Journal of Neurolinguistics 61 (2022) 101039

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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

Distinct but integrated processing of lexical tones, vowels, and consonants in tonal language speech perception: Evidence from mismatch negativity

Keke Yu[a,1], Yuan Chen[a,1], Menglin Wang[a], Ruiming Wang[a,*], Li Li[b,**]

[a] *Key Laboratory of Brain Cognition and Education Science of Ministry of Education, 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, China*

[b] *The Key Laboratory of Chinese Learning and International Promotion, and College of International Culture, South China Normal University, Guangzhou, China*

ARTICLE INFO

Keywords:
Tonal languages
Lexical tones
Vowels
Consonants
Speech perception
MMN

ABSTRACT

The processing of lexical tones, vowels, and consonants is significant in tonal language speech perception. However, it remains unclear whether their processing is similar or distinct concerning the extent and time course and whether their processing is independent or integrated. Thus in the present study, we conducted two event-related potential (ERP) experiments to explore how native speakers of Cantonese process lexical tones (including level and contour tones), vowels, and consonants in real vs. pseudo-Cantonese words with mismatch negativity (MMN). The MMN amplitudes and latencies showed that lexical tones and vowels were processed similarly in extent and time course. Lexical tones and consonants were processed differently in extent and time course. Vowels and consonants were processed to similar extents but over different time courses. Lexicality (real words vs. pseudowords) and tonal type (level vs. contour tones) modulated the differences in the extent and time courses of processing between lexical tones/vowels and consonants. The MMN additivity analyses further suggested that the processing of lexical tones and vowels, lexical tones and consonants, and vowels and consonants were integrated regardless of lexicality and tonal type. The results revealed that distinct but integrated processing occurs for lexical tones, vowels, and consonants in the speech perception of tonal languages. The findings provided neurophysiological evidence for the mechanism underlying tonal language spoken word recognition.

1. Introduction

In tonal languages such as Mandarin and Thai, syllables usually consist of consonants, vowels, and lexical tones (Gussenhoven & Jacobs, 2013). Both consonants and vowels also exist in non-tonal languages such as English. However, lexical tones are special phonological features that only exist in tonal languages (Yip, 2002). Different word meanings can be expressed when distinct tonal

Abbrevation: MMN, Mismatch negativity.

[*] Corresponding author.

[**] Corresponding author.

E-mail addresses: wangrm@scnu.edu.cn (R. Wang), liliwrm@163.com (L. Li).

[1] Co-first authors.

https://doi.org/10.1016/j.jneuroling.2021.101039
Received 2 February 2021; Received in revised form 1 October 2021; Accepted 1 November 2021
Available online 6 November 2021

categories are superimposed on the same syllable (e.g., Mandarin syllable /ya/ with Tone 1 means duck, while the same syllable with Tone 3 means elegant). Thus, lexical tones, like consonants and vowels, play significant roles in tonal language speech perception. An appropriate spoken word recognition model to the tonal languages should consider all of the three phonological features (e.g., Tong, McBride, & Burnham, 2014; Ye & Connine, 1999). Many previous studies have been conducted on the cognitive and neural mechanism of tonal language processing, especially the neural processing of lexical tones in native and nonnative speakers (e.g., Chandrasekaran, Krishnan, & Gandour, 2007; Deng, Chandrasekaran, Wang, & Wong, 2016; Yu et al., 2019; Zatorre & Gandour, 2008). But it remains unclear how lexical tones, vowels, and consonants are processed during tonal language speech perception.

1.1. Similar vs. distinct processing

Many studies have explored how native speakers of tonal languages process lexical tones, vowels, and consonants regarding whether their cognitive and neural processing features are similar or different. Schirmer, Tang, Penney, Gunter, and Chen (2005) found that adult native Cantonese (a tonal language) speakers showed similar early frontal negativity and late centro-parietal positivity (in both latency and amplitude) to the Cantonese lexical tone and rhyme variations in the last words of Cantonese sentences. In event-related potential (ERP) studies, the amplitude of an ERP component has been shown to reflect the extent of cognitive processing. On contrast, the latency of an ERP component has been shown to reflect the time course of cognitive processing Duncan et al., 2009. Thus, the previous study suggested that the processing of lexical tones and that of vowels are similar in terms of both extent and time course. Lee et al. (2012) revealed that the processing of Mandarin lexical tones and vowels yield similar ERP responses (adult-like mismatch negativity (MMN) or positive mismatch response (P-MMR)) in native Mandarin preschool children. Choi, Tong, Gu, Tong, and Wong (2017) showed that adult native Cantonese speakers exhibit similar MMN responses (peak latency and topography) to lexical tone and vowel variations in Cantonese lexical tone and vowel combinations (vowels superimposed onto lexical tones). The results of all these studies suggested that the processing of lexical tones and vowels were similar. However, Hu et al.'s (2012) found that when the vowels and lexical tones in the last words of Mandarin idioms varied, adult native Mandarin speakers showed different ERP responses to vowel and lexical tone variations. Compared with lexical tone variations, vowel variations elicited a larger and early negative component and larger and early N400 but a smaller and late positive component. The results indicated that the processing of lexical tones and vowels differed in terms of both extent and time course.

In Lee et al. (2012), the more significant lexical tone or vowel variations in Mandarin syllables elicited adult-like mismatch negativity (MMN), while the minor lexical tone or vowel variations elicited a positive mismatch response (P-MMR). However, both larger and smaller variations of consonants consistently elicited P-MMRs. Thus the study further indicated that the processing of consonants might be different from that of lexical tones and vowels, as they elicited different ERP responses. Tong et al. (2014) also suggested there are differences in processing between lexical tones and consonants. In their study, native Cantonese children (second graders) showed P-MMRs to consonants and MMN to lexical tones. Moreover, Luo et al.'s (2006) early study conducted in adult native Mandarin speakers showed that the MMN amplitude of consonants in the left hemisphere was larger than that in the right hemisphere, while that of lexical tones in the left hemisphere was smaller than that in the right hemisphere. The results indicate that the processing of lexical tones and consonants differ in terms of extent and hemispheric lateralization.

Based on the findings from these studies, the processing between lexical tones and consonants, between vowels and consonants seem to be distinct (Lee et al., 2012; Luo et al., 2006; Tong et al., 2014). However, it remains controversial whether lexical tones, especially the extent and time course of processing, are different from vowels (Choi et al., 2017; Hu, Gao, Ma, & Yao, 2012; Lee et al., 2012; Schirmer et al., 2005).

1.2. Independent vs. integrated processing

Previous studies have also investigated whether the processing of lexical tones is independent of or integrated with that of vowels or consonants. They have tried to propose revised models to illustrate the spoken word recognition mechanism specific to tonal languages based on the TRACE model (Mcclelland & Elman, 1986). The TRACE model is a classical model that mainly focuses on non-tonal languages. It considers that the spoken words consist of three levels, the features level, the phonemes level, and the words level. The features level contains the acoustic features of vowels and consonants (e.g., burst, friction). The phonemes level have specific vowels and consonants (e.g.,/a/,/t/). The words level includes the input word and its related words (e.g., cat, hat). The model hypothesizes that each level has the detectors for its inclusive elements. When a particular detector is activated during spoken word recognition, it will inhibit the activation of detectors at the same level. In contrast, it will activate and interact with the detectors at different levels. According to this view, the processing of vowels and consonants would be independent of each other. Nevertheless, the model does not explain how lexical tones are stored in the mental lexicon and whether their processing is independent of or integrated with vowels and consonants' processing.

Therefore, in the revised models, researchers tried to finger out further how lexical tones are processed and interacted with vowels and consonants (Choi et al., 2017; Gao et al., 2019; Shuai & Malins, 2016; Tong, McBride, & Burnham, 2014; Ye & Connine, 1999). Ye and Connine (1999) first supplemented a tonemes level in the TRACE model, which is parallel to the phoneme level and stores the lexical tones (e.g., for Mandarin, Tone 1, Tone 2, Tone 3, and Tone 4). Consistent with the TRACE model, they considered that the processing of lexical tones, vowels, and consonants is independent. It is similar to the view presented in Shuai and Malins (2016). But Tong, McBride, and Burnham (2014) suggested that the phonemes (also referred to as phonemes & tonemes) level also includes the lexical tones, and the processing of lexical tones, vowels, and consonants is integrated. Choi et al. (2017) further considered lexical tones and vowels as phonological units and supported that lexical tones and vowels are processed integrally. However, the recent study by Gao et al. (2019) proposed that vowels and consonants are composed of atonal syllables and processed integrally, whereas the processing of atonal syllables (consonants or vowels) occurs independently of the processing of lexical tones.

As seen from the points in the revised models, it remains controversial whether the processing of lexical tones, vowels, and consonants is independent or integrated. Ye and Connine (1999) and Shuai and Malins (2016) (referred to as the independent models) both hypothesized independent processing of these phonological features, whereas the other models (referred to as the integration models) considered integrated processing for them. However, even these integration models still have distinct views, especially on whether processing lexical tones and vowels, lexical tones and consonants are integrated (Choi et al., 2017; Gao et al., 2019; Tong et al., 2014).

1.3. The present study

How lexical tones, consonants, and vowels are processed is a significant issue for tonal language speech perception and is also a fundamental question on spoken word recognition. Although many previous studies have explored this question, findings from these studies still could not provide a clear picture to understand the mechanism underlying tonal language speech processing. Firstly, whether the processing of lexical tones and vowels differ in terms of the extent or time course is controversial. A potential reason for the inconsistent findings may be the speech materials used in these studies. Previous studies adopted several types of materials, including the sentences (Schirmer et al., 2005), idioms (Hu et al., 2012), syllables (Lee et al., 2012), and lexical tone and vowel combinations (Choi et al., 2017), the semantic contexts differentiated by these speech materials may modulate lexical tones and vowels processing. The TRACE model posits that the words and phonemes levels interact, which also suggests the role of word meanings in processing phonological features.

