Language and Cognition (2024), 1–23
doi:10.1017/langcog.2023.75

ARTICLE

The effects of word and beat priming on Mandarin lexical stress recognition: an event-related potential study

Wenjing Yu[1,2], Yu-Fu Chien[3], Bing Wang[4], Jianjun Zhao[5] and Weijun Li[1,2]

[1]Research Center of Brain and Cognitive Neuroscience, Liaoning Normal University, Dalian, China; [2]Key Laboratory of Brain and Cognitive Neuroscience, Liaoning Province, Dalian, China; [3]Department of Chinese Language and Literature, Fudan University, Shanghai, China; [4]School of Music, Liaoning Normal University, Dalian, China; [5]School of Chinese Language and Literature, Liaoning Normal University, Dalian, China
Corresponding authors: Weijun Li and Jianjun Zhao; Emails: liwj@lnnu.edu.cn; zhaojianjun768@163.com

(Received 09 May 2023; Revised 17 December 2023; Accepted 21 December 2023)

Abstract
Music and language are unique communication tools in human society, where stress plays a crucial role. Many studies have examined the recognition of lexical stress in Indo-European languages using beat/rhythm priming, but few studies have examined the cross-domain relationship between musical and linguistic stress in tonal languages. The current study investigates how musical stress and lexical stress influence lexical stress recognition in Mandarin. In the auditory priming experiment, disyllabic Mandarin words with initial or final stress were primed by disyllabic words or beats with either congruent or incongruent stress patterns. Results showed that the incongruent condition elicited larger P2 and the late positive component (LPC) amplitudes than the congruent condition. Moreover, the Strong-Weak primes elicited larger N400 amplitudes than the Weak-Strong primes, and the Weak-Strong primes yielded larger LPC amplitudes than the Strong-Weak primes. The findings reveal the neural correlates of the cross-domain influence between music and language during lexical stress recognition in Mandarin.

Keywords: beat; event-related potential; lexical stress; music and language; priming

1. Introduction

Music and language are complex sound systems and play important roles in our daily lives. The protolanguage hypothesis, proposed even in Darwin's time, states that the two might have the same origin (Darwin, 1871). Evidence from musicology (Feld & Fox, 1994), acoustics (Besson et al., 2011; Patel, 2007, 2011) and cognitive neuroscience (Bidelman et al., 2011; Koelsch et al., 2002; Maess et al., 2001; Patel, 2007; Slevc et al., 2009) suggests that music and language share many similarities. For example, both are hierarchically organized, rule-governed system and unfolding with acoustic events over time (Lerdahl & Jackendoff, 1985; Lutz, 2012; Patel, 2011). In addition, language-specific brain regions, such as Broca’s area and Wernik’s area, are activated when music is being processed (Koelsch et al., 2002; Maess et al., 2001), indicating the cross-domain relationship between music and language.

Stress refers to a prominent syllable or note marked by pitch, intensity, duration, or other enhanced acoustic characteristics (e.g., speech intelligibility) (Breen et al., 2010; Cutler, 1976; Lin et al., 1984; Xu, 1956). Stress carries important information in both music and language (Plack et al., 2005), and its regularity makes auditory signals highly predictable (Ellis & Jones, 2010; Jones et al., 2002). In music, the stress on different scales (i.e., beats) repeats cyclically in a certain order to organize music, and the beats largely determine the sense of harmony of tonal music (Cooper & Meyer, 1971; Krumhansl, 1991). In language, stress is often used by listeners for word recognition (Ye & Connine, 1999). Stressed (i.e., strong) and unstressed (i.e., weak) syllables of words help listeners form the rhythm of the language, making the upcoming information predictable to a certain extent (Pitt & Samuel, 1990). The current study employed event-related potentials (ERPs) to examine the cross-domain relationship between music and language in an auditory priming experiment. Specifically, we utilized Mandarin Chinese as a test case to investigate how lexical stress and musical stress influence the cortical processing of the upcoming lexical stress.

Mandarin Chinese, henceforth Mandarin, is a tonal language, which uses four types of pitch variations to distinguish word meanings (Kratochvil & Chao, 1970; Yip, 2002). A famous example is that the syllable “ma” means mother, hemp, horse, and scold when carrying T1 (level tone), T2 (rising tone), T3 (falling-rising tone) and T4 (falling tone), respectively. In addition, Mandarin has a neutral tone that signals an unstressed syllable and always occurs in the final syllable of a word. The syllable carrying a neutral tone is shorter than that carrying one of the four citation tones (Chen & Xu, 2006; Lee & Zee, 2010; Lin, 1985). For example, the word [’tʊ̄ŋ ’ɕīī] (meaning west and east) and the word [’tʊ̄ŋ ɕi] (meaning stuff) contrast in the stress pattern, with the unstressed second syllable of [’tʊ̄ŋ ɕi] (i.e., stuff) having a shorter duration and a mid-falling contour (Chao, 1968; Duanmu, 2007).

In addition to the words with neutral tone, there are unstressed and stressed syllables in Mandarin disyllabic words. Although unstressed and stressed syllables in Mandarin can be judged based on the listener’s subjective perception of phonological prominence, the judgment is closely related to acoustic parameters such as pitch, intensity and duration (Cao, 1986; Lin, 1962; Lin et al., 1984). Compared to unstressed syllables, stressed syllables have a longer duration and a wider F0 range (Lu, 1984; Sun, 1999). Mandarin disyllabic words have two stress patterns, namely, the Strong-Weak and the Weak-Strong pattern (Lu, 1984; Yin, 1982). The first syllable of the Strong-Weak pattern is more prominent in terms of intensity, duration, pitch and energy, while the second syllable of the Weak-Strong pattern is more prominent in terms of the same acoustic parameters (Xu, 1956). In recent years, research on the lexical stress patterns of Mandarin has shown contradictory results. Some researchers believe that both Strong-Weak and Weak-Strong patterns are equally common in Mandarin, while others argued that the Strong-Weak pattern is predominant (Wang & Feng, 2006). There are also researchers considering that the Weak-Strong pattern is the principal pattern (Lin et al., 1984; Xu, 1982), although it is argued that there is no fixed stress pattern in Mandarin (Liu, 2007; Zhou, 2018). The controversy in the literature may arise from the fact that Mandarin is a tonal language. Therefore, lexical stress in Mandarin has receded to a relatively minor function (Yin, 2021). Therefore, it is often difficult to distinguish the stressed and unstressed syllables of words when no syllable bears a neutral tone. The primary function of Mandarin lexical stress is to distinguish linguistic symbols and semantic information (Lu, 1984; Yin, 1982). Mandarin lexical stress also has a metrical rhythmic function, as most syllables are tonalized, creating a syllable-timed rhythmic feature with syllables as units (Luo & Wang, 2002).

As for the role of stress in rhythm processing, Jones and Boltz (1989) proposed the Dynamic Attentional Theory (DAT). According to the theory, internal oscillators would synchronize with external rhythms, and such coupling would prevent listeners from evenly distributing attentional resources to every part of the auditory signals. Instead, attention is selectively focused on the important information to facilitate auditory signal processing and prediction of upcoming information (Jones & Boltz, 1989; Koelsch et al., 2002). DAT is equally applicable in music (Large et al., 2015; Large & Palmer, 2002) and language (Kotz & Schwartze, 2010; Pitt & Samuel, 1990). For example, when hearing a triple meter sequence with the accented pulse falling on the last beat (12**3**, 12**3**, 12**3**), listeners would predict that a "12**3**" (weak-weak-strong) rhythm will occur next (Mirka, 2004); when hearing a number of trisyllabic words with final stress, listeners would expect the upcoming words to have the same stress pattern (Pitt & Samuel, 1990). In short, prediction is the main factor that makes listeners pay attention to a particular part of the speech signal. These findings are also consistent with the predictive coding theory and the theory of expectancy-driven speech processing. Predictive coding theory states that the brain does not passively receive bottom-up inputs, but uses predictive coding to process hierarchical information. During information processing, comprehenders would receive both bottom-up input and top-down prediction. Predictive errors would be generated when the predicted results are inconsistent with the actual input. In this case, the error signal would be transmitted upward and adjust the following prediction. The adjusted prediction would then be transmitted downward to generate expectations at a lower level, suppressing the prediction error (Garrido et al., 2007, 2009; Parmentier et al., 2011; Todorović et al., 2011). Therefore, processing unexpected information consumes more cognitive resources (Friston, 2005; Rubin et al., 2016). The theory of expectancy-driven speech processing also suggests that the temporal elements of auditory input and the subcortical speech perception system in neural networks work together to optimize the prediction of speech acts. The neural network can coordinate speech perception and production in a timely and accurate manner. The auditory rhythm-driven prediction may help enhance the perception of stressed syllables, thus facilitating language processing at these time points (Kotz & Schwartze, 2010, 2015).

