the extent and the time course of lexical tone processing.

As mentioned in Introduction, there were two competing views on the time course of acoustic and phonological processing (Luo et al., 2006 vs. Xi et al., 2010; Yu et al., 2014). In the present study, the results showed clear MMNs elicited by four types of deviant stimuli with tonal category or pitch type variations, which was clearly inconsistent with the two-stage model in Luo et al. (2006) but supported Xi et al. (2010) and Yu et al. (2014)’ analyses. Moreover, in addition to contour tones investigated in previous studies, we further examined level tones and revealed that both acoustic and phonological information in level tones could also be processed at pre-attentive stage.

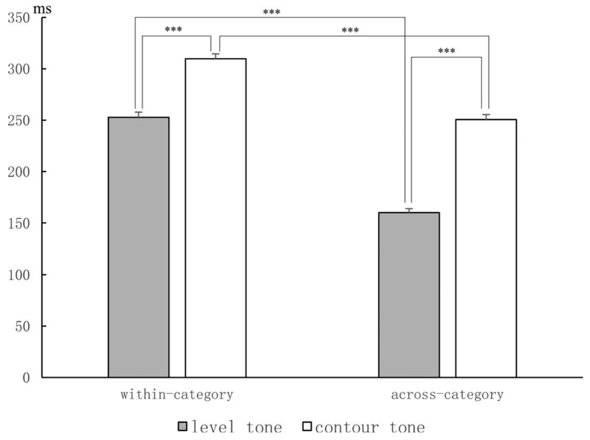

Figure 6. The average of MMN peak latencies of different kinds of deviant stimuli in nine electrodes (F3, F4, FZ, FC3, FC4, FCZ, C3, C4 and CZ). Error bars represent one standard error, * $p < .05$, *** $p < .001$.

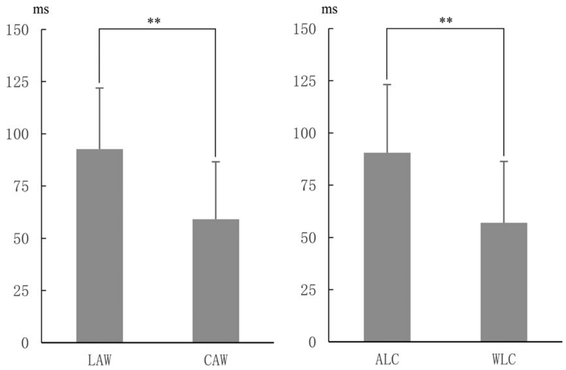

Figure 7. The interaction between tonal category and pitch type in MMN peak latency. LAW (Level Tones, Across vs. Within Category) refers to the MMN peak latency difference between across- and within-category level tones; CAW (Contour Tones, Across vs. Within Category); ALC (Across Category, Level vs. Contour Tone); WLC (Within Category, Level vs. Contour Tone). Error bars represent one standard error, **$p < .01$.

However, as to the specific time course of acoustic and phonological processing at pre-attentive stage, Xi et al. (2010) considered that these two types of processing took place in parallel. But Yu et al. (2014) found that phonological information could be processed earlier than acoustic information. In our study, across-category tonal stimuli were found to be processed earlier than within-category tonal stimuli when both pitch types of lexical tones were considered. Thus our data are consistent with those from Yu et al. (2014). As interpreted in Yu et al. (2014), the result may be due to the enhanced difference in across-category tones, because the across-category tones differed in both phonological information and acoustic information, whereas the within-category tones differed only in acoustic information. Furthermore, the top-down effect elicited by different phonological information in across-category tones may also facilitate tone processing, given that the different phonological information expresses distinct meanings of words.

In this study, we also compared the MMN peak latency difference between within- vs. across-category conditions for level tones vs. contour tones. The results showed that the difference between within- and across-category tones was larger for level tones than for contour tones, suggesting the modulation of pitch type on the time course of acoustic and phonological processing in lexical tones. The reason may lie in the pitch features of level tones versus contour tones, as the pitch of level tones differs from the beginning of the syllable and remains the same over the whole syllable, whereas the pitch of contour tones are similar at the beginning but differ gradually along the syllable (as shown in Figure 1). Thus the variation of phonological information, along with pitch change, between level tones may turn up earlier than that between contour tones. This feature allows the listener to detect the phonological difference of level tones earlier, which can result in a larger MMN peak latency difference between within- and across-category level tones.