Secondly, we now have evidence for the different processing between lexical tones and consonants (Lee et al., 2012; Luo et al., 2006; Tong, Mcbride, et al., 2014). However, except for Lee et al. (2012), little evidence demonstrated the distinct processing between vowels and consonants. Moreover, the TRACE model hypothesizes that the features level interacts with the phonemes level. For lexical tones, Tong, McBride, and Burnham (2014)'s revised model considered that the features level also contained the acoustic features of lexical tones like pitch height and pitch contour. Previous studies have suggested that the processing of pitch height is distinct from that of pitch contour (e.g., Chandrasekaran, Gandour, & Krishnan, 2007; Tsang, Jia, Huang, & Chen, 2011; Wang, Wang, & Chen, 2013). These tonal features also affected native speakers' processing of lexical tones (Yu et al., 2017). However, previous studies concerning lexical tone, vowel, and consonant processing did not differentiate pitch height from pitch contour (e.g., Choi et al., 2017; Lee et al., 2012; Schirmer et al., 2005). It remains unclear whether and how the tonal features affect the processing of these phonological features.

Finally, the TRACE model does not have specific hypotheses on the independent or integrated processing between lexical tones and vowels/consonants. In order to resolve the fundamental issue for tonal languages, previous studies have proposed several revised models (Choi et al., 2017; Gao et al., 2019; Shuai & Malins, 2016; Tong, McBride, & Burnham, 2014; Ye & Connine, 1999). But these models suggested distinct views, i.e., independent vs. integrated view, on processing these phonological features.

In order to explore the unresolved issues in previous studies and models, the present study aimed to examine the processing of lexical tones, vowels, and consonants during tonal language speech perception and provide new evidence to the mechanism underlying tonal language spoken word recognition from the neurophysiological aspect. Specifically, we explored native Cantonese speakers' processing of lexical tones, vowels, and consonants regarding the features (extent and time course) and integration with the ERP technique, especially MMN.

In Cantonese, the tonal system consists of six tones (Matthews & Yip, 2011). Based on their pitch features, these six tones can be divided into two types, level (Tones 1, 3, and 6) and contour tones (Tones 2, 4, and 5). The level tones mainly differ in pitch height, while the contour tones mainly differ in pitch contour. By differentiating the level from contour tones in Cantonese, we could detect whether the tonal features modulate the processing between lexical tones and vowels/consonants in the study. Considering the potential role of semantic context, we also manipulated the lexicality of Cantonese words (real words vs. pseudowords) in the experiments. The lexical tones/vowels/consonants in the real words can signify different meanings, while those in pseudowords cannot. Some previous studies suggested that semantic/phonological information affected the processing of lexical tones in tonal languages (e.g., Shuai & Gong, 2014; Xi, Zhang, Shu, Zhang, & Li, 2010; Yu, Wang, Li, & Li, 2014). Manipulating the words' lexicality can reveal the role of semantic/phonological information in processing lexical tones, vowels, and consonants. It can also indicate the influence of the words level on the phonemes level under the framework of the TRACE model.

MMN is a classic ERP component that could reflect the automatic processing of speech stimuli at the pre-attentive stage. It usually peaks at approximately 150–250 ms after stimulus onset and is mainly distributed in the frontal-central area of the scalp (Näätänen, Paavilainen, Rinne, & Alho, 2007). It is usually elicited by deviant stimuli in the oddball paradigm (the paradigm usually consists of two types of stimuli, frequent standard stimuli (typically 70–90% of the total stimuli) and infrequent deviant stimuli (typically 10–30% of the total stimuli)). As the amplitude and peak latency of ERP components can reflect the extent and time course of cognitive processing (Duncan et al., 2009), we planned to detect the extent and time course of Cantonese lexical tones, vowels, and consonants processing via the MMN amplitude and peak latency.

Moreover, we applied the MMN additivity approach to explore the integration between the processing of lexical tones and vowels, lexical tones and consonants, and vowels and consonants. This approach has been used in several previous studies to effectively detect independent vs. integrated processing between different auditory cognitive processes (e.g., Caclin et al., 2006; Choi et al., 2017; Lidji et al., 2010). The logic of this approach is that if the sum of the MMN amplitudes of two single-dimensional deviants (e.g., "tone"-MMN + "vowel"-MMN) is larger than the MMN amplitude of a double-dimensional deviant (e.g., "tone + vowel"-MMN), the processing of the two dimensions is integrated. If the sum of the MMN amplitudes of two single-dimensional deviants (e.g., "tone"-MMN + "vowel"-MMN) is smaller than or the same as the MMN amplitude of a double-dimensional deviant (e.g., "tone + vowel"-MMN), the processing of these two dimensions is independent.

In summary, we conducted two ERP experiments in the present study. In Experiment 1, we adopted the passive oddball paradigm and MMN to explore how native speakers process level tones, vowels, and consonants in real vs. pseudo-Cantonese words. In Experiment 2, we explored how native speakers process contour tones, vowels, and consonants in real vs. pseudo-Cantonese words with the same method as in Experiment 1. Based on previous studies, we hypothesized that if the MMN amplitudes and/or peak latencies of lexical tone, vowel, and consonant deviants differ, the extent and/or time course of lexical tone, vowel, and consonant processing differ. If the sum of the MMN amplitudes of lexical tone and vowel deviants is larger than the MMN amplitude of lexical tone + vowel deviants, the processing of lexical tones and vowels may be integrated. The integration between lexical tone and consonant processing, and vowel and consonant processing is similar to that between lexical tone and vowel processing. Moreover, if the extent, time course, or integration differ between real words and pseudowords, the lexicality may affect the processing of lexical tones, vowels, and consonants. Last, as the tonal type is related to lexical tones, if the MMN amplitude and/or peak latency differences between lexical tones and vowels and/or between lexical tones and consonants are different between Experiments 1 and 2, the tonal type may affect the differences in the extent and/or time courses of processing between lexical tones and vowels/consonants. If the integration between lexical tone and vowel/consonant processing differs between Experiments 1 and 2, the tonal type may also influence the integration between lexical tone and vowel/consonant processing.

2. Experiment 1

2.1. Participants

Twenty-four undergraduate students from South China Normal University participated in the experiment (10 males, mean age: 20, age range: 18–25). All the participants were native speakers of Cantonese. They could also speak Mandarin, the official language in Mainland China. The participants were asked to complete the Language History Questionnaire (LHQ, version 2.0, Chinese) (Li, Zhang, Tsai, & Puls, 2014) before they participated in the study. According to the results of the questionnaire, all the participants began to learn Cantonese when they could speak and began to learn Mandarin when they attended primary schools. The results of the participants' self-evaluations on the proficiency of Cantonese and Mandarin (on a seven-point scale from 1 (very poor) to 7 (very proficient)) showed that the Cantonese listening proficiency of the participants was significantly higher than their Mandarin listening proficiency (Cantonese: 6.79 ± 0.42, Mandarin: 5.83 ± 0.57, F (1, 46) = 44.90, $p < 0.001$, $\eta_p^2 = 0.49$).

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

2.2. Materials

We adopted eight real Cantonese words and eight pseudo-Cantonese words in the experiment. The real Cantonese words were /si1/ (means "poem"), /se1/(means "some"), /fu1/(means "skin"), /ji1/(means "doctor"), /si6/(means "thing"), /se6/(means "shoot"), /fu6/(means "pay"), /ji6/(means "two"). The pseudo-Cantonese words were/bi1/, /bu1/, /di1/, /du1/, /bi6/, /bu6/, /di6/, /du6/. They consisted of Cantonese vowels and consonants that conform to the rules of Cantonese phonology but do not have corresponding meanings in Cantonese.

We first recorded these words by a male native Cantonese speaker with Cool Edit Pro software (Adobe Systems Incorporated,

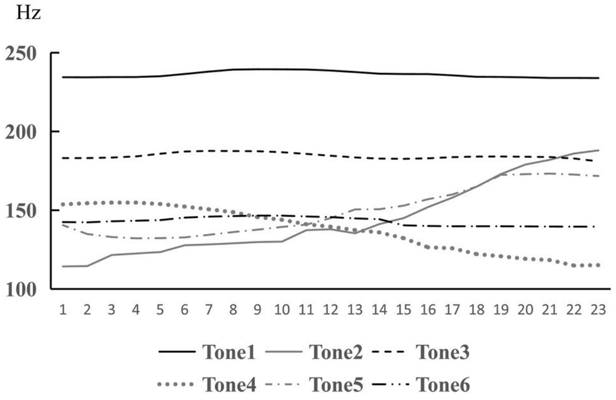

Fig. 1. The fundamental frequency (F0) features of Cantonese lexical tones superimposed on the syllable /si/.

United States), sampling at 44100 Hz. The duration of each word was 400 ms. Then, we re-superimposed Cantonese Tone 1 and Tone 6 (level tones; the F0 features of these two tones are shown in Fig. 1) to these recorded words with Praat software (http://www.fon.hum.uva.nl/praat/). Lastly, these words were normalized to 75 dB with Praat software.

Before the ERP experiment, we recruited another 16 native Cantonese speakers who did not participate in the ERP experiment to determine whether these words have corresponding Cantonese meanings. All the participants reported that the real words have meanings, while the pseudo words have no meaning. These confirmed that the manipulation of lexicality in the experiment was effective.

2.3. Procedure

We used a passive oddball paradigm in the experiment (e.g., Näätänen et al., 2007). The experiment contained 14 conditions, and each condition consisted of one type of standard stimulus and one type of deviant stimulus (as shown in Table 1). There were 28 blocks in total, with each condition being applied twice. The order of these blocks was counterbalanced among participants, and the conditions of any two adjacent blocks were different. Each block consisted of 192 standard stimuli and 40 deviant stimuli. The standard and deviant stimuli were presented pseudo-randomly. Each stimulus was 400 ms, and the interstimulus interval (ISI) was 800 ms.