Auditory rhythm, such as speech rhythm or music with a significant metrical structure, are characterized by stress patterns of syllables and notes; these syllables and notes are repeated (quasi-) periodically and are considered to be highly regular (Lehiste, 1977; London, 2004). Repeated and regular stress makes the rhythm predictable and enhances listeners' attention (Ellis & Jones, 2010; Jones et al., 2002). Previous studies using the fragment priming paradigm have shown that listeners responded more quickly when the stress patterns of prime fragments and that of target words were matched (e.g., mu-muSEUM) than when they were mismatched (e.g., mu-MUsic) (Cooper et al., 2002; Cutler & Donselaar, 2001; Soto-Faraco et al., 2001). At the cortical level, previous studies have found larger

P350 amplitudes when the stress patterns of prime fragments and target words were mismatched. Given that the P350 component is associated with facilitating word recognition, the observed P350 effect suggests that stress is used in spoken language recognition (Friedrich et al., 2004). In addition to syllable priming experiments, the word priming paradigm has also been employed to investigate the role of stress in word recognition. For example, Böcker et al. (1999) examined the effect of word priming on stress recognition in Dutch by manipulating the congruency of the stress pattern between the prime word sequences and target words. Their behavioral results showed shorter reaction times and higher accuracy rates in the congruent stress condition. For the ERP results, greater LPC amplitudes were elicited in the incongruent stress condition relative to those yielded in the congruent stress condition. In addition, the participants' performance was modulated by the stress pattern. The Weak-Strong primes elicited greater P2, N325, and N400 amplitudes than the Strong-Weak primes (Böcker et al., 1999). However, this study did not control the semantic relationship between the primes and targets. Therefore, it is possible that the N400 effect was due to their semantic relatedness rather than due to their stress patterns. Based on Böcker et al. (1999), Magne et al. (2016) controlled the semantic relationship between the primes and targets. They observed larger negative ERP components in the 288–567 ms and 398–594 ms time windows when the Strong-Weak primes and Weak-Strong primes had incongruent stress patterns with their corresponding targets, respectively. These ERP components were still argued to be N400 in this study even though the semantic relatedness between the primes and targets was eliminated, since previous studies have shown that the N400 effect can also be elicited by disharmony or unexpected stress in rhythm/prosody (Bohn et al., 2013; Domahs et al., 2008; Magne et al., 2007; Marie et al., 2011; McCauley et al., 2013; Rothermich et al., 2012; Schmidt-Kassow & Kotz, 2009).

Given a plethora of similarities between music and language, a large amount of research has been conducted to investigate the interactions between these two domains. For example, some studies have shown that long-term musical training has a positive effect on listeners' speech recognition, memory, and segmentation (Elmer et al., 2014, 2015; François et al., 2013; François & Schön, 2014). Others have shown that music can shape the coding of language in a relatively short period of time (Cason et al., 2015a, 2015b; Cason & Schön, 2012; Kotz & Gunter, 2015). Of particular interest to the current study is Cason and Schön (2012), in which participants were first presented with a few musical beats as primes, then with pseudowords as targets, while their behavioral (reaction times) and EEG data were recorded. Results showed that participants' reaction times to the targets were facilitated when the beat primes and pseudoword targets shared the same stress pattern. When they had incongruent stress patterns, N100 and P300 effects were observed. These results suggest that subsequent speech recognition can be enhanced if listeners have been exposed to the musical beats with the same stress pattern as the following words or short sentences (Cason et al., 2015a; Cason & Schön, 2012).

Fotidzis et al. (2018) conducted a cross-modal priming experiment to examine whether auditory beat primes would affect visual word processing. Results showed that target words with inconsistent stress patterns with the beat primes elicited a greater negative ERP component with a fronto-central distribution (Fotidzis et al., 2018). Consistent with Fotidzis et al. (2018), Hilton and Goldwater (2020) obtained similar results using sentences as the target stimuli. These findings support the idea that music and language interact in stress processing, and the effect can be cross-modal.

As shown above, previous studies on lexical stress recognition mainly used words and beats as primes and found that the stress pattern of both types of primes influenced the stress recognition of word/pseudoword targets similarly (Böcker et al., 1999; Cason & Schön, 2012, Cason et al., 2015a, 2015b; Fotidzis et al., 2018; Magne et al., 2016). For example, Böcker et al. (1999) and Magne et al. (2016) showed ERP components indicating processing difficulty at the cortical level when word prime sequences and word targets shared different stress patterns relative to when they shared the same stress pattern. Cason and Schön (2012, Cason et al. (2015a, 2015b), and Fotidzis et al. (2018) demonstrated higher demand for cognitive resources when the stress patterns of beat primes and pseudoword/word targets were incongruent. Although it is relatively certain that word primes with the same stress pattern as word targets facilitate the recognition of target lexical stress (Böcker et al., 1999; Magne et al., 2016), the facilitative effect of beat priming on lexical stress recognition is still unclear for tone language speakers (Bidelman et al., 2011, 2013; Peretz et al., 2011; Zheng & Samuel, 2018). The majority of studies examining the effect of musical stress on lexical stress recognition have been conducted in non-tonal languages such as French, German and English. Few studies investigated these issues in tonal languages, such as Mandarin, Cantonese and Thai. Given that some studies have demonstrated that non-musicians of tonal languages and musicians of non-tonal languages have similar abilities in processing stress (Bidelman et al., 2011, 2013; Hutka et al., 2015; Tong et al., 2018), while others revealed that listeners of tonal languages have more difficulty in distinguishing the contours of musical stress (Bent et al., 2006; Chang et al., 2016; Peretz et al., 2011; Zheng & Samuel, 2018), it is warranted to further investigate the effect of musical stress and lexical stress on the processing of lexical stress in tonal languages, such as Mandarin.

Taken together, the present study used ERPs to investigate the effects of word priming and beat priming on the lexical stress recognition of Mandarin disyllabic words. At the behavioral level, we predict that the recognition of the stress pattern of disyllabic targets would be facilitated if beat and word prime sequences have the same stress pattern as the target words (Cason et al., 2015a; Cason & Schön, 2012; Fotidzis et al., 2018; Hilton & Goldwater, 2020). We also predict that shorter reaction times and higher accuracy rates would be observed if the stress pattern of the prime sequences is congruent with that of the target words (Cooper et al., 2002; Cutler & Donselaar, 2001; Soto-Faraco et al., 2001). At the cortical level, we predict that greater N400 and LPC amplitudes would be elicited if the stress pattern of the prime sequences is incongruent with that of the target words (Böcker et al., 1999; Fotidzis et al., 2018; Magne et al., 2016). In addition, given that Brochard et al. (2003) and Potter et al. (2009) showed greater P300 amplitudes for the Weak-Strong pattern than for the Strong-Weak pattern (Brochard et al., 2003; Potter et al., 2009), we therefore predict that Weak-Strong prime sequences would induce larger P300 amplitudes than Strong-Weak prime sequences for beat priming (Brochard et al., 2003; Potter et al., 2009); for word priming, Weak-Strong prime sequences may induce larger P2 and N400 amplitudes, as well as stronger late positive components (P600 or LPC) than Strong-Weak prime sequences (Böcker et al., 1999; Breen et al., 2019; Marie et al., 2011; McCauley et al., 2013).

2. Methods

2.1. Participants

Before participant recruitment, G∗Power 3.1 software (Faul et al., 2009) was used to calculate the required number of participants. At the significance level $\alpha = 0.05$ and the medium effect ($f = 0.25$), the total sample size required to reach the 80% statistical power level was predicted to be at least 24 participants. We randomly recruited 27 non-music major students from Liaoning Normal University to participate in the experiment. Among them, the data of 3 participants were excluded due to excessive artifacts or incorrect responses (the error rate of an experimental condition was greater than 35%). The remaining data of the 24 participants were subject to subsequent analyses (10 male students, ranging in age from 18 to 27, with an average age of 20.52 ± 2.77 years). According to their self-reports, all the participants were right-handed and native speakers of Mandarin Chinese. They all had normal vision or corrected-to-normal vision, no hearing impairment, dyslexia, or any neurological disorder. The study was approved by the Ethics Committee of Liaoning Normal University. The participants signed an informed consent form before the experiment and were paid after the experiment.

2.2. Materials

The priming materials include word primes and beat primes. A total of 84 word primes were selected from the *Mandarin Concise Light and Heavy Format Dictionary* (Song, 2009), half of which had the Strong-Weak stress pattern. The other half of them had the Weak-Strong stress pattern. In order to reduce the influence of tones on stress perception, we balanced the number of times each tone occurred at the initial syllable or final syllable positions, resulting in 16 combinations (T1 + T1, T1 + T2, T1 + T3, T1 + T4, T2 + T2, T2 + T3, T2 + T4, T3 + T1, T3 + T3, T3 + T4, T4 + T1, T4 + T2, T4 + T3, T4 + T1, T4 + T2, T4 + T3, T4 + T4, see Table 1). The beat primes were created by combining 7 notes (‘do’, ‘re’, ‘mi’, ‘fa’, ‘sol’, ‘la’ and ‘si’) into all possible pair combinations with both Strong-Weak and Weak-Strong patterns, resulting in 84 tokens. The 40 pairs of disyllabic target words were also selected from the *Mandarin Concise Light and Heavy Format Dictionary*. They were all inverse morpheme words (for example, 中期 ‘middle stage’ and 期中 ‘midterm’, with the bold characters representing stressed syllables). In addition, another 40 words with initial stress and 40 words with final stress were selected as filler targets (e.g., 西瓜 ‘watermelon’, 血脉 ‘blood relationship’).