With regard to pitch type, the present study revealed that pitch height was processed earlier than pitch contour at pre-attentive stage. The different pitch features of level tones versus contour tones can also account for this finding, because the pitch difference of level tones (pitch height) emerges earlier than that of contour tones (pitch contour). Wang et al. (2013) also found that pitch height was processed earlier than pitch contour by using level and contour tones, but the level tones or the contour tones used in their study were carried by a single vowel /a/. For tonal languages, pitch variations occur at the whole syllable level, consisting of consonant plus vowel. Wang et al.'s study may therefore not reflect the entire process of lexical tone during comprehension. To correct this problem, the stimuli used in the present study were real Cantonese syllables, which involve both an onset consonant plus a vowel. Our study therefore provided more consistent and reliable evidence to the time course of pitch height and pitch contour. Moreover, we used tonal category (within-/across-category) to distinguish acoustic information from phonological information in level tones and contour tones, and further confirmed that regardless of the phonological information in tone contrasts (within-category or across-category tones), pitch height was always processed prior to pitch contour. However, there was a clear interaction effect such that the difference between level tones and contour tones was larger for the across-category condition than for the within-category condition. The result further suggests that, in addition to the modulation of pitch type on phonological information, phonological information also modulate the time course of the processing of pitch type, although the latter seems to play a greater role on the time course of level tone processing. Such interactions may also reflect an earlier top-down effect elicited by phonological information on the processing of pitch height than pitch contour.

In terms of the extent of processing, our data indicated that contour tones produced larger MMN mean amplitude for the across-category than for the within-category conditions, which clearly reflects the effect of phonological information on pitch contour. This finding was consistent with Yu et al. (2014) and Xi et al. (2010) in which contour tones were used. However, the present study further showed that the extent of level tone processing, in contrast to that of contour tone processing, was not affected by the phonological information. Moreover, when examined within each category, the across-category tones produced smaller amplitude for the level tones than for the contour tones, whereas the within-category tones produced larger amplitude for the level tones than for the contour tones, indicating a clear interaction between phonological information and pitch type.

Such interaction between phonological information and pitch type reflected in the processing extent may be due to the categorical features of level and contour tones: for example, some previous studies indicates that category perception of level tones was not as obvious as that of contour tones (Francis et al., 2003; Xu, Gandour, & Francis, 2006). Category perception of lexical tones means that listeners can perceive the stimuli with continuous pitch variations as discrete tonal categories. Listeners are more sensitive to the stimuli in different tonal categories, but are insensitive to the stimuli in the same tonal category. Because the contour tones are more obvious due to changing temporal characteristics, they may be more easily perceived than the level tones across categories; however, they may be more difficult to detect within the same category, as compared with level tones. Thus, the MMN mean amplitude differences shown in our data might reflect that more neural resource were involved in detecting the difference between contour tones than between level tones for across-category conditions, but more neural resource were involved in detecting the difference between level tones than between contour tones for within-category conditions.

To conclude, previous studies confounded the effects of pitch type with phonological information on lexical tone processing. In the present study, we disassociated the effects due to pitch type from those due to phonological information and further clarified the auditory processing of lexical tones at pre-attentive stage. On the basis of these results, we proposed an integrated mechanism for tonal processing. The mechanism considered acoustic and phonological information in both level tones and contour tones, and incorporated the extent and the time course of pitch type (pitch height and pitch contour) and phonological features along with their interaction. First, our findings suggest that acoustic information (more specifically pitch height and pitch contour) and phonological information in the speech stimuli could all be processed at pre-attentive stage for native listeners. Second, they are processed in an interactive way: for the extent of processing, phonological information only influence the processing of pitch contour but not pitch height; for the time course of processing, pitch height is always processed earlier than pitch contour regardless of the phonological information, but the phonological information plays a greater role on the processing of pitch height. In particular, phonological information is processed earlier than acoustic information regardless of pitch type, but the phonological variation signalled by pitch height difference can be processed earlier than that by pitch contour difference. Future studies should examine whether the processing of lexical tones at attentive stage is similar to pre-attentive stage, and whether different levels of attention contribute differently to the processing of specific acoustic and phonological features in lexical tones.

Note

1. The numbers in brackets indicate the tone pitch of each tone: the first number refers to the pitch onset of the tone and the second number refers to the pitch offset of the tone (1 means the lowest and 5 the highest pitch).

Acknowledgements

We thank Huimin Zheng and Manqiong Shen for assisting with preparing the experimental materials and collecting ERP data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was funded by the National Natural Science Foundation of China [31571142], the Excellent Young Teacher Foundation of University in Guangdong [HS2015006]. Partial support was also provided by the National Science Foundation [BCS-1349110].

References

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology & Neuroscience*, *25*(3–4), 195–210.

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

Chao, Y. R. (1947). *Cantonese primer*. Cambridge: Cambridge University Press.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., … Van Petten, C. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, *120*(11), 1883–1908. doi:10.1016/j.clinph.2009.07.045

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

Gandour, J. (1983). Tone perception in far Eastern-Languages. *Journal of Phonetics*, *11*, 149–175.