The auditory stimuli were presented via E-prime 2 software (Psychology Software Tools, https://pstnet.com/products/e-prime/). The participants were instructed to watch a silent movie and ignore the auditory stimuli. They did not need to respond to the auditory stimuli. To ensure that the participants watched the movie attentively, they were asked to answer five questions about the content of the movie after the experiment. The whole experiment lasted about 1.5 h.

2.4. Electroencephalogram (EEG) recording

EEG signals were recorded using a 64-channel (Ag–AgCl) NeuroScan system (NeuroScan, http://www.neuroscan.com/). The electrodes were positioned following the 10–20 system convention. The reference electrode was placed at the tip of the nose, and the ground electrode was placed between the FPz and Fz electrodes. The supra- and infra-orbital bipolar electrodes near the left eye were used to record the vertical electro-oculography (EOG) signals, and the left and the right orbital rim was used to record the bipolar horizontal EOG signals. The impedance of each electrode remained below 5 kΩ. The EEG and EOG signals were digitized online at 500 Hz and band-pass filtered online from 0.05 to 100 Hz.

2.5. Data analysis

Off-line signal processing was carried out using Scan 4.5 (NeuroScan, http://www.neuroscan.com/). The reference electrode was converted to the average signal of the two mastoids (M1 and M2). The interference from the horizontal and vertical eye movements was then automatically detected and corrected. Then, the data were segmented with a 900 ms epoch window, including a 100 ms prestimulus baseline. After baseline correction was performed, any trials with artefact activities beyond the range of −80 to 80 mV were excluded. The trials that were included were then filtered at 1–30 Hz with a finite impulse response filter. The averaged ERPs elicited by the standard and deviant stimuli in each condition were obtained. The MMN of each condition was obtained by subtracting the ERPs of the standard stimuli from that of the deviant stimuli.

Based on the distribution of MMNs and results of previous MMN studies (e.g., Näätänen et al., 2007; Yu et al., 2017), we selected nine electrodes (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4) to further analyze the MMNs. After examining the grand average waveforms, we chose 150–350 ms after the stimuli onset as the time window of the MMNs for the real word conditions and chose 100–300 ms after the stimuli onset as the time window of the MMNs for the pseudoword conditions. We first detected the peak latencies of the MMNs for each electrode selected during the corresponding time window and then calculated the mean amplitudes of the MMNs with a moving

Table 1
The experimental conditions and the standard and deviant stimuli for each condition.

Experimental conditions			
Word types	Deviant types	Standard stimuli	Deviant stimuli
Real words	Tone	/si1/(/si6/)	/si6/(/si1/)
	Vowel	/si1/(/se1/)	/se1/(/si1/)
	Consonant	/si1/(/ji1/)	/ji1/(/si1/)
	Tone + Vowel	/si1/(/se6/)	/se6/(/si1/)
	Tone + Consonant	/si1/(/ji6/)	/ji6/(/si1/)
	Vowel + Consonant	/si1/(/fu1/)	/fu1/(/si1/)
	Tone + Vowel + Consonant	/si1/(/fu6/)	/fu6/(/si1/)
Pseudowords	Tone	/bi1/(/bi6/)	/bi6/(/bi1/)
	Vowel	/bi1/(/bu1/)	/bu1/(/bi1/)
	Consonant	/bi1/(/di1/)	/di1/(/bi1/)
	Tone + Vowel	/bi1/(/bu6/)	/bu6/(/bi1/)
	Tone + Consonant	/bi1/(/di6/)	/di6/(/bi1/)
	Vowel + Consonant	/bi1/(/du1/)	/du1/(/bi1/)
	Tone + Vowel + Consonant	/bi1/(/du6/)	/du6/(/bi1/)

time window ranging from 20 ms before the detected peak to 20 ms after that peak for each electrode. The mean MMN peak latencies and amplitudes of the nine electrodes that were selected were calculated for each condition for the statistical analyses.

2.6. Results

The grand average waveforms of the standard and deviant stimuli in each condition at the Fz electrode site are shown in Fig. 2. The MMN waveforms elicited by the deviant stimuli in each condition at the Fz electrode site are shown in Fig. 3. As shown in Fig. 3, all types of deviant stimuli elicited clear MMNs.

2.6.1. MMN amplitude and peak latency

We first conducted two two-factor repeated-measures ANOVAs with the word type (real words vs. pseudowords) and deviant type (tone vs. vowel vs. consonant) as within-subject factors for the MMN amplitude and peak latency, respectively. Bonferroni correction was performed for all multiple comparisons. Fig. 4 shows the mean MMN amplitudes and peak latencies on the nine selected electrodes for the different types of deviant conditions.

2.6.1.1. MMN amplitude. The ANOVA results showed that the main effect of deviant type was significant ($F(2, 46) = 7.44$, $p = 0.002$, $\eta_p^2 = 0.24$). Post hoc analysis showed that the MMN amplitude of the consonant deviants was significantly smaller than those of the tone deviants ($p = 0.007$) and vowel deviants ($p = 0.03$). However, the MMN amplitude of the tone deviants was not different from that of the vowel deviants ($p = 0.99$). The main effect of word type and the interaction effect between word and deviant types were not significant (F(1, 23) = 0.03, $p = 0.88$, $\eta_p^2 = 0.001$; F(1, 23) = 2.30, $p = 0.11$, $\eta_p^2 = 0.11$).

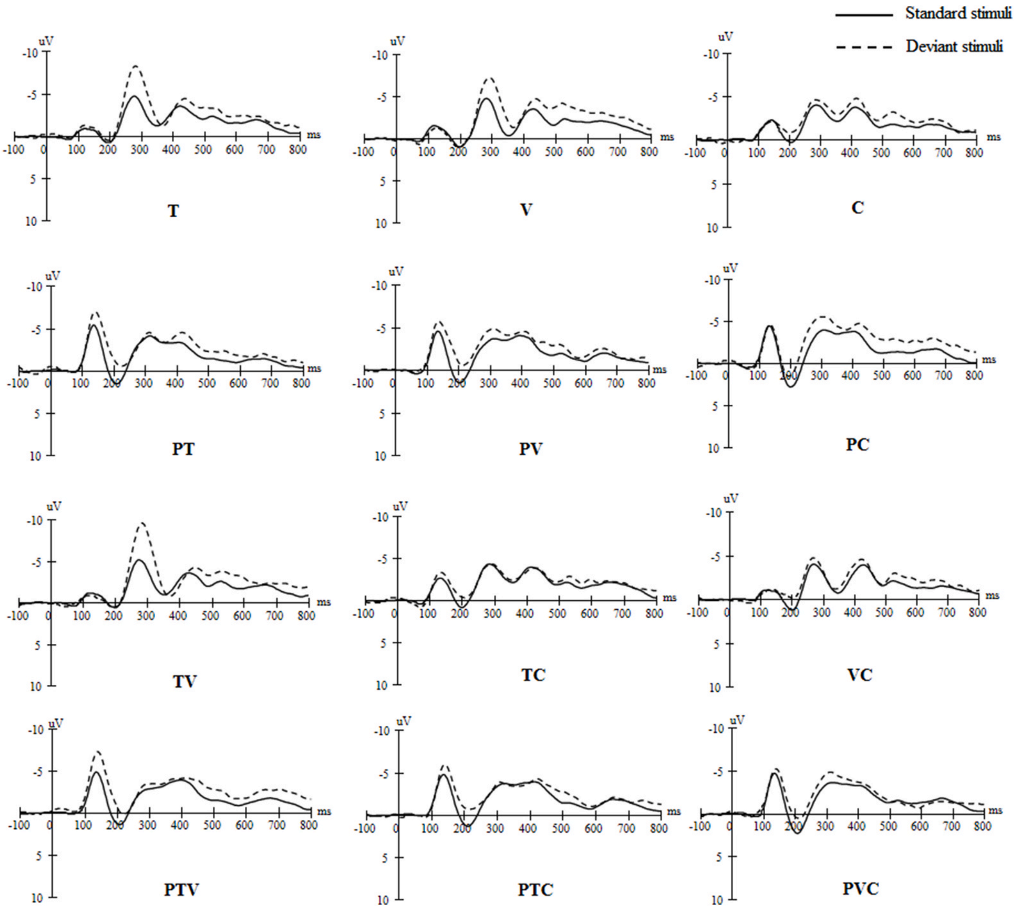

Fig. 2. The grand average waveforms of the standard and deviant stimuli in each condition at the Fz electrode site in Experiment 1 (T: real words with the tone deviant condition; V: real words with the vowel deviant condition; C: real words with the consonant deviant condition; PT: pseudowords with the tone deviant condition; PV: pseudowords with the vowel deviant condition; PC: pseudowords with the consonant deviant condition; TV: real words with the tone + vowel deviant condition; TC: real words with the tone + consonant deviant condition; VC: real words with the vowel + consonant deviant condition; PTV: pseudowords with the tone + vowel deviant condition; PTC: pseudowords with the tone + consonant deviant condition; PVC: pseudowords with the vowel + consonant deviant condition).