Table 1. Examples of experimental materials

Condition	Word1/Beat1	Word2/Beat2	Word3/Beat3	Target word
Strong-Weak, congruent	饥荒	亲情	乐器	中期
	do-la	**mi**-sol	**fa**-si	中期
Weak-Strong, congruent	地基	退回	心虚	期中
	do-**la**	mi-**sol**	fa-**si**	期中
Strong-Weak, incongruent	饥荒	亲情	乐器	波长
	do-la	**mi**-sol	**fa**-si	波长
Weak-Strong, incongruent	地基	退回	心虚	长波
	do-**la**	mi-**sol**	fa-**si**	长波

Note: Bolded words or beats indicate stressed conditions.

Each word/beat prime was presented at a different position of a prime sequence of three words using the Latin Square design, producing a total of 160 word-prime sequences and 160 beat-prime sequences. Within a sequence, the three word/beat primes had the same stress pattern, with either initial stress or final stress. The congruency of the stress pattern between the prime sequence and the target was also manipulated, with half of the targets having a congruent stress pattern with the prime sequences and half of the targets having an incongruent stress pattern with the prime sequences. To prevent the participants from hearing the same target words in both congruent and incongruent conditions, all the experimental materials were divided into two lists, with each participant receiving only one of the lists. There was no semantic relatedness between the prime words and target words. All word materials were recorded in a soundproof studio at 44.1 kHz sampling rate and 16-bit resolution by a phonetically trained female native speaker of Mandarin. The musical stimuli were sung by a female musician and recorded at 44.1 kHz sampling rate and 16-bit resolution (see Figure 1 for the examples). All the recorded materials were saved as. wav files. The intensity of the materials was normalized to 70 dB using Praat (http://www.fon.hum.uva.nl/praat/).

To ensure the validity of the recorded experimental materials, 20 participants (10 males, 19–24 years old, mean age = 22.2) who did not participate in the priming experiment were recruited to evaluate the materials. The auditory stimuli were presented using the E-prime 2.0 software, and the participants were asked to judge the stress pattern of the materials (initial stress or final stress). Whenever there was a stimulus that did not have a 90%-or-higher accuracy rate, this particular stimulus was re-recorded and rated until it met the standard. To ensure that all the word stimuli were familiar to the participants, a familiarity rating experiment was conducted in which the participants were asked to score the familiarity of words on a 7-point scale (1 being very unfamiliar and 7 being very familiar). The rating results showed that the participants were pretty familiar with the word stimuli (M = 5.09 ± 1.46). Finally, acoustic analyses were also conducted on the prime words, target words, and prime beats to evaluate the acoustic features of the stressed and unstressed syllables, and

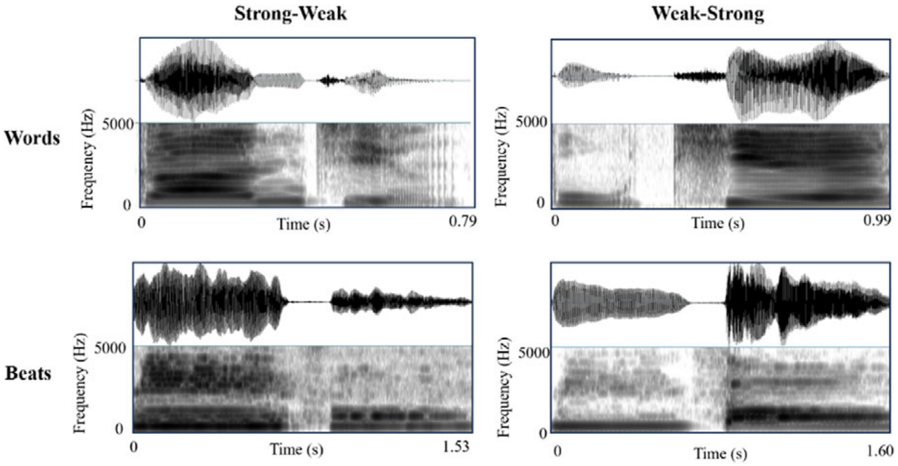

Figure 1. Oscillograms and spectrograms of the stimulus sample.

those of the strong and weak beats. Duration, F0 range, and intensity data were analyzed using paired sample *t*-tests (see Table 2). Results showed that the stressed syllables/strong beats had longer duration, wider F0 range, and greater intensity than the unstressed syllables/weak beats. These results are consistent with the acoustic characteristics observed for stressed/unstressed syllables and strong/weak beats in previous studies (Chen & Gussenhoven, 2008; Li et al., 2018; Sun et al., 2022).

2.3. Procedure

The auditory–auditory priming experiment was conducted in a dim-lit, cozy and soundproof room. The participants were seated in a comfortable chair in front of the LCD screen (23 inches, 60 Hz refresh rate), with headphones in both ears and the volume adjusted to suit individuals' hearing needs. In each trial, the fixation cross (+) was first presented for 500 ms. Then the fixation cross stayed in the center of the screen to reduce the eye movement of the participants while three-word primes/beat primes and a target word were being presented auditorily one after another. The participants were required to pay attention to the stress pattern of the target word. The time interval between the stimuli was 600 ms, which is considered the most spontaneous and natural interval without bias (Fraisse, 1982; Krumhansl, 2000). The participants' task was to determine whether or not the stress pattern of the prime sequence and that of the target word were congruent. Half of the participants would need to press "F" for the congruent condition and "J" for the incongruent condition. The other half of the participants would need to press "F" for the incongruent condition and "J" for the congruent condition on the keyboard (see Figure 2 for the experimental procedure). The whole experiment consisted of 4 blocks of word priming and 4 blocks of beat priming. Each block comprised 40 trials, with each experimental condition containing 10 trials (see Table 1). To minimize the possibility that prior exposure to music may enhance cognitive processing (Ladányi et al., 2021) and prevent such effects from influencing our results, the orders of the word priming and beat priming were counterbalanced across participants. Before the main experiment, 8 practice trials were provided to familiarize the participants with the procedure.

We followed up with a stress type judgment task and a word recall test. In the stress type judgment task, we selected a total of 50 words from the formal experiment, of which 25 were Strong-Weak words and 25 were Weak-Strong words. In the word recall test, 10 words from the stress type judgment task (5 Strong-Weak words, 5 Weak-Strong words) were selected, and another 10 words from the formal

Table 2. Acoustic parameters of words and beats

	Words			Beats		
	Stressed *M*(*SD*)	Unstressed *M*(*SD*)	*t* (1,327)	Strong *M*(*SD*)	Weak *M*(*SD*)	*t* (1,83)
Duration (ms)	595 (126)	370 (83)	22.43***	909 (199)	698 (143)	7.99***
Range (Hz)	117 (102)	89 (86)	5.04***	27 (33)	19 (25)	1.58
Intensity (dB)	69 (4)	64 (4)	12.69***	981 (1)	75 (3)	20.68***

*** $p < 0.001$.

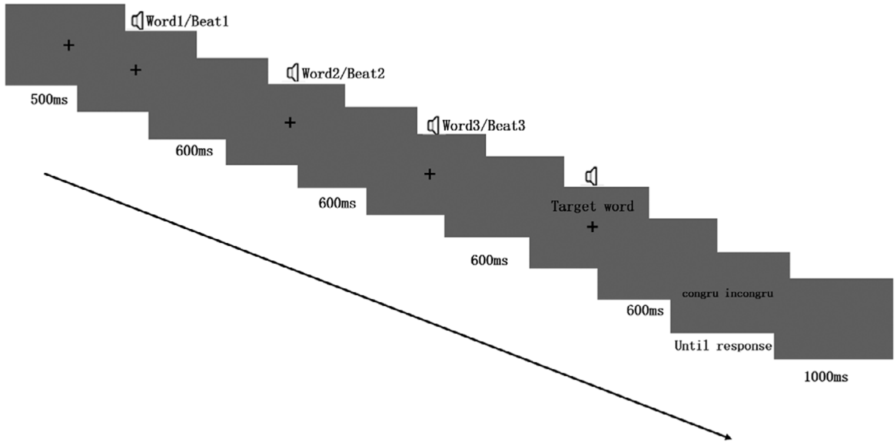

Figure 2. Schematic illustration of the trial schemes.

experiment (5 Strong-Weak words, 5 Weak-Strong words) were selected as new vocabulary. We recruited 20 additional subjects (9 males, mean age 21.85 ± 2.21 years) to conduct a stress type judgment task and a word recall test in order. In the stress type judgment task, a fixation cross (+) was first presented for 500 ms, followed by an auditory presentation of a word that required subjects to focus only on the stress type of the word. Then they were asked to determine whether the word belonged to the Strong-Weak word or Weak-Strong word by pressing the key on a keyboard. Half of the participants would need to press "F" for the Strong-Weak word and "J" for the Weak-Strong word. The other half of the participants would need to press "F" for the Weak-Strong word and "J" for the Strong-Weak word. In the word recall test, a 500-ms fixation cross was presented, followed by an auditory presentation of a word, in which subjects were asked to determine whether the word they had just heard appeared in the stress type judgment task. They had to press the "F" key if the word appeared, or the "J" key if it did not. The other half of the subjects made the opposite keystroke.