Gandour, J. (2006). Brain mapping of Chinese speech prosody. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics (Vol. 1: Chinese)* (pp. 308–319). New York, NY: Cambridge University Press.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., … Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage*, *23*(1), 344–357. doi:10.1016/j.neuroimage.2004.06.004

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A cross-linguistic PET study of tone perception. *Journal of Cognitive Neuroscience*, *12*(1), 207–222. doi:10.1162/089892900561841

Gandour, J., Xu, Y., Wong, D., Dzemidzic, M., Lowe, M., & Li, X. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping*, *20*(4), 185–200. doi:10.1002/hbm.10137

Gu, F., Zhang, C., Hu, A., & Zhao, G. (2013). Left hemisphere lateralization for lexical and acoustic pitch processing in Cantonese speakers as revealed by mismatch negativity. *Neuroimage*, *83* (5), 637–645. doi:10.1016/j.neuroimage.2013.02.080

Jia, S., Tsang, Y. K., Huang, J., & Chen, H. C. (2015). Processing Cantonese lexical tones: Evidence from oddball paradigms. *Neuroscience*, *305*, 351–360. doi:10.1016/j.neuroscience.2015.08.009

Jiang, A., Yang, J., & Yang, Y. (2014). MMN responses during implicit processing of changes in emotional prosody: An ERP study using Chinese pseudo-syllables. *Cognitive Neurodynamics*, *8*(6), 499–508. doi:10.1007/s11571-014-9303-3

Jongman, A., Wang, Y., Moore, C., & Sereno, J. (2006). Perception and production of Chinese tones. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *Handbook of East Asian Psycholinguistics (Vol. 1: Chinese)* (pp. 209–218). New York, NY: Cambridge University Press.

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research*, *1148*(4), 113–122. doi:10.1016/j.brainres.2007.02.019

Kubovy, M., Cohen, D. J., & Hollier, J. (1999). Feature integration that routinely occurs without focal attention. *Psychonomic Bulletin & Review*, *6*(2), 183–203. doi:10.3758/BF03212326

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

Li, X., Yang, Y., & Hagoort, P. (2008). Pitch accent and lexical tone processing in Chinese discourse comprehension: An ERP study. *Brain Research*, *1222*(3), 192–200. doi:10.1016/j.brainres.2008.05.031

Light, G. A., Swerdlow, N. R., & Braff, D. L. (2007). Preattentive sensory processing as indexed by the MMN and P3a brain responses is associated with cognitive and psychosocial functioning in healthy adults. *Journal of Cognitive*

Neuroscience, *19*(10), 1624–1632. doi:10.1162/jocn.2007.19.10.1624

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. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(51), 19558–19563. doi:10.1073/pnas.0607065104

Näätänen, R., & Alho, K. (1997). Mismatch negativity–the measure for central sound representation accuracy. *Audiology & Neurotology*, *2*(5), 341–353. doi:10.1159/000259255

Näätänen, R., Gaillard, A. W. K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Actapsychologica*, *42*(4), 313–329. doi:10.1016/0001-6918(78)90006-9

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. doi:10.1016/j.clinph.2007.04.026

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm – clinical neurophysiology. *Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology*, *115*(1), 140–144. doi:10.1016/j.clinph.2003.04.001

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.

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

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148. doi:10.1016/j.clinph.2007.04.019

Ren, G. Q., Yang, Y., & Li, X. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience*, *162*(1), 87–95. doi:10.1016/j.neuroscience.2009.04.021

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Shuai, L., & Gong, T. (2014). Temporal relation between top-down and bottom up processing in lexical tone perception. *Frontiers in Behavioral Neuroscience*, *8*, 97. doi:10.3389/fnbeh.2014.00097

Tsang, Y. K., Jia, S., Huang, J., & Chen, H. C. (2011). ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height. *Neuroscience Letters*, *487*(3), 268–272. doi:10.1016/j.neulet.2010.10.035

Wang, X. D., Wang, M., & Chen, L. (2013). Hemispheric lateralization for early auditory processing of lexical tones: Dependence on pitch level and pitch contour. *Neuropsychologia*, *51*(11), 2238–2244. doi:10.1016/j.neuropsychologia.2013.07.015

Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of mandarin lexical tone. *Journal of Cognitive Neuroscience*, *15*(7), 1019–1027. doi:10.1162/089892903770007407

Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience the Official Journal of the Society for Neuroscience*, *24*(41), 9153–9160. doi:10.1523/JNEUROSCI.2225-04.2004

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. doi:10.1016/j.neuroscience.2010.06.077

Xu, Y., Gandour, J. T., & Francis, A. L. (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. doi:10.1121/1.2213572

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Yu, K., Wang, R., Li, L., & Li, P. (2014). Processing of acoustic and phonological information of lexical tones in Mandarin Chinese revealed by mismatch negativity. *Frontiers in Human Neuroscience*, *8*(3), 729. doi:10.3389/fnhum.2014.00729

Zhang, L., Xi, J., Wu, H., Shu, H., & Li, P. (2011). Electrophysiological evidence of categorical perception of Chinese lexical tones in attentive condition. *Neuroreport*, *23*(1), 35–39. doi:10.1097/WNR.0b013e32834e4842

Zhang, L., Xi, J., Xu, G., Shu, H., Wang, X., & Li, P. (2011). Cortical dynamics of acoustic and phonological processing in speech perception. *PLoS one*, *6*(6), e20963. doi:10.1371/journal.pone.0020963

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 Psychiatry and Psychology*, *53*, 874–882. doi:10.1111/j.1469-7610.2012.02528.x