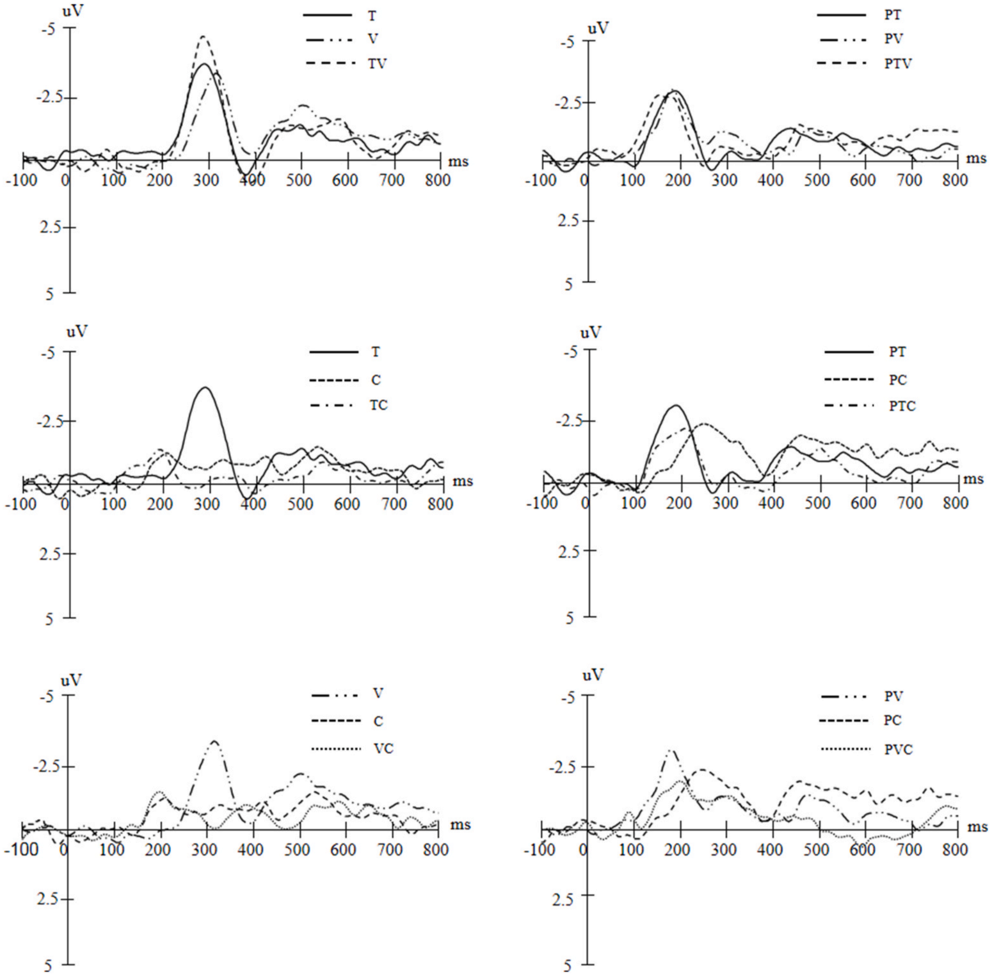

Fig. 3. The MMN waveforms elicited by the deviant stimuli in each condition at the Fz electrode site in Experiment 1 (T: real words with the tone deviants; V: real words with the vowel deviants; C: real words with the consonant deviants; PT: pseudowords with the tone deviants; PV: pseudowords with the vowel deviants; PC: pseudowords with the consonant deviants; TV: real words with the tone + vowel deviants; TC: real words with the tone + consonant deviants; VC: real words with the vowel + consonant deviants; PTV: pseudowords with the tone + vowel deviants; PTC: pseudowords with the tone + consonant deviants; PVC: pseudowords with the vowel + consonant deviants).

2.6.1.2. MMN peak latency. The ANOVA results revealed that the main effect of word type was significant (pseudowords < real words; $F(1, 23) = 191.27$, $p < 0.001$, $\eta_p{}^2 = 0.89$). The main effect of deviant type was also significant ($F(2, 46) = 3.42$, $p = 0.04$, $\eta_p{}^2 = 0.13$). Post hoc analysis showed that the MMN peak latency of tone deviants was significantly earlier than that of the vowel deviants ($p = 0.017$). There were no significant differences between the tone and consonant deviants ($p = 0.236$) or between the vowel and consonant deviants ($p = 0.99$). The interaction effect between word and deviant types was significant ($F(2, 46) = 32.57$, $p < 0.001$, $\eta_p{}^2 = 0.59$). Simple effect analysis further showed that for the real words, the MMN peak latency of the consonant deviants was significantly earlier than those of the tone deviants ($p = 0.04$) and vowel deviants ($p < 0.001$), while the MMN peak latency of the tone deviants was similar to that of the vowel deviants ($p = 0.06$). For the pseudowords, the MMN peak latency of the consonant deviants was significantly longer than those of the tone deviants and vowel deviants ($ps < 0.001$). However, the MMN peak latency of the tone deviants was not different from that of the vowel deviants ($p = 0.99$). In addition, for both the tone deviants and vowel deviants, the MMN peak latency of the real words was significantly longer than that of the pseudowords ($ps < 0.001$). However, for the consonant deviants, the MMN latency of the real words was not different from that of the pseudowords ($p = 0.128$).

2.6.2. MMN additivity analyses

We first calculated the sums of the MMN amplitudes (referred to as added amplitudes) based on the MMN additivity approach by adding the MMN amplitudes of the tone and vowel deviants, those of the tone and consonant deviants, and those of the vowel and consonant deviants. Then, we conducted a three-factor repeated-measures ANOVA on the MMN amplitude with type (tone and vowel (TV) vs. tone and consonant (TC) vs. vowel and consonant (VC)), word (real words vs. pseudowords) and amplitude (original vs. added) as within-subject factors. The original MMN amplitudes were referred to as the MMN amplitudes elicited by tone + vowel

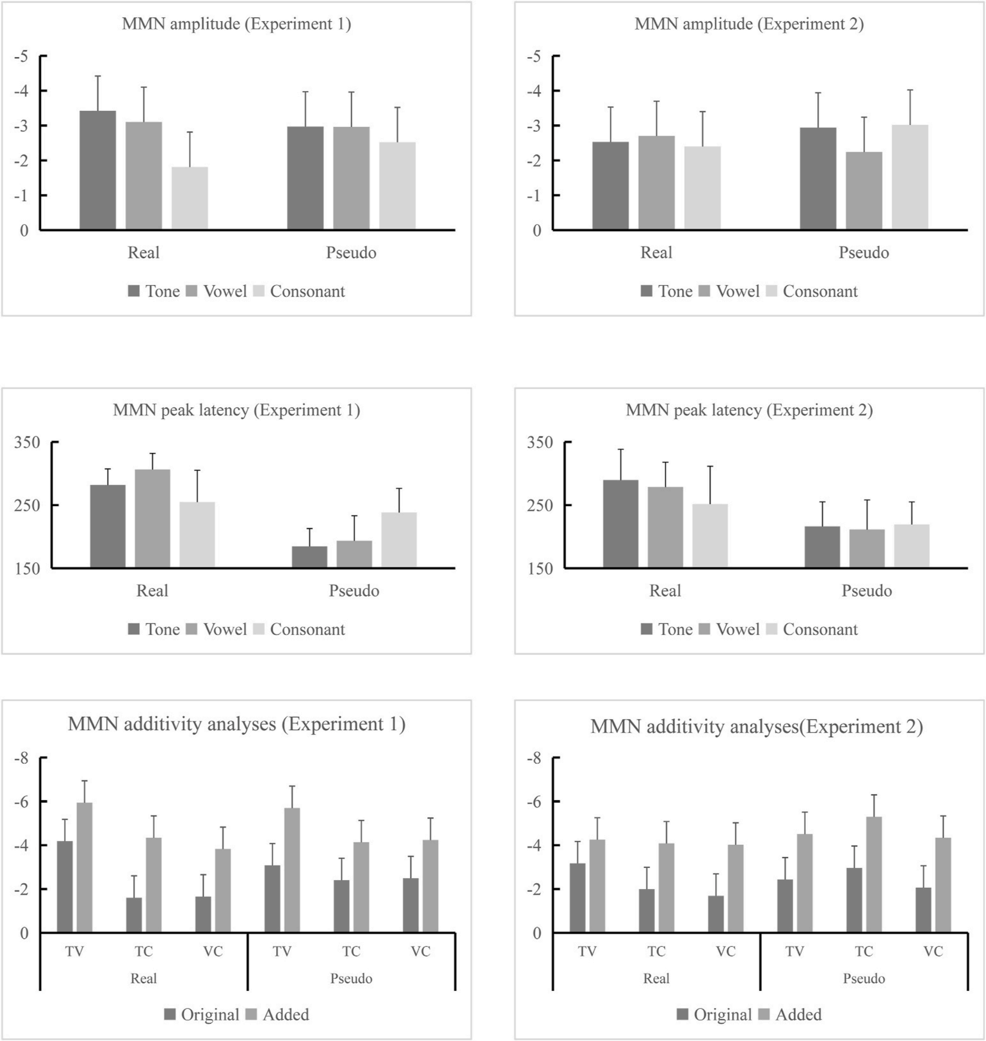

Fig. 4. The mean MMN amplitudes and peak latencies on the nine selected electrodes in Experiments 1 and 2, and the original and added MMN amplitudes of the different conditions in Experiments 1 and 2. The original MMN amplitudes of TV, TC, and VC refer to the MMN amplitudes elicited by the tone + vowel deviants, tone + consonant deviants, and vowel + consonant deviants, respectively. The added MMN amplitudes of TV, TC, and VC correspond to the sums of the MMN amplitudes elicited by the tone and vowel deviants, tone and consonant deviants, and vowel and consonant deviants, respectively (Real: real words; Pseudo: pseudowords; vertical bars represent standard deviants).

deviants, tone + consonant deviants and vowel + consonant deviants. Bonferroni correction was performed for all multiple comparisons. Fig. 4 shows the original and added MMN amplitudes for TV, TC, and VC deviants in real words and pseudowords.