2.4. EEG recording and data processing

Continuous EEG data were recorded using Brain Products actiCHamp in accordance with a 64 Ag/AgCl electrodes cap modified by the 10–20 international system. The signal was recorded at a sampling rate of 1000 Hz and FCz was used as an online reference. The electrodes TP9 and TP10 were placed on the left and right mastoid processes, respectively, and the impedance between the electrodes and the scalp was less than 5 kΩ.

The data were preprocessed using EEGLAB v.13.5.4b (MathWorks, Natick, USA). Off-line filtering was performed by high-pass filtering at 0.01 Hz and low-pass filtering at 30 Hz, and the mean of bilateral mastoid processes was subtracted from the EEG data of each lead as the re-reference. Artifact correction was performed using independent component analysis in EEGLAB (ICA). The onset of EEG analysis lock was at the moment the target words were presented, intercepting data 200 ms before and 800 ms after the target word presentation, and the baseline correction time was

200 ms to 0 ms before the target stimulus presentation. After excluding the trials whose amplitudes were greater than ±80 µV, the effective trials under each condition were more than 25. Finally, the ERPs of the remaining trials within each condition were averaged.

By visual inspection of the average amplitudes and previous studies on the P2 (Böcker et al., 1999; Marie et al., 2011), N400 (Breen et al., 2019; Marie et al., 2011; Rothermich et al., 2012) and LPC (Böcker et al., 1999; McCauley et al., 2013) components, we selected the electrodes and time windows corresponding to each component. For the P2 component, the selected time window was 50–150 ms, and the ROIs were FC1, FCz, FC2, F1, Fz, F2, C1, Cz, C2; for the N400 component, the selected time window was 300–450 ms, and the ROIs were P1, Pz, P2, CP1, CPz, CP2; for the LPC component, the selected time window was 500–700 ms, and the ROIs were CP1, CPz, CP2, P1, Pz, P2, PO3, POz, PO4.

3. Results

3.1. Behavioral results

A three-way repeated measures ANOVA was conducted on participants' accuracy rates with Prime Type (word priming vs. beat priming), Stress Pattern (Strong-Weak vs. Weak-Strong), and Congruency (congruent vs. incongruent stress patterns between primes and targets) as independent variables. A significant main effect of Congruency was found, $F(1, 23) = 16.158$, $p = 0.001$, $\eta^2 = 0.413$, with the accuracy rates yielded in the congruent condition ($M = 92.444$, 95%CI [90.924, 93.964]) being significantly higher than those elicited in the incongruent condition ($M = 89.179$, 95%CI [86.899, 91.459]). A significant interaction between Stress Pattern and Prime Type was also observed, $F(1, 23) = 8.524$, $p = 0.008$, $\eta^2 = 0.270$. Subsequent simple effects analysis showed that the accuracy rates of word primes were significantly higher than those of beat primes for the Strong-Weak condition, $F(1, 23) = 3.968$, $p = 0.058$, $\eta^2 = 0.147$, while there was no significant difference between the two types of primes for the Weak-Strong condition, $F(1, 23) = 1.861$, $p = 0.186$, $\eta^2 = 0.075$. Moreover, Congruency and Prime Type had a significant interaction, $F(1, 23) = 4.566$, $p = 0.043$, $\eta^2 = 0.166$. Subsequent simple effects analysis showed that the accuracy rates in the congruent condition were significantly higher than those in the incongruent condition only for the beat primes, $F(1, 23) = 22.898$, $p < 0.001$, $\eta^2 = 0.499$, while there was no significant difference of the accuracy rates between the congruent and incongruent conditions for the word primes, $F(1, 23) = 1.448$, $p = 0.241$, $\eta^2 = 0.059$. There was no significant difference between the word primes and beat primes for the congruent condition, $F(1, 23) = 1.366$, $p = 0.255$, $\eta^2 = 0.056$. Neither was there significant difference between the two prime types for the incongruent condition, $F(1, 23) = 1.435$, $p = 0.243$, $\eta^2 = 0.059$ (see Figure 3A).

A three-way repeated measures ANOVA was conducted on reaction times (RTs) with Prime Type (word priming vs. beat priming), Stress Pattern (Strong-Weak vs. Weak-Strong) and Congruency (congruent vs. incongruent stress patterns between primes and targets) as independent variables. A significant main effect of Stress Pattern was found, $F(1, 23) = 15.776$, $p = 0.001$, $\eta^2 = 0.407$, showing that the RTs elicited by the Strong-Weak primes ($M = 947.124$, 95%CI [867.544, 1026.705]) were significantly longer than those yielded by the Weak-Strong primes ($M = 877.019$, 95%CI [958.065, 955.974]). There was a significant main effect of

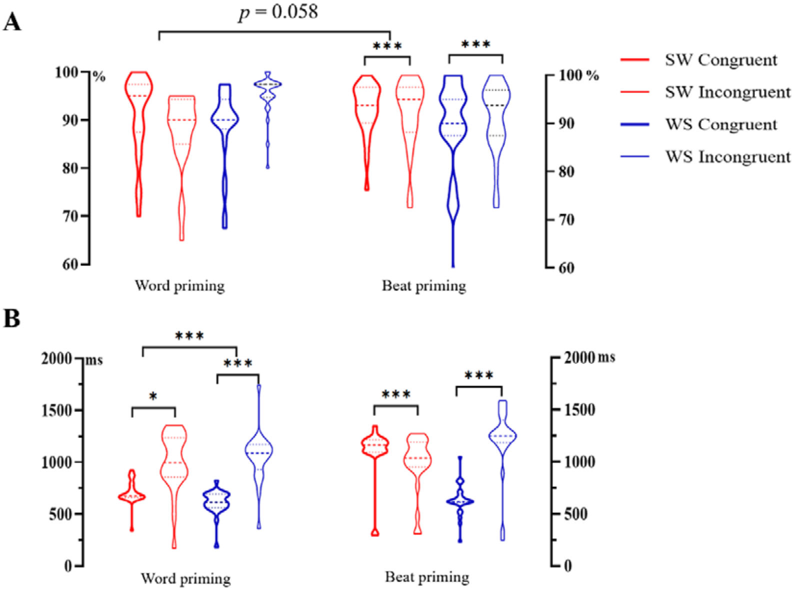

Figure 3. Accuracy (A) and reaction times (B) for the four experimental conditions under word priming and beat priming. SW, Strong-Weak; WS, Weak-Strong.

Congruency, $F(1, 23) = 84.771$, $p < 0.001$, $\eta^2 = 0.787$, revealing that the RTs elicited by the congruent condition ($M = 761.982$, 95%CI [702.100, 821.864]) were significantly shorter than those yielded by the incongruent condition ($M = 1062.162$, 95%CI [959.260, 1165.064]). Prime Type also reached significance, $F(1, 23) = 28.416$, $p < 0.001$, $\eta^2 = 0.553$, with the RTs elicited by the beat primes ($M = 842.313$, 95% CI [770.885, 913.742]) being significantly shorter than those yielded by the word primes ($M = 981.830$, 95%CI [890.824, 1072.736]). There was a significant interaction between Stress Pattern and Congruency, $F(1, 23) = 232.358$, $p < 0.001$, $\eta^2 = 0.910$. Follow-up simple effects analysis showed that while the RTs of the congruent condition were significantly shorter than those of the incongruent condition for both the Weak-Strong prime condition, $F(1, 23) = 202.613$, $p < 0.001$, $\eta^2 = 0.898$, and the Strong-Weak prime condition, $F(1, 23) = 6.659$, $p = 0.017$, $\eta^2 = 0.225$, the RT difference between the congruent and incongruent conditions was larger for the Weak-Strong prime condition than for the Strong-Weak prime condition. There was also a significant interaction between Stress Pattern and Prime Type, $F(1, 23) = 12.691$, $p = 0.002$, $\eta^2 = 0.356$. Follow-up simple effects analysis demonstrated that the RTs elicited by the Strong-Weak primes were significantly longer than those yielded by the Weak-Strong primes only for word priming, $F(1, 23) = 17.544$, $p < 0.001$, $\eta^2 = 0.433$; however, there was no significant difference of RTs between the Strong-Weak and Weak-Strong prime conditions for beat priming, $F(1, 23) = 2.395$, $p = 0.135$, $\eta^2 = 0.094$. In addition, there was a significant interaction between Congruency and Prime Type, $F(1, 23) = 12.000$, $p = 0.002$, $\eta^2 = 0.343$. Simple effects analysis showed that while the RTs elicited by the incongruent condition was significantly longer than those of the congruent condition for the word primes, $F(1, 23) = 38.214$, $p < 0.001$, $\eta^2 = 0.624$, the RTs difference between the incongruent and congruent conditions was greater for the beat primes, with the incongruent condition producing significantly longer RTs than the congruent condition, $F(1, 23) = 93.835$, $p < 0.001$, $\eta^2 = 0.803$. There was a significant three-way interaction between Prime Type, Stress Pattern, and Congruency, $F(1, 23) = 23.5003$, $p < 0.001$, $\eta^2 = 0.535$. Subsequent simple effects analysis showed that for the word primes with the Strong-Weak stress pattern, the RTs in the incongruent condition were longer than those in the congruent condition, $F(1, 23) = 7.946$, $p = 0.010$, $\eta^2 = 0.257$. For the word primes with the Weak-Strong stress pattern, the RTs in the incongruent condition were also longer than those in the congruent condition, $F(1, 23) = 22.270$, $p < 0.001$, $\eta^2 = 0.492$. For the beat primes with the Strong-Weak stress pattern, the RTs in the congruent condition were longer than those in the incongruent condition, $F(1, 23) = 26.127$, $p < 0.001$, $\eta^2 = 0.532$. For the beat primes with the Weak-Strong stress pattern, the RTs in the incongruent condition were longer than those in the congruent condition, $F(1, 23) = 18.486$, $p < 0.001$, $\eta^2 = 0.446$ (see Figure 3B).