The ANOVA results showed that the main effect of type was significant ($F(2, 46) = 26.18, p < 0.001, {\eta_p}^2 = 0.53$). Post hoc analysis showed that the MMN amplitude of TV was significantly larger than those of TC and VC ($ps < 0.001$). The MMN amplitude of TC was not different from that of VC ($p = 0.99$). The main effect of amplitude was significant (original < added, $F(1, 23) = 20.44, p < 0.001, {\eta_p}^2 = 0.47$). The interaction effect between type and word was significant ($F(2, 46) = 4.97, p = 0.01, {\eta_p}^2 = 0.18$). Simple effect analysis revealed that for the real words, the MMN amplitude of TV was significantly larger than those of TC and VC ($p = 0.001, p < 0.001$). The MMN amplitude of TC was not different from that of VC ($p = 0.99$). However, for the pseudowords, no significant differences were found between any two types (TV vs. TC: $p = 0.14$; TV vs. VC: $p = 0.21$; TC vs. VC: $p = 0.99$). In addition, there were no significant differences between the real words and pseudowords in the TV, TC, and VC conditions ($p = 0.10; p = 0.50; p = 0.10$). The main effect of word and the interaction effects between type and amplitude, between word and amplitude, and among type, word and amplitude were not significant ($F(1, 23) = 0.07$, $p = 0.79$, ${\eta_p}^2 = 0.003$; $F(2, 46) = 0.3, p = 0.74, {\eta_p}^2 = 0.01$; $F(1, 23) = 0.12, p = 0.73, {\eta_p}^2 = 0.005$; $F(2, 46) = 2.13, p = 0.13, {\eta_p}^2 = 0.09$).

2.7. Discussion

We investigated how native speakers process Cantonese level tones, vowels, and consonants with MMN in this experiment. We first examined whether the processing of these phonological features is similar or distinct. The MMN amplitude of level tones was similar to that of vowels, while that of consonants was significantly smaller than that of both level tones and vowels. Regarding MMN peak latency, we found differences between consonants and level tones/vowels, and the differences were mediated by the word type. The MMN peak latency of the consonants was significantly earlier than those of the level tones and vowels in real words but significantly longer than those of the level tones and vowels in pseudowords. However, the MMN peak latency was similar between the level tones and vowels, regardless of whether they were in the real words or pseudowords. Thus, based on the MMN amplitude and peak latency results, the extent and time course of level tone processing are similar to vowel processing. The extent of both level tone and vowel processing was larger than consonant processing. But the time course differences between level tone/vowel and consonant processing were mediated by the lexicality.

We also found that the lexicality affected the MMN peak latency of the level tones and vowels (level tones/vowels: real words > pseudowords), whereas it did not play a role in consonants. The results suggested that the time courses of level tone and vowel processing were influenced by the semantic/phonological information differentiated by the real words and pseudowords. The similar effect of semantic/phonological information also indicated the similar processing between level tones and vowels. The semantic/ phonological information in real words seemed to delay the processing of level tones and vowels, which may just cause the time course differences between level tone/vowel and consonant processing.

Concerning the integration between processing, in the comparisons between original and added MMN amplitudes, the added MMN amplitude was significantly larger than the original MMN amplitude for level tones and vowels, level tones and consonants, and vowels and consonants. According to the MMN additivity approach, the processing of level tones and vowels, level tones and consonants, and vowels and consonants were all integrated. Such integrated processing was not affected by the lexicality and seemed to be stable.

But for the original MMN amplitudes, the word type affected the MMN amplitudes of double-dimensional (level tone + vowel, level tone + consonant, vowel + consonant) deviants. In real words, the level tone + vowel deviants elicited larger MMNs than did the level tone + consonant and vowel + consonant deviants. However, there were no MMN amplitude differences between any two types of double-dimensional deviants in the pseudowords. Nonetheless, we did not find an effect of word type on the MMN amplitudes of single-dimensional deviants. The distinct roles of lexicality between the single vs. double-dimensional deviants may result from the integrated processing of these phonological features in the double-dimensional deviants.

In summary, Experiment 1 revealed similar processing between level tones and vowels, whereas distinct processing between level tones/vowels and consonants. Moreover, we also found the role of lexicality in the time course differences between level tones/vowels and consonants. The processing of level tones and vowels, level tones and consonants, and vowels and consonants was stably integrated. Considering another type of lexical tone, contour tones, we explored how native speakers process contour tones, vowels, and consonants in real vs. pseudo-Cantonese words in Experiment 2. By comparing the results between Experiments 1 and 2, we further examined the potential role of tonal type on processing different phonological features in real and pseudowords.

3. Experiment 2

3.1. Participants

Another group of 24 native Cantonese speakers (11 males, mean age: 20, age range: 18–24) participated in the experiment. They were also undergraduate students at South China Normal University. All the participants could also speak Mandarin. Two participants did not complete the whole experiment because of fatigue, and their data were excluded from analyses. According to the rest of the participants' responses to the Language History Questionnaire (LHQ Version 2.0, Chinese) (Li et al., 2014), the participants began to learn Cantonese from birth and began to learn Mandarin when they attended primary schools. Moreover, their self-evaluations on the proficiency of Cantonese and Mandarin (on a seven-point scale from 1 (very poor) to 7 (very proficient)) showed that their Cantonese listening proficiency was significantly higher than their Mandarin listening proficiency (Cantonese: 7.00 ± 0.01, Mandarin: 6.05 ± 0.49, $F(1, 42) = 84.963$, $p < 0.001$, $\eta_p{}^2 = 0.80$).

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

3.2. Materials

We adopted another eight real Cantonese words and eight pseudo-Cantonese words in Experiment 2. The syllables of these words were the same as that in Experiment 1. But these words were all superimposed on Cantonese contour tones. The real words were /si2/ (means "history"), /se2/(means "write"), /fu2/(means "tiger"), /ji2/(means "chair"), /si4/(means "time"), /se4/(means "snake"), /fu4/(means "symbol"), /ji4/(means "sun"). The pseudowords were /bi2/, /bu2/, /di2/, /du2/, /bi4/, /bu4/, /di4/, /du4/.

Similar to Experiment 1, we first recorded these words by a male native speakers of Cantonese with Cool Edit Pro software (Adobe Systems Incorporated, United States), sampling at 44100 Hz. The duration of each word was also 400 ms. Then, we re-superimposed Cantonese Tone 2 and Tone 4 (contour tones; the F0 features of these two tones are also shown in Fig. 1) to these recorded words with

Praat software (http://www.fon.hum.uva.nl/praat/). Lastly, these words were also normalized to 75 dB with Praat software.

Before the ERP experiment, we asked the same group of 16 native Cantonese speakers who participated in Experiment 1 to judge whether these words have corresponding Cantonese meanings. All the participants reported that the real words have meanings, while the pseudowords have no meaning. These also demonstrated the effectiveness of the lexicality manipulation.

3.3. *Procedure*

We also adopted the passive oddball paradigm in the experiment. The experimental conditions were the same as those used in Experiment 1. But we used contour tones instead of level tones in the experiment. The presentation of deviant and standard stimuli in each condition and the task for the participants were also the same as those used in Experiment 1.

3.4. *EEG recording*

The EEG recording protocol was the same as that used in Experiment 1.

3.5. *Data analyses*

The analyses of the EEG and ERP data were the same as those performed in Experiment 1.

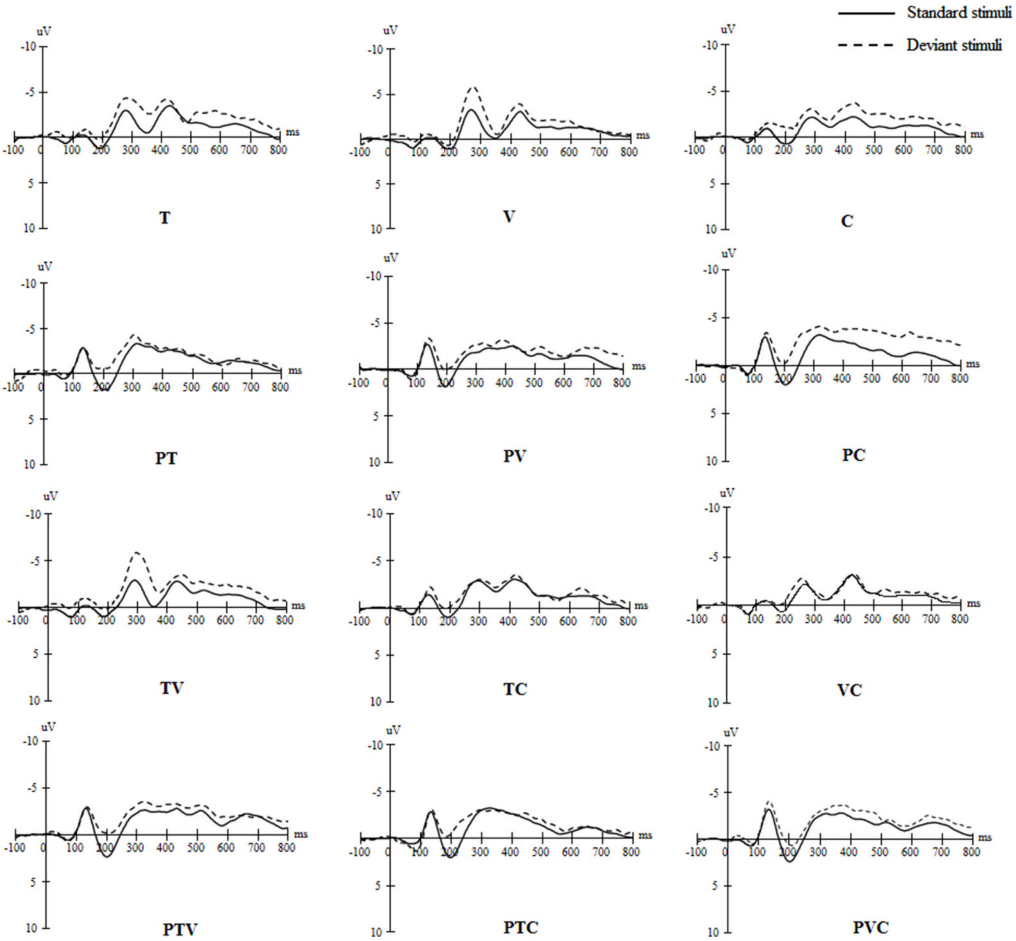

Fig. 5. The original waveforms of the standard and deviant stimuli of each condition at the Fz electrode site in Experiment 2 (T: real words with the tone deviant condition; V: real words with the vowel deviant condition; C: real words with the consonant deviant condition; PT: pseudowords with the tone deviant condition; PV: pseudowords with the vowel deviant condition; PC: pseudowords with the consonant deviant condition; TV: real words with the tone + vowel deviant condition; TC: real words with the tone + consonant deviant condition; VC: real words with the vowel + consonant deviant condition; PTV: pseudowords with the tone + vowel deviant condition; PTC: pseudowords with the tone + consonant deviant condition; PVC: pseudowords with the vowel + consonant deviant condition).

3.6. Results

Fig. 5 shows the grand average waveforms of the standard and deviant stimuli under different conditions at the Fz electrode site. Fig. 6 shows the MMN waveforms elicited by the deviant stimuli under different conditions at the Fz electrode site.

3.6.1. MMN amplitude and peak latency

We conducted similar ANOVAs to the MMN amplitude and peak latency as in Experiment 1. Fig. 4 shows the mean MMN amplitudes and peak latencies for the nine selected electrodes and the different types of deviant conditions.

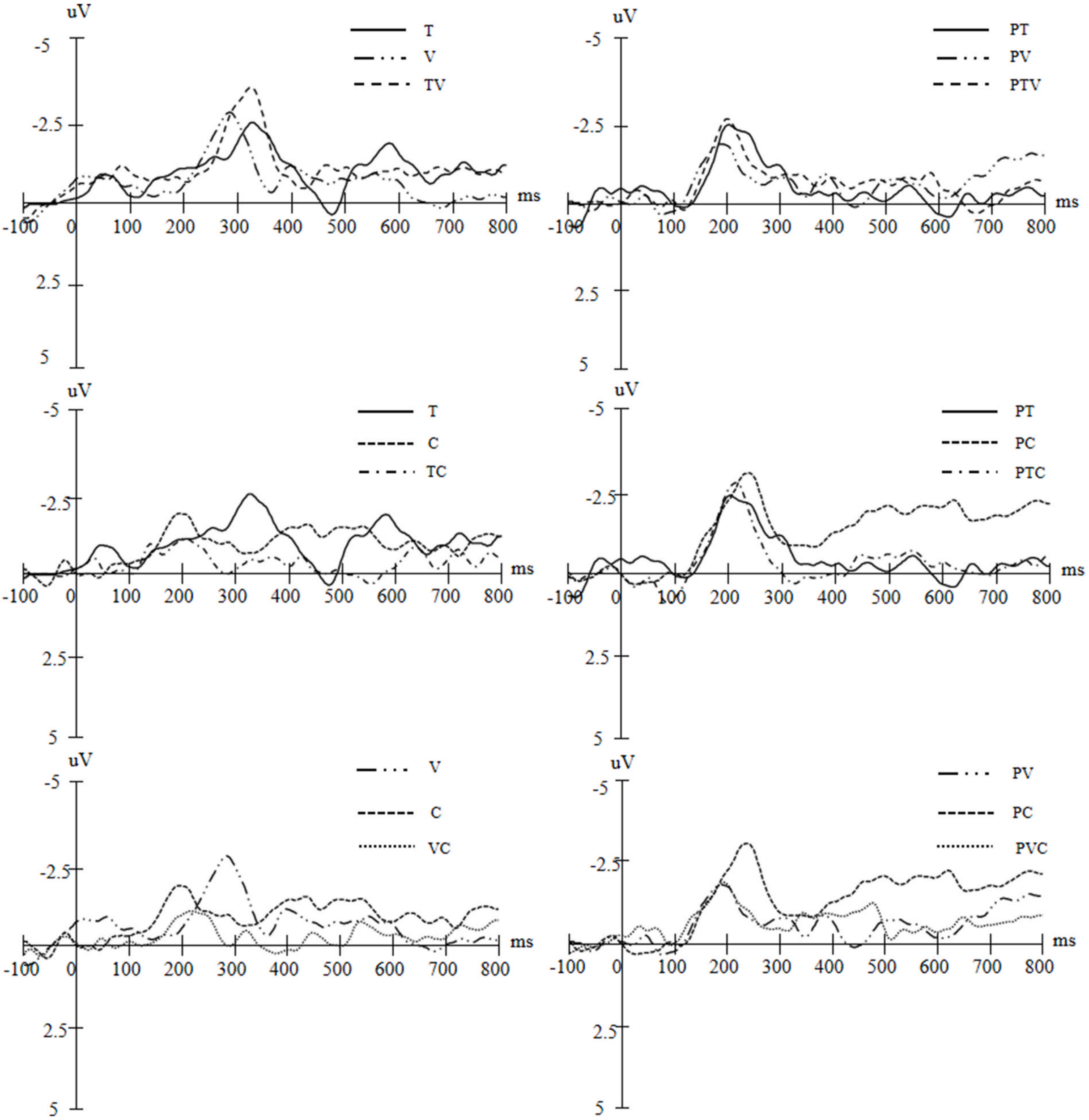

Fig. 6. The MMN waveforms elicited by each condition of deviant stimuli in Experiment 2 (T: real words with the tone deviants; V: real words with the vowel deviants; C: real words with the consonant deviants; PT: pseudowords with the tone deviants; PV: pseudowords with the vowel deviants; PC: pseudowords with the consonant deviants; TV: real words with the tone + vowel deviants; TC: real words with the tone + consonant deviants; VC: real words with the vowel + consonant deviants; PTV: pseudowords with the tone + vowel deviants; PTC: pseudowords with the tone + consonant deviants; PVC: pseudowords with the vowel + consonant deviants).

3.6.1.1. MMN amplitude. The ANOVA results showed that the main effects of word and deviant types and the interaction effect between word and deviant types were not significant (word type: $F(1, 21) = 0.36, p = 0.56, \eta_p{}^2 = 0.02$; deviant type: $F(2, 42) = 0.58, p = 0.57, \eta_p{}^2 = 0.03$; interaction effect: $F(2, 42) = 1.11, p = 0.34, \eta_p{}^2 = 0.05$).

3.6.1.2. MMN peak latency. The ANOVA results revealed that the main effect of word type was significant (pseudowords < real words; $F(1, 21) = 72.90, p < 0.001, \eta_p{}^2 = 0.78$). However, the main effect of deviant type and the interaction effect between word and deviant types were not significant ($F(2, 42) = 1.73, p = 0.19, \eta_p{}^2 = 0.08$; $F(2, 42) = 2.21, p = 0.12, \eta_p{}^2 = 0.10$).

3.6.2. MMN additivity analyses

The analyses were also similar to that in Experiment 1. Fig. 4 shows the original and added MMN amplitudes for each condition.

The ANOVA results showed that the main effect of type was significant ($F(2, 42) = 5.00, p = 0.01, \eta_p{}^2 = 0.19$). Post hoc analysis showed that the MMN amplitude of VC was significantly smaller than those of TV and TC ($p = 0.04$; $p = 0.04$). The MMN amplitude of TV was not different from that of TC ($p = 0.99$). The main effect of amplitude was significant (original < added, $F(1, 21) = 47.39, p < 0.001, \eta_p{}^2 = 0.69$). The interaction effect between type and word was significant ($F(2, 42) = 5.69, p = 0.01, \eta_p{}^2 = 0.21$). Simple effect analysis revealed that for the real words, the MMN amplitude of TV was marginally significantly larger than that of VC ($p = 0.06$). The MMN amplitude of TC was not different from those of TV and VC ($p = 0.19$; $p = 0.99$). However, for the pseudowords, no significant differences were found between any two types (TV vs. TC: $p = 0.77$; TV vs. VC: $p = 0.99$; TC vs. VC: $p = 0.33$). In addition, there were no significant differences between the real words and pseudowords in the TV and VC conditions ($p = 0.46$; $p = 0.39$). However, in the TC condition, the MMN amplitude of the real words was marginally significantly smaller than that of the pseudowords ($p = 0.06$). The main effect of word and the interaction effects between type and amplitude, between word and amplitude, and among type, word and amplitude were not significant ($F(1, 21) = 1.19, p = 0.29, \eta_p{}^2 = 0.05$; $F(2, 42) = 1.83, p = 0.17, \eta_p{}^2 = 0.08$; $F(1, 21) = 0.42, p = 0.53, \eta_p{}^2 = 0.02$; $F(2, 42) = 0.57, p = 0.57, \eta_p{}^2 = 0.03$).

3.7. Analyses of both Experiments 1 (level tones) and 2 (contour tones)

To explore the potential effect of tonal type (level vs. contour tone), we further conducted analyses of both Experiments 1 and 2. Here, we mainly reported the significant results related to tonal type.

3.7.1. MMN amplitude and peak latency

We conducted two three-factor mixed ANOVAs on the MMN amplitude and peak latency, with tonal type (level vs. contour) as a between-subject factor and word (real words vs. pseudowords) and deviant (tone vs. vowel vs. consonant) as within-subject factors. Bonferroni correction was performed for all multiple comparisons.

3.7.1.1. MMN amplitude. The ANOVA results showed that the interaction between tonal type and deviant was significant ($F(2, 88) = 4.72, p = 0.011, \eta_p{}^2 = 0.10$). Simple effect analyses showed that for the words with level tones, the MMN amplitude of the tone deviants was marginally significantly larger than that of the consonant deviants ($p = 0.05$). However, the MMN amplitudes between tone and vowel deviants and between vowel and consonant deviants were not significant ($p = 0.99$; $p = 0.12$). For the words with contour tones, there were no significant differences between any two types of deviants ($ps = 0.99$). The MMN amplitudes of the tone, vowel, and consonant deviants did not differ between the words with level vs. contour tones (tone: $p = 0.29$; vowel: $p = 0.12$; consonant: $p = 0.14$).