3.2. ERP results

A series of three-way repeated measures ANOVAs were conducted on the average amplitudes of the ERP components, with Prime Type (word priming vs. beat priming), Stress Pattern (Strong-Weak vs. Weak-Strong), and Congruency (congruent vs. incongruent stress patterns between primes and targets) as independent variables. Whenever sphericity was not met in a repeated measures ANOVA, the Greenhouse–Geisser was used to correct the degrees of freedom.

3.2.1. P2 (50–150 ms)

Statistical results showed that the interaction between Priming Type, Stress Pattern, and Congruency was significant, $F(1, 23) = 5.684$, $p = 0.026$, $\eta^2 = 0.198$. Simple effects analysis showed that for the word primes with the Strong-Weak stress pattern, the incongruent condition elicited significantly greater P2 amplitudes than the congruent condition, $F(1, 23) = 7.426$, $p = 0.012$, $\eta^2 = 0.244$. For the word primes with the Weak-Strong stress pattern, there was no significant difference between the incongruent and congruent conditions, $F(1, 23) = 0.051$, $p = 0.823$, $\eta^2 = 0.002$. For the beat primes with the Strong-Weak and Weak-Strong stress patterns, there was no significant difference between the incongruent and congruent conditions (Strong-Weak: $F(1, 23) = 0.553$, $p = 0.465$, $\eta^2 = 0.023$, Weak-Strong: $F(1, 23) = 0.750$, $p = 0.395$, $\eta^2 = 0.032$). In addition, for word priming, the Weak-Strong congruent condition elicited larger P2 amplitudes than the Strong-Weak congruent condition, $F(1, 23) = 3.952$, $p = 0.059$, $\eta^2 = 0.147$, whereas for beat priming, there was no significant difference in the P2 amplitudes elicited by the Weak-Strong congruent and Strong-Weak congruent conditions, $F(1, 23) = 3.146$, $p = 0.088$, $\eta^2 = 0.120$ (Figure 4).

3.2.2. N400 (300–450 ms)

Statistical results revealed that the main effect of stress pattern was significant, $F(1, 23) = 6.895$, $p = 0.015$, $\eta^2 = 0.231$. The Strong-Weak primes ($M = -0.641$, 95%CI [−1.559, 0.277]) elicited significantly greater N400 amplitudes than the Weak-Strong primes ($M = 0.135$, 95%CI [−0.776, 1.046]) (Figure 5).

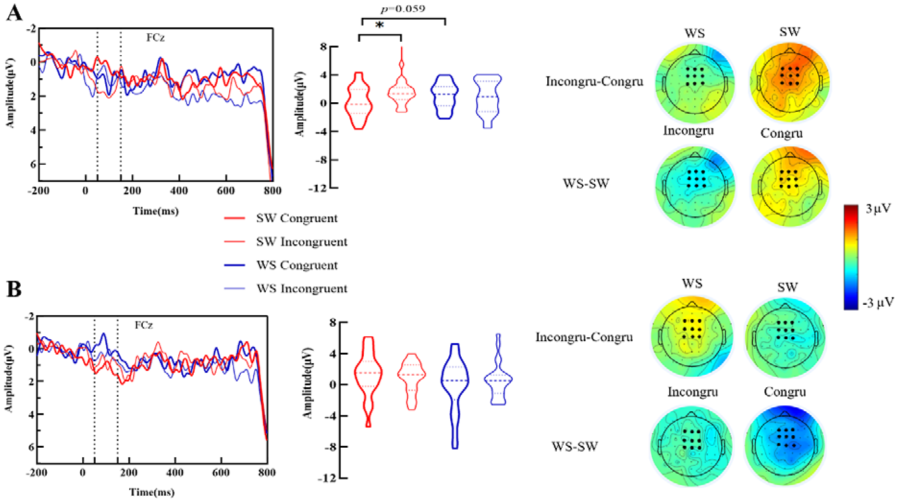

Figure 4. Word priming (A) and beat priming (B) elicited waveforms (left), violin plots of P2 wave amplitudes for each experimental condition (middle), and topography of differences for the four experimental conditions (right), where the small black dots indicate ROIs (C1, Cz, C2, FC1, FCz, FC2, F1, Fz, F2). SW, Strong-Weak; WS, Weak-Strong.

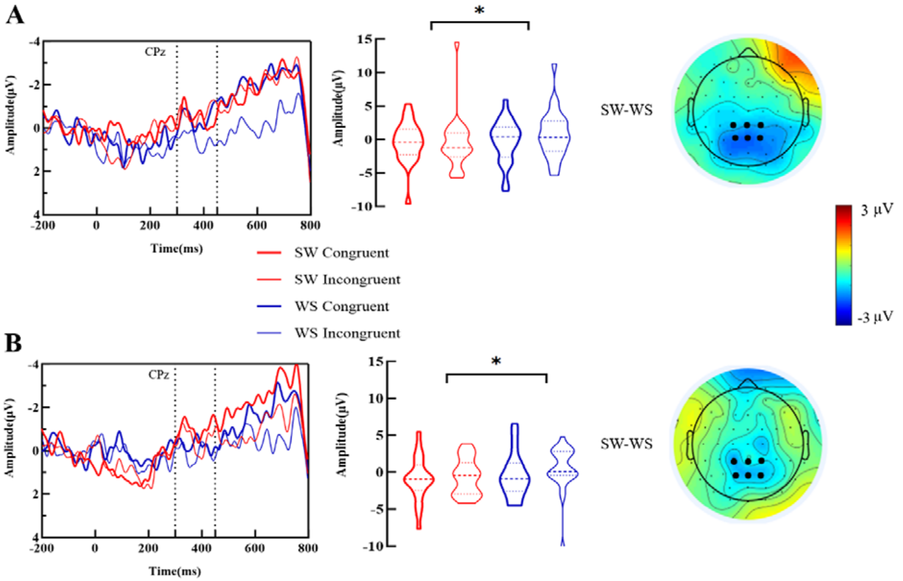

Figure 5. Word priming (A) and beat priming (B) elicited waveforms (left), violin plots of N400 wave amplitudes for each experimental condition (middle), and topography of differences for the experimental conditions (right), where the small black dots indicate ROIs (P1, Pz, P2, CP1, CPz, CP2). SW, Strong-Weak; WS, Weak-Strong.

3.2.3. LPC (500–700 ms)

Statistical results illustrated that the main effect of stress pattern was significant, $F(1, 23) = 10.235$, $p = 0.004$, $\eta^2 = 0.308$, with the Weak-Strong primes ($M = -0.995$, 95%CI [−2.028, 0.038]) elicited significantly larger LPC amplitudes than the Strong-Weak primes ($M = -1.822$, 95%CI [−2.842, −0.803]). The main effect of Congruency was also significant, $F(1, 23) = 4.696$, $p = 0.041$, $\eta^2 = 0.170$, with the incongruent stress patterns between the primes and targets showing significantly larger LPC amplitudes than the congruent stress pattern (incongruent: $M = -0.781$, 95%CI [−1.997, 0.434]; congruent: $M = -2.036$, 95%CI [−1.997, −0.434]) (Figure 6).

4. Discussion

This study explored the effects of word priming and beat priming on Mandarin lexical stress recognition by manipulating the stress pattern and the congruency of stress between the prime sequences and target words. Behavioral results showed that the RTs elicited by the Strong-Weak primes were longer than those yielded by the Weak-Strong primes. In the Strong-Weak condition, the accuracy rates produced by the word primes were higher than those by the beat primes. In the Weak-Strong condition, the accuracy rates produced by the word and beat primes had no significant difference. Furthermore, the accuracy rates and RTs elicited by the congruent stress pattern between the primes and targets were higher/shorter than those yielded by the incongruent stress pattern. The ERP results revealed that the incongruent stress pattern between the primes and targets elicited larger P2 and LPC amplitudes than the congruent stress pattern (Böcker et al., 1999; Bohn et al., 2013; Domahs et al., 2008; Marie et al., 2011; McCauley et al., 2013;

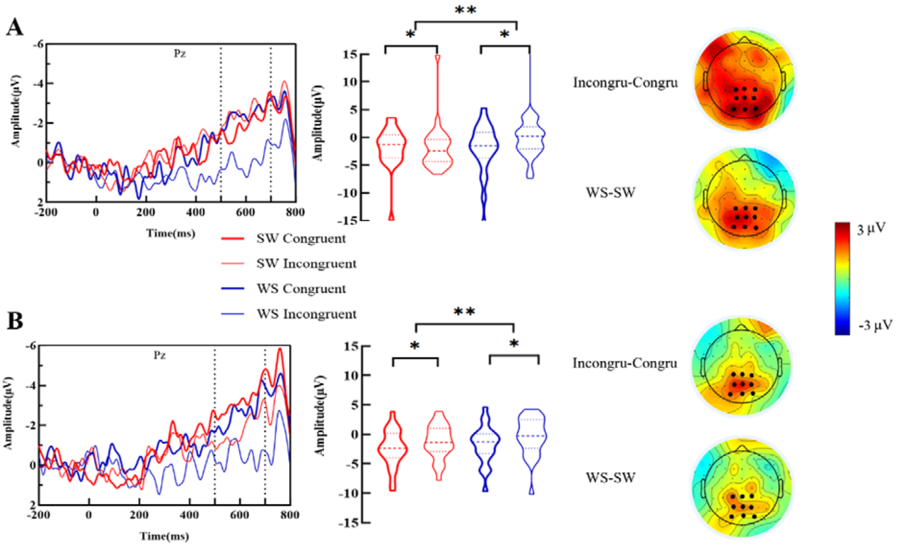

Figure 6. Word priming (A) and beat priming (B) elicited waveforms (left), violin plots of LPC wave amplitudes for each experimental condition (middle), and topography of differences for the experimental conditions (right), where the small black dots indicate ROIs (CP1, CPz, CP2, P1, Pz, P2, PO3, Poz, PO4). SW, Strong-Weak; WS, Weak-Strong.

Schmidt-Kassow & Kotz, 2009). Moreover, the Strong-Weak prime sequences led to larger N400 amplitudes than the Weak-Strong prime sequences (Bohn et al., 2013; Magne et al., 2016; Marie et al., 2011). For the comparisons of beat priming and word priming, except for the difference in the early time window (50–150 ms), no effect of Prime Type was found in either the middle (300–450 ms) or late time windows (500–700 ms). The present study showed that there were cross-domain interactions between music and language in lexical stress recognition, although the neural mechanisms regarding how they influence lexical stress recognition of Mandarin were different.

For the word primes with the Strong-Weak stress pattern, significantly larger P2 amplitudes were found when the stress patterns between the primes and targets were incongruent, while for the word primes with the Weak-Strong stress pattern, there was no P2 difference between the incongruent and congruent stress patterns. Given that the P2 component is related to the early processing load, the larger P2 amplitudes for the Strong-Weak incongruent priming indicate increasing difficulty in processing the target words with the unexpected Weak-Strong stress pattern (Huang et al., 2018; Marie et al., 2011). As the brain processes information, it constantly predicts new information based on the characteristics of previously presented signals and incorporates the predicted information into the pre-existing representational structures. Therefore, when the new information is incongruent with the regularity of the previous information, the pre-existing representational structure would not be able to accommodate the new information, resulting in processing difficulties. In addition, some studies have illustrated that the P2 component may be related to the process of rapid attention (Fan et al., 2016; Kanske et al., 2011). In the current study, greater P2 elicited by the Strong-Weak incongruent priming may indicate that stress violation of the Weak-Strong targets attracts more attentional resources, leading to an increase of P2 amplitudes (Fan et al., 2016; Zhang et al., 2021).

In addition, greater P2 amplitudes were also observed for the Weak-Strong congruent priming condition relative to the Strong-Weak congruent priming condition. This effect may be related to the subjective stress pattern. Previous studies have found that listeners have strong preference for the strong beat/strong syllable falling in the odd position (i.e., 1, 3, 5,…), showing the strong-weak binary structure (Bolton, 1894; Drake, 1993; Fraisse, 1982). Such stress preference has been attested in Indo-European languages such as Dutch (Böcker et al., 1999), where approximately 88% of words carry the Strong-Weak stress pattern (Vroomen & de Gelder, 1995), and in tonal languages such as Mandarin, where native Mandarin listeners showed preference for trochaic blocks in their perception of sound intensity (Yu et al., 2019). The accuracy results of the current study also confirm native Mandarin listeners' preference for the lexical stress pattern of Strong-Weak, in that the accuracy rates elicited by the word primes with the Strong-Weak stress pattern were higher than those yielded by the beat primes with the same stress pattern, but there was no significant accuracy difference between the word and beat primes having the Weak-Strong stress pattern, indicating that the Strong-Weak stress pattern may be the stored representation in Mandarin listeners' mental lexicon. The Weak-Strong stress pattern of words did not conform to the default binary structure pattern of Mandarin listeners, thus producing larger P2 amplitudes, reflecting an automatic and early process of lexical stress awareness (Olson et al., 2001).

In Mandarin, the second syllable of some words loses its original tone, becoming the so-called unaccented, or neutral tone (Lin, 1962; Lin & Yan, 1980). The larger P2 amplitudes elicited by the Weak-Strong words than by the Strong-Weak words may also indicate that Mandarin listeners require more resources to process words having the relatively atypical Weak-Strong pattern. However, there are controversies over the stress pattern of Mandarin disyllabic words. Some scholars believe that the Strong-Weak pattern is more typical (Wang & Feng, 2006), while others claim that the Weak-Strong pattern is more common (Lin et al., 1984; Xu, 1982), and still others propose that the two are at random (Liu, 2007; Zhou, 2018). The results of the present study suggest that the Strong-Weak stress pattern may be more typical in Mandarin disyllabic words.

Compared to studies of non-tonal languages (Böcker et al., 1999; Marie et al., 2011), the P2 time window in this study is relatively early, indicating that Mandarin listeners may be more sensitive to stress (Bidelman et al., 2011, 2013; Hutka et al., 2015; Tong et al., 2018), so that they were able to detect stress violation more rapidly whenever there was one.

With regard to the N400 results, for both word priming and beat priming, we found that the Strong-Weak primes elicited greater N400 amplitudes than the Weak-Strong primes. Although the semantic relatedness between the prime sequences and targets was controlled in the current study, some studies have shown that semantic processing is human intuition (i.e., default) for language processing (Kutas & Federmeier, 2011). In the experiment of the present study, the participants' task was quite simple, which only required them to pay attention to the stress pattern of the stimuli. Therefore, the participants may have extra attentional resources to process the meanings of the words. Moreover, the behavioral results demonstrated that the RTs induced by the beat primes were significantly shorter than those elicited by the word primes, indicating that more information, such as lexical information, for the word primes than for the beat primes has to be processed by the listeners, thus, slowing down the speed of lexical stress recognition. Importantly, the behavioral results also showed that the RTs elicited by the Strong-Weak primes were longer than those yielded by the Weak-Strong primes, which further confirms that the participants may have carried out additional semantic processing for the Strong-Weak words.

Lexical processing depends not only on the suitable semantic context, but also on regular prosody (Rothermich et al., 2010; Rothermich & Kotz, 2013). Violation of subtle rhythmic preferences can consume additional processing resources (Bohn et al., 2013; Henrich et al., 2014). Furthermore, atypical stress patterns are not stored in the mental lexicon, so processing these stress patterns may hinder semantic understanding at the same time (Marie et al., 2011; Schiller et al., 2004). Given that the Weak-Strong stress pattern may be more atypical and does not conform to listeners' default binary structure, making it harder for listeners to form rhythm expectations (Cason et al., 2015a), more cognitive resources should be needed to process this stress pattern (Schröger, 1996). As a result, the lexical information of the Weak-Strong words may not be able to be processed so thoroughly that smaller N400 amplitudes were obtained relative to the Strong-Weak words. The greater N400 amplitudes for the Strong-Weak words may be due to the fact that the listeners not only focused on the stress patterns of the words. The greater N400 amplitudes for the Strong-Weak words (irrespective of the congruency) may be due to the fact that the listeners not only focused on the stress patterns of the words, but most likely accessed the semantic information them. Further, a paired-sample *t*-test on the accuracy of the word recall test revealed that the recall rate of the Strong-Weak words (87.00%) was significantly higher than that of the Weak-Strong words (77.50%), $t\ (1,19) = 2.89$, $p < 0.01$. This result reinforces the explanation that the participants were likely to process the semantic information of the Strong-Weak words more deeply than that of the Weak-Strong words, leading to a larger N400.

Consistent with Böcker et al. (1999), our behavioral results showed that the accuracy rates were higher and the RTs were shorter when the prime sequences and the targets shared the same stress pattern. Our ERP results showed that incongruent priming elicited larger LPC amplitudes than congruent priming. Moreover, both word and beat primes with the Weak-Strong stress pattern produced larger LPC amplitudes than those with the Strong-Weak stress pattern. Studies on explicit and implicit prosodic processing have shown that prosodic violations are likely to induce late positive components (Kriukova & Mani, 2016; Rothermich et al., 2012; Schmidt-Kassow & Kotz, 2009). The LPC is thought to be related to error detection (Kolk et al., 2003), reanalysis, and re-attention processes (Zhang et al., 2010), reflecting the integration process of tasks (Bornkessel & Schlesewsky, 2006). The larger LPC amplitudes for incongruent priming may be due to the fact that the participants devoted more attentional resources to reanalyzing the stress violation of the targets and integrating the unexpected stress pattern into the stress pattern of the prime sequence. This explanation is consistent with the predictive coding theory and the theory of expectancy-driven speech processing in that individuals need to readjust their predictions and re-generate expectations at a lower level to suppress predictive errors when there is unpredicted information (Friston, 2005; Rubin et al., 2016). When the auditory input is consistent with the predicted information, the auditory rhythm drives the information processing, consuming less cognitive resources (Kotz & Schwartze, 2010, 2015). As the P2 results, larger LPC amplitudes were yielded by the Weak-Strong primes compared to the Strong-Weak primes, revealing that the Weak-Strong stress pattern may not be in line with Mandarin listeners' expectations of the Strong-Weak binary structure and preference for rhythm perception. Hence, the listeners needed to put more psychological effort into readjusting expectations to accept new pieces of information while processing the stimuli (Friston, 2002). Our LPC results support Jones' dynamic attention theory (Jones et al., 2002; Jones & Boltz, 1989; Large & Jones, 1999), which states that attention resources would be re-adjusted to optimize the dynamic prediction and processing of events when unexpected events occur. The fact that the effect of Stress Pattern was still observed at a later time window reflects that stress processing is not only influenced by earlier attention, but also by later cognitive processing at a higher level.

There are some limitations in this study that warrant future research. First, the duration of the beat prime sequences (5061 ± 693 ms) and that of the word prime sequences (2967 ± 441 ms) were not normalized. Future studies are needed to investigate the effects of prime duration on lexical stress recognition. Second, the present study did not separate the factors of pitch, intensity, and duration; neither did it test the individual contributions of these factors to lexical stress recognition. Future studies can further investigate the contribution of each of these factors to lexical stress recognition. Finally, the cross-domain bidirectional influence between music and language, especially whether and how language affects musical processing, should be further examined.

5. Conclusions

This study found that although both word priming and beat priming facilitated the recognition of Mandarin lexical stress, beat priming had no effect on the recognition of lexical stress in the early processing stage. In the late time window, there was no significant difference between the two prime types. In addition, priming effects were also modulated by stress congruency between primes and targets, as well as stress patterns. We propose that Mandarin listeners need to consume more cognitive resources when processing the incongruent stress and Weak-Strong stress patterns. To sum up, stress in music and language can influence each other across domains.

Data availability statement. The data from the study can be found on the Open Science Framework at https://osf.io/xgjyq/?view_only=0cf01be4e6754a97816823d69fbeb9fc.

References

Bent, T., Bradlow, A. R., & Wright, B. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception Performance*, 32(1), 97–103. https://doi.org/10.1037/0096-1523.32.1.97

Besson, M., Chobert, J., & Marie, C. (2011). Language and music in the musician brain. *Language: Linguistics Compass*, 5, 617–634. https://doi.org/10.1111/j.1749-818X.2011.00302.x

Bidelman, G. M., Gandour, J., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, 23, 425–434. https://doi.org/10.1162/jocn.2009.21362

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PLoS ONE*, 8(4), e60676. https://doi.org/10.1371/journal.pone.0060676

Böcker, K. B. E., Bastiaansen, M. C. M., Vroomen, J. H. M., Brunia, C. H. M., & de Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology*, 36(6), 706–720. https://doi.org/10.1111/1469-8986.3660706

Bohn, K., Knaus, J., Wiese, R., & Domahs, U. (2013). The influence of rhythmic (ir)regularities on speech processing: Evidence from an ERP study on German phrases. *Neuropsychologia*, 51(4), 760–771. https://doi.org/10.1016/j.neuropsychologia.2013.01.006

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology*, 6, 145–238.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, 113(4), 787–821. https://doi.org/10.1037/0033-295X.113.4.787

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language Cognitive Processes*, 25(7), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Breen, M., Fitzroy, A. B., & Oraa Ali, M. (2019). Event-related potential evidence of implicit metric structure during silent reading. *Brain Sciences*, 9(8), 192. https://doi.org/10.3390/brainsci9080192

Brochard, R., Abecasis, D., Potter, D., Ragot, R., & Drake, C. (2003). The "Ticktock" of our internal clock direct brain evidence of subjective accents in isochronous sequences. *Psychological Science*, 14(4), 362–366. https://doi.org/10.1111/1467-9280.24441

Cao, J. (1986). Characterization of Mandarin light syllables (in Chinese). *Journal of Applied Acoustics*, 4, 1–6.

Cason, N., Astésano, C., & Schön, D. (2015a). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica*, 155, 43–50. https://doi.org/10.1016/j.actpsy.2014.12.002

Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015b). Rhythmic priming enhances speech production abilities: Evidence from prelingually deaf children. *Neuropsychology*, 29(1), 102–107. https://doi.org/10.1037/neu0000115

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, 50(11), 2652–2658. https://doi.org/10.1016/j.neuropsychologia.2012.07.018

Chang, D., Hedberg, N., & Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America*, 139(5), 24–32. https://doi.org/10.1121/1.4947497

Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Phonetics*, 36 (4), 724–746. https://doi.org/10.1016/j.wocn.2008.06.003

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech–evidence from F_0 patterns of neutral tone in Standard Chinese. *Phonetica*, 63(1), 47–75. https://doi.org/10.1159/000091406

Cooper, G. W., & Meyer, L. B. (1971). *The rhythmic structure of music*. The University of Chicago.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, 45, 207–228. https://doi.org/10.1177/00238309020450030101

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception Psychophysics*, 20(1), 55–60. https://doi.org/10.3758/BF03198706

Cutler, A., & Donselaar, W. (2001). Voornaam is not (really) a Homophone: Lexical Prosody and Lexical Access in Dutch. *Language and Speech*, 44, 171–195. https://doi.org/10.1177/00238309010440020301

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. John Murray.

Domahs, U., Wiese, R., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). The processing of German word stress: evidence for the prosodic hierarchy. *Phonology*, 25(1), 1–36. https://doi.org/10.1017/S0952675708001383

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics*, 53, 25–33. https://doi.org/10.3758/BF03211712

Duanmu, S. (2007). *The phonology of Standard Chinese*. Oxford University Press.

Elena, F., Luisa, L., Chiara, T., Marcella, M., Stefania, Z., & Daniele, S. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS ONE*, 10(9), e0138715. https://doi.org/10.1371/journal.pone.0138715

Ellis, R. J., & Jones, M. R. (2010). Rhythmic context modulates foreperiod effects. *Attention, Perception Psychophysics*, 72(8), 2274–2288. https://doi.org/10.3758/BF03196701

Elmer, S., Klein, C., Kühnis, J., Liem, F., Meyer, M., & Jäncke, L. (2014). Music and language expertise influence the categorization of speech and musical sounds: behavioral and electrophysiological measurements. *Journal of Cognitive Neuroscience*, 26(10), 2356–2369. https://doi.org/10.1162/jocn_a_00632

Fan, W., Zhong, Y., Li, J., Yang, Z., Zhan, Y., Cai, R., & Fu, X. (2016). Negative emotion weakens the degree of self-reference effect: evidence from ERPs. *Frontiers in Psychology*, 7, 1048. https://doi.org/10.3389/fpsyg.2016.01408

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Feld, S., & Fox, A. (1994). Music and language. *Annual Review of Anthropology*, 23, 25–53.

Fotidzis, T. S., Moon, H., Steele, J. R., & Magne, C. L. (2018). Cross-modal priming effect of rhythm on visual word recognition and its relationships to music aptitude and reading achievement. *Brain Sciences*, 8(12), 210. https://doi.org/10.3390/brainsci8120210

Fraisse, P. (1982). *The psychology of music: Rhythm and tempo*. Academic Press.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, 23(9), 2038–2043. https://doi.org/10.1093/cercor/bhs180

François, C., & Schön, D. (2014). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: the role of musical practice. *Hearing Research*, 308, 122–128. https://doi.org/10.1016/j.heares.2013.08.018

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience*, 16(4), 541–552. https://doi.org/10.1162/089892904323057281

Friston, K. J. (2002). Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience*, 25, 221–250. https://doi.org/10.1146/ANNUREV.NEURO.25.112701.142846

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Garrido, M. I., Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proceedings of the National Academy of Sciences*, 104(52), 20961–20966. https://doi.org/10.1073/pnas.0706274105

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, 120(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Henrich, K., Alter, K., Wiese, R., & Domahs, U. (2014). The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain and Language*, 136(2014), 19–30. https://doi.org/10.1016/j.bandl.2014.07.003

Hilton, C. B., & Goldwater, M. B. (2020). Linguistic syncopation: Meter-syntax alignment affects sentence comprehension and sensorimotor synchronization. *Cognition*, 217, 104880. https://doi.org/10.31234/osf.io/hcngm

Huang, X., Liu, X., Yang, J.-C., Zhao, Q., & Zhou, J. (2018). Tonal and vowel information processing in Chinese spoken word recognition: An event-related potential study. *NeuroReport*, 29(5), 356–362. https://doi.org/10.1097/WNR.0000000000000973

Hutka, S., Bidelman, G. M., & Moreno, S. (2015). Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. *Neuropsychologia*, 71, 52–63. https://doi.org/10.1016/j.neuropsychologia.2015.03.019

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychology review*, 96(3), 459–491. https://doi.org/10.1037/0033-295X.96.3.459

Jones, M. R., MacKenzie, H., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, 13(4), 313–319. https://doi.org/10.1111/1467-9280.00458

Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia*, 49(11), 3121–3129. https://doi.org/10.1016/j.neuropsychologia.2011.07.022

Koelsch, S., Gunter, T. C., Cramon, D. Y. v., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *NeuroImage*, 17, 956–966. https://doi.org/10.1006/nimg.2002.1154

Kolk, H. H., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and language*, 85(1), 1–36. https://doi.org/10.1016/s0093-934x(02)00548-5

Kotz, S. A., & Gunter, T. C. (2015). Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease? *Annals of the New York Academy of Sciences*, 1337, 62–68. https://doi.org/10.1111/nyas.12657

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: A timely subcortico-cortical framework. *Trends in Cognitive Sciences*, 14(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Kotz, S. A., & Schwartze, M. (2015). Motor-timing and sequencing in speech production: A general-purpose framework. In G. Hickok, & S. L. Small (Eds.), *Neurobiology of language* (pp. 717–724). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00057-2

Kratochvil, P., & Chao, Y. R. (1970). A grammar of spoken Chinese. *Language*, 46(2), 513–524. https://doi.org/10.2307/412300

Kriukova, O., & Mani, N. (2016). Processing metrical information in silent reading: An ERP study. *Frontiers in Psychology*, 7, 1432. https://doi.org/10.3389/fpsyg.2016.01432

Krumhansl, C. (1991). *Cognitive foundations of musical pitch*. Oxford University Press.

Krumhansl, C. L. (2000). Rhythm and pitch in music. *Cognition Psychological Bulletin*, 126, 159–179.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Ladányi, E., Lukács, Á., & Gervain, J. (2021). Does rhythmic priming improve grammatical processing in Hungarian-speaking children with and without developmental language disorder? *Developmental Science*, 24(6), e13112. https://doi.org/10.1111/desc.13112

Large, E. W., Herrera, J., & Velasco, M. J. (2015). Neural networks for beat perception in musical rhythm. *Frontiers in Systems Neuroscience*, 9, 159. https://doi.org/10.3389/fnsys.2015.00159

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119–159. https://doi.org/10.1037/0033-295X.106.1.119

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26(1), 1–37. https://doi.org/10.1016/S0364-0213(01)00057-X

Lee, W.-S., & Zee, E. (2010). Articulatory characteristics of the coronal stop, affricate, and fricative in Cantonese. *Journal of Chinese Linguistics*, 38(2), 336–372. https://doi.org/10.2307/23754137

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5, 253–263. https://doi.org/10.1016/S0095-4470(19)31139-8

Lerdahl, F., & Jackendoff, R. (1985). *A generative theory of tonal music* (Vol. 9, pp. 72–73). MIT Press. https://doi.org/10.2307/843535

Li, W., Deng, N., Yang, Y., & Wang, L. (2018). Process focus and accentuation at different positions in dialogues: An ERP study. *Language, Cognition and Neuroscience*, 33(2), 255–274. https://doi.org/10.1080/23273798.2017.1387278

Lin, D. (1962). The relationship between modern Chinese allophones and syntactic structure (in Chinese). *Zhongguo Yuwen*, 7, 1.

Lin, M., & Yan, J. (1980). The acoustic properties of the Beijing light voice (in Chinese). *Dialect*, 3, 166–178.

Lin, M., Yan, J., & Sun, G. (1984). Preliminary experiments on the normal stress of two character sets in Beijing dialect (in Chinese). *Dialect*, 1, 57–73.

Lin, T. (1985). *Tantao Beijinghua qingyin xingzhi de chubu shiyan [On neutral tone in Beijing Mandarin]*. Peking University Press.

Liu, X. (2007). A study on the rhythmic key of modern Chinese (in Chinese). *Language Teaching and Linguistic Studies*, 3, 56–62.

London, J. (2004). *Hearing in time*. Oxford University Press.

Lu, Z. (1984). A preliminary study on the acoustic properties of Mandarin diphthongs of the "heavy-medium" and "medium-heavy" forms (in Chinese). *Chinese Language Learning*, 6, 41–48.

Luo, C., & Wang, J. (2002). *Outline of general phonetics*. The Commercial Press.

Lutz, J. (2012). The relationship between music and language. *Frontiers in Psychology*, 3(3), 123. https://doi.org/10.3389/fpsyg.2012.00123

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, 4, 540–545. https://doi.org/10.1038/87502

Magne, C. L., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. *Cerebral Cortex*, 17(11), 2659–2668. https://doi.org/10.1093/CERCOR/BHL174

Magne, C. L., Jordan, D. K., & Gordon, R. L. (2016). Speech rhythm sensitivity and musical aptitude: ERPs and individual differences. *Brain and Language*, 153–154, 13–19. https://doi.org/10.1016/j.bandl.2016.01.001

Marie, C., Magne, C. L., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, 23(2), 294–305. https://doi.org/10.1162/jocn.2010.21413.

McCauley, S. M., Hestvik, A., & Vogel, I. (2013). Perception and bias in the processing of compound versus phrasal stress: Evidence from event-related brain potentials. *Language and Speech*, 56(1), 23–44. https://doi.org/10.1177/0023830911434277

Mirka, D. (2004). Hearing in time: Psychological aspects of musical meter. *Journal of Music Theory*, 48(2), 325–336. https://doi.org/10.1215/00222909-48-2-325

Olson, I. R., Chun, M. M., & Allison, T. (2001). Contextual guidance of attention: human intracranial event-related potential evidence for feedback modulation in anatomically early temporally late stages of visual processing. *Brain: A Journal of Neurology*, 124(7), 1417–1425. https://doi.org/10.1093/BRAIN/124.7.1417

Parmentier, F. B. R., Elsley, J. V., Andrés, P., & Barceló, F. (2011). Why are auditory novels distracting? Contrasting the roles of novelty, violation of expectation and stimulus change. *Cognition*, 119(3), 374–380. https://doi.org/10.1016/j.cognition.2011.02.001

Patel, A. D. (2007). *Music, language, and the brain*. Oxford University Press.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, 2(142), 1–14. https://doi.org/10.3389/fpsyg.2011.00142

Peretz, I., Nguyen, S., & Cummings, S. (2011). Tone language fluency impairs pitch discrimination. *Frontiers in Psychology*, 2, 145. https://doi.org/10.3389/fpsyg.2011.00145

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception & Performance*, 16(3), 564–573. https://doi.org/10.1037//0096-1523.16.3.564

Plack, C. J., Oxenham, A. J., Fay, R. R., & Popper, A. N. (2005). Pitch: Neural coding and perception. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Springer handbook of auditory research*. Springer.

Potter, D. D., Fenwick, M., Abecasis, D., & Brochard, R. (2009). Perceiving rhythm where none exists: Event-related potential (ERP) correlates of subjective accenting. *Cortex*, 45(1), 103–109. https://doi.org/10.1016/j.cortex.2008.01.004

Rothermich, K., & Kotz, S. A. (2013). Predictions in speech comprehension: fMRI evidence on the meter-semantic interface. *NeuroImage*, 70(2013), 89–100. https://doi.org/10.1016/j.neuroimage.2012.12.013

Rothermich, K., Schmidt-Kassow, M., & Kotz, S. A. (2012). Rhythm's gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, 50(2), 232–244. https://doi.org/10.1016/j.neuropsychologia.2011.10.025

Rothermich, K., Schmidt-Kassow, M., Schwartze, M., & Kotz, S. A. (2010). Event-related potential responses to metric violations: rules versus meaning. *NeuroReport*, 21(8), 580–584. https://doi.org/10.1097/WNR.0b013e32833a7da7

Rubin, J., Ulanovsky, N., Nelken, I., & Tishby, N. (2016). The representation of prediction error in auditory cortex. *PLoS Computational Biology*, 12(8), e1005058. https://doi.org/10.1371/journal.pcbi.1005058

Schiller, N., Fikkert, P., & Levelt, C. C. (2004). Stress priming in picture naming: An SOA study. *Brain and Language*, 90(2004), 231–240. https://doi.org/10.1016/S0093-934X(03)00436-X

Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, 21(9), 1693–1708. https://doi.org/10.1162/jocn.2008.21153

Schröger, E. (1996). The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes. *Electroencephalography and Clinical Neurophysiology*, 100(6), 517–526. https://doi.org/10.1016/S0168-5597(96)95576-8

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin Review*, 16, 374–381. https://doi.org/10.3758/16.2.374

Song, H. (2009). *Mandarin concise light and heavy format dictionary* (in Chinese). Shanghai Music Press.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory Language*, 45, 412–432. https://doi.org/10.1006/JMLA.2000.2783

Sun, G. (1999). Acoustic cues of stressed vowel in sentences. In *Proceedings of the 4th congress of phonetic sciences* (pp. 109–112). Jin Cheng Press.

Sun, Y., Sommer, W., & Li, W. (2022). How accentuation influences the processing of emotional words in spoken language: An ERP study. *Neuropsychologia*, 166, 108144. https://doi.org/10.1016/j.neuropsychologia.2022.108144

Todorović, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *The Journal of Neuroscience*