3.7.1.2. MMN peak latency. The ANOVA results revealed significant interactions between tonal type and word ($F(1, 44) = 4.33, p = 0.04, \eta_p{}^2 = 0.04$), between tonal type and deviant ($F(2, 88) = 4.09, p = 0.02, \eta_p{}^2 = 0.09$), and among tonal type, word, and deviant ($F(2, 88) = 3.36, p = 0.04, \eta_p{}^2 = 0.07$). The level and contour tones analyses yielded similar results to those of the analyses of MMN peak latencies in Experiments 1 and 2, respectively. Moreover, simple effect analysis revealed that the MMN peak latency for the vowel deviants in the real words with contour tones was significantly earlier than that with level tones ($p = 0.01$). The MMN peak latency of the level tone deviants in the pseudowords was significantly earlier than that of the contour tone deviants in the pseudowords ($p = 0.003$).

3.7.2. MMN additivity analyses

A four-factor mixed ANOVA was conducted to determine the effect of tonal type on the original and added MMN amplitudes, with tone (level vs. contour) as a between-subject factor and type (TV vs. TC vs. VC), word (real words vs. pseudowords), and amplitude (original vs. added) as within-subject factors. Bonferroni correction was performed for all multiple comparisons.

The ANOVA results showed that the interaction between tone and type was significant ($F(2, 88) = 11.80, p < 0.001, \eta_p{}^2 = 0.21$). Simple effect analysis indicated that the MMN amplitude of TV in words with level tones was significantly larger than that with contour tones ($p = 0.03$). For the deviants in words with level tones, the MMN amplitude of TV was significantly larger than those of TC and VC ($p = 0.003$; $p = 0.002$). The MMN amplitude of TC was not different from that of VC ($p = 0.99$). For the deviants in words with contour tones, there were no significant differences between any two types (TV vs. TC: $p = 0.99$; TV vs. VC: $p = 0.39$; TC vs. VC: $p = 0.41$).

3.8. Discussion

Experiment 2 further explored the processing of contour tones, vowels, and consonants by native Cantonese speakers. The MMN amplitudes of contour tone, vowel, and consonant deviants did not differ, regardless of the lexicality. There were no significant differences between them in the MMN peak latencies either. However, the MMN peak latencies for the contour tone, vowel, and consonant deviants in real words were longer than pseudowords. These results revealed the similar extent and time course among contour tone, vowel, and consonant processing, and the lexicality affected the time course of processing.

By comparing the results from Experiments 1 and 2, we further found that the tonal type of lexical tones affected the difference in processing extent between lexical tones and consonants. The extent of level tone processing was larger than that of consonant processing, while the extent of contour tone processing was similar to that of consonant processing. Moreover, the time courses between lexical tone/vowel and consonant processing were affected by the interaction between lexicality and tonal type. In the real words with level tones, the consonant processing occurred earlier than the level tone and vowel processing. In contrast, in the pseudowords with level tones, the consonant processing occurred later than the level tone and vowel processing. However, the consonant processing did not differ from the contour tone and vowel processing with either real words or pseudowords. The two experiments' analyses also revealed the interaction effect on the time course of lexical tone and vowel processing, respectively. The level tones were processed earlier than the contour tones in the pseudowords. The vowels in the real words with contour tones were processed earlier than those in the real words with level tones.

For the integrated processing analyses, regardless of word type, the original MMN amplitudes of the double-dimensional deviants were smaller than the added MMN amplitudes of single-dimensional deviants for the contour tones and vowels, contour tones and consonants, and vowels and consonants. The results suggested that the contour tones and vowels, contour tones and consonants, and vowels and consonants are all integrally processed. The lexicality did not affect the integrated processing. Analyses of Experiments 1 and 2 further revealed that the integrated processing between these phonological features was not influenced by the tonal type either.

Similar to Experiment 1, we also found that the lexicality affected the processing of double-dimensional deviants. The MMN amplitudes of contour tone + vowel deviants were marginally significantly larger than those of vowel + consonant deviants in the real words, while they were similar in the pseudowords. Besides, the contour tone + consonant deviants in the pseudowords were marginally significantly larger than those in the real words. In the analyses of both Experiments 1 and 2, the tonal type also affected the processing between these double-dimensional deviants. The MMN amplitude of the level tone + vowel deviants was larger than those of the level tone + consonant deviants and vowel + consonant deviants in words with level tones. However, there were no differences between any two types of double-dimensional deviants in words with contour tones. For each type of double-dimensional deviant, the tonal type only affected the processing of lexical tone + vowel deviants; that is, the MMN amplitude of level tone + vowel deviants was larger than that of contour tone + vowel deviants. The effect here may be the potential cause of the distinctions between the processing of lexical tone + vowel deviants and lexical tone/vowel + consonant deviants in words with level vs. contour tones.

Therefore, Experiment 2 indicated similar processing among contour tones, vowels, and consonants in the extent and time course and the integrated processing between these phonological features. Analyses of Experiments 1 and 2 further revealed the role of tonal type in the extent difference between lexical tone and consonant processing. The tonal type also played an interaction role with lexicality in processing lexical tones and vowels and the time course differences between lexical tone/vowel and consonant processing.

4. General discussion

We conducted two ERP experiments in the present study to explore the processing of lexical tones (level and contour tones), vowels, and consonants in native speakers' speech perception of Cantonese real and pseudowords via the MMN amplitudes and peak latencies. Overall, results from the two experiments revealed the distinct but integrated processing of lexical tones, vowels, and consonants and the roles of tonal type and lexicality in the processing differences between lexical tones/vowels and consonants. All these findings provided implications to the mechanism underlying tonal language spoken word recognition.

4.1. Similar processing between lexical tones vs. vowels

Whether the processing of lexical tones vs. vowels is similar or distinct remains controversial in previous studies. The present study found that the extent and time course of level/contour tone processing was similar to vowel processing, regardless of real words vs. pseudowords. The results are consistent with those reported by Schirmer et al. (2005), Lee et al. (2012), and Choi et al. (2017). However, in these previous studies, Schirmer et al. (2005) and Lee et al. (2012) mixed the level tones with contour tones, and Choi et al. (2017) used only level tones. We differentiated these two types of lexical tones and demonstrated that the processing of both level and contour tones is similar to that of vowels. Previous studies did not examine the semantic/phonological information's potential role in processing lexical tones and vowels. Our study revealed that the semantic/phonological information would not affect the similarity in processing between lexical tones and vowels. We consider that the similar processing may be that the lexical tones are superimposed on vowels. The acoustic variations in lexical tones are synchronous with those of vowels.

Hu et al. (2012) used Mandarin idioms as the materials and observed differences in the extent and time course of processing between lexical tones and vowels. But in the study by Schirmer et al. (2005), the materials were general Cantonese sentences. Although both of these two studies controlled the last words in the materials, the semantic variations elicited by the last words in the idioms may be larger than those in general sentences, as the words in idioms may be more fixed than those in general sentences (Hu et al., 2012). Therefore, the degree of semantic variation may be one potential reason for the inconsistent findings between Hu et al. (2012) and

Schirmer et al. (2005). Combined with the findings in our study, the semantic information at the word level would not affect the processing of lexical tones and vowels, but it may play a role at a higher level (e.g., sentence level).

4.2. Distinct processing between lexical tones vs. consonants

Consistent with previous studies, we also found a difference between lexical tone and consonant processing (Lee et al., 2012; Tong, Mcbride, et al., 2014). Moreover, our findings revealed that the tonal type modulated their difference. The extent of level tone processing was larger than consonant processing, while the processing of contour tones was similar to consonants. The reason may lie in the F0 variation difference between the level and contour tones. The F0 variation among level tones mainly appears at the onset of a syllable and remains the same during the whole syllable. The acoustic variation among consonants also appears at the onset of a syllable. Therefore, listeners may need more neural resources to detect lexical tone variations in words with level tones. However, the F0 variation among contour tones exists during a whole syllable. Listeners can detect the difference during the whole syllable and do not have to use many neural resources. The results also implied that the features level in the TRACE model consisted of the tonal features, especially pitch height and pitch contour, which influence the processing of different types of lexical tones (level vs. contour tones) in the phonemes & tonemes level (suggested by Tong et al., 2014).

The time course also differed between lexical tone and consonant processing. However, the time course difference was influenced by the interaction between the lexicality and tonal type, mainly in the real words and pseudowords with level tones. The reason may be that the semantic/phonological information did not facilitate but rather interfered with level tone processing. Experiment 1 also showed that listeners could detect the level tones in pseudowords earlier than real words. The F0 onset of a syllable is essential for the detection of level tone differences. The processing of level tones may require more neural resources than consonants. When processing the real words that differ in level tones, listeners would automatically spare neural resources to process the semantic/phonological information. This process may result in insufficient neural resources for the processing of level tones in real words. Thus, semantic/phonological information may interfere with level tone processing. The results also indicated that the words level in the TRACE model might include the pseudowords, and both the features and words levels play roles in processing lexical tones in the phonemes & tonemes level.

4.3. Distinct processing between vowels vs. consonants

Few previous studies discussed the differences in processing between vowels and consonants in tonal languages because these two types of phonological features also exist in non-tonal languages. In the present study, we found that the extent of processing is similar for vowels and consonants. But the time course differences between them were mediated by the interaction between the lexicality and tonal type. This interaction effect was similar to that on the time course between lexical tone and consonant processing. The vowel and consonant processing differences also occurred in the real words and pseudowords with level tones. As lexical tones are superimposed on vowels and the integrated processing of lexical tones and vowels, the reason for this interaction may be the same as that for the interaction effect on lexical tone and consonant processing. Besides, the role of tonal type on the time course difference between the vowel and consonant processing indicated a potential unique mechanism that underlies the processing of vowels and consonants in tonal languages.

4.4. Integrated processing of lexical tones and vowels, lexical tones and consonants, and consonants and vowels

The independent vs. integrated processing between the phonological features is also a fundamental but controversial issue in the literature. Our study revealed that the processing of lexical tones and vowels, lexical tones and consonants, vowels and consonants are integrally processed irrespective of real words vs. pseudowords, words with level vs. contour tones. Choi et al. (2017) also observed integrated level tones and vowels processing. Our study further revealed that the processing of contour tones and vowels is also integrated. For lexical tones and consonants, Gao et al. (2012) suggested that consonants and pitch are processed integrally. However, lexical tones in tonal languages consist of acoustic information like pitch features and phonological information (Xi et al., 2010). Tong, McBride, and Burnham (2014) hypothesized integrated processing of consonants and lexical tones while lacking of direct evidence for the integrated processing. The present study used Cantonese words with consonant and lexical tone variations as the speech materials. Our findings provided direct neurophysiological evidence showing that the consonants and lexical tones processing is integrated. The integration between vowel and consonant processing was consistent with the findings reported by Gao et al. (2019). Token together, our study supported the integration view on the processing of phonological features during tonal language speech perception.

4.5. Implications for the spoken word recognition model

The present study explored how lexical tones, vowels, and consonants were processed in native speakers' processing of Cantonese spoken words. Based on our findings and previous revised TRACE models (especially Tong et al., 2014), we proposed further explanations on the processing mechanism underlying tonal languages' spoken word recognition. Firstly, consistent with Tong, McBride, and Burnham (2014), the features level stored the acoustic features of lexical tones, vowels, and consonants, and the phonemes level (the phonemes & tonemes level in Tong et al., 2014) consisted of vowels, consonants, and lexical tones. For the words level, we considered that it contained real words and pseudowords. Secondly, the extent of level and contour tone processing (the phonemes level) was influenced by the tonal features like pitch height and pitch contour (the features level). In contrast, the time courses of lexical tone, vowel, consonant processing were affected by the tonal features (the features level) and words' lexicality (the words level) simultaneously. Lastly, lexical tones, vowels, and consonants in the phonemes level could be stably integrally processed.

We did not hypothesize units for these phonological features like the lexical tone and vowel combinations (Choi et al., 2017) and atonal syllables (Gao et al., 2019). The phonological units may increase listeners' memory load and constrain lexical tone, vowel, and consonant processing flexibility. With separate representation, listeners could process lexical tones, vowels, and consonants with the bottom-up or top-down approach more efficiently and automatically in different experimental tasks and contexts. Nevertheless, considering the similar extent and time course between lexical tone and vowel processing while different extents and time courses between lexical tone/vowel and consonant processing, the representation of lexical tones and vowels may be closer than that of lexical tones and consonants or vowels and consonants.

4.6. Limitations of the present study

Although the present study provided implications to our understanding of tonal languages' spoken word recognition, some issues still need to be investigated in the future. We only examined the role of tonal type in processing different phonological features by controlling the acoustic features of lexical tones (pitch height vs. pitch contour). It remains unclear how vowels and consonants' acoustic features affect phonological processing. Exploring it could further reveal the influence of the features level and the combined effect of the features and words levels on the phonological processing in the revised TRACE models. Moreover, how experimental tasks and semantic context modulate lexical tones, vowels, and consonants processing should also be investigated. The present study used the passive oddball paradigm and revealed the automatic processing of lexical tones, vowels, and consonants in the context of monosyllabic words. Future studies could adopt tasks (e.g., lexical decision task) that require the participants' attention and disyllabic words, idioms, and sentences as the experimental materials, contributing to indicating a dynamic model of spoken word recognition.

In conclusion, the present study indicated the lexical tone, vowel, and consonant processing in tonal language speech perception. Native speakers processed lexical tones and vowels similarly regarding both extent and time course. The processing of lexical tones and consonants differed in terms of both of these two aspects. Moreover, the extent difference was affected by tonal type, while the time course was affected by the interaction between lexicality and tonal type. The extent of vowel and consonant processing was similar, whereas the time course between them differed, and the interaction between lexicality and tonal type influenced the differences. Although the processing of lexical tones/vowels and consonants differed in terms of extent and/or time course, lexical tones and vowels, lexical tones and consonants, and vowels and consonants were processed integrally, regardless of real words or pseudowords with level or contour tones. These findings provided neurophysiological evidence for the mechanism underlay speech perception and the spoken word recognition models specific to tonal languages.

Funding

This work was supported by the National Social Science Foundation of China (18BYY121).

CRediT authorship contribution statement

Keke Yu: Conceptualization, Writing – original draft, Writing – review & editing. **Yuan Chen:** Conceptualization, Methodology, Investigation. **Menglin Wang:** Investigation. **Ruiming Wang:** Conceptualization, Writing – review & editing. **Li Li:** Conceptualization, Writing – review & editing.

Declaration of competing interest

None.

Acknowledgement

We thank Professor Ping Li for helping with the experimental design and writing. We also thank Jing'an Su for helping to record the speech materials.

References

Caclin, A., Brattico, E., Tervaniemi, M., Naeaetaenen, R., Morlet, D., Giard, M. H., et al. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience, 18*(12), 1959–1972. https://doi.org/10.1162/jocn.2006.18.12.1959

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

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

Choi, W., Tong, X., Gu, F., Tong, X., & Wong, L. (2017). On the early neural perceptual integrality of tones and vowels. *Journal of Neurolinguistics, 41*, 11–23. https://doi.org/10.1016/j.jneuroling.2016.09.003

Deng, Z., Chandrasekaran, B., Wang, S., & Wong, P. C. (2016). Resting-state low-frequency fluctuations reflect individual differences in spoken language learning. *Cortex, 76*, 63–78. https://doi.org/10.1016/j.cortex.2015.11.020

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

Gao, S., Hu, J., Gong, D., Chen, S., Kendrick, K. M., & Yao, D. (2012). Integration of consonant and pitch processing as revealed by the absence of additivity in mismatch negativity. *PLoS One, 7*(5), Article e38289. https://doi.org/10.1371/journal.pone.0038289

Gao, X., Yan, T., Tang, D., Huang, T., Shu, H., Nan, Y., et al. (2019). What makes lexical tone special: A reverse accessing model for tonal speech perception. *Frontiers in Psychology, 10*, 2830. https://doi.org/10.1371/journal.pone.0038289

Gussenhoven, C., & Jacobs, H. (2013). Understanding phonology. In *Understanding phonology* (3rd ed.). Routledge. https://doi.org/10.4324/978020377708.

Hu, J., Gao, S., Ma, W., & Yao, D. (2012). Dissociation of tone and vowel processing in Mandarin idioms. *Psychophysiology, 49*(9), 1179–1190. https://doi.org/10.1111/j.1469-8986.2012.01406.x

Lee, C. Y., Yen, H. L., Yeh, P. W., Lin, W. H., Cheng, Y. Y., Tzeng, Y. L., et al. (2012). Mismatch responses to lexical tone, initial consonant, and vowel in Mandarin-speaking preschoolers. *Neuropsychologia, 50*(14), 3228–3239. https://doi.org/10.1016/j.neuropsychologia.2012.08.025

Lidji, P., Jolicoeur, P., Kolinsky, R., Moreau, P., Connolly, J. F., & Peretz, I. (2010). Early integration of vowel and pitch processing: A mismatch negativity study. *Clinical Neurophysiology, 121*(4), 533–541. https://doi.org/10.1016/j.clinph.2009.12.018

Li, P., Zhang, F., Tsai, E., & Puls, B. (2014). Language History questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17*(3), 673–680. https://doi.org/10.1017/S1366728913000606

Luo, H., Ni, J. T., Li, Z. H., Li, X. O., Zhang, D. R., Zeng, F. G., et al. (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. https://doi.org/10.1073/pnas.0607065104

Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar* (2nd ed.). Routledge.

Mcclelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology, 18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

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. https://doi.org/10.1016/j.clinph.2007.04.026

Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., & Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience, 17*(1), 1–12. https://doi.org/10.1162/0898929052880057

Shuai, L., & Gong, T. (2014). Temporal relation between top-down and bottom-up processing in lexical tone perception. *Frontiers in Behavioral Neuroscience, 79*(8), 97. https://doi.org/10.3389/fnbeh.2014.00097

Shuai, L., & Malins, J. G. (2016). Encoding lexical tones in jTRACE: A simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behavior Research Methods, 49*(1), 230–241. https://doi.org/10.3758/s13428-015-0690-0

Tong, X., McBride, C., & Burnham, D. (2014). Cues for lexical tone perception in children: Acoustic correlates and phonetic context effects. *Journal of Speech, Language, and Hearing Research, 57*, 1589–1605. https://doi.org/10.1044/2014_JSLHR-S-13-0145

Tong, X., Mcbride, C., Lee, C. Y., Zhang, J., Shuai, L., Maurer, U., et al. (2014). Segmental and suprasegmental features in speech perception in Cantonese-speaking second graders: An ERP study. *Psychophysiology, 51*(11), 1158–1168. https://doi.org/10.1111/psyp.12257

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. https://doi.org/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. https://doi.org/10.1016/j.neuropsychologia.2013.07.015

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

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language & Cognitive Processes, 14*(5–6), 609–630. https://doi.org/10.1080/016909699386202

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

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

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. https://doi.org/10.3389/fnhum.2014.00729

Yu, K., Zhou, Y., Li, L., Su, J., Wang, R., & Li, P. (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. https://doi.org/10.1080/23273798.2017.1310909

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences, 363*(1493), 1087–1104. https://doi.org/10.1098/rstb.2007.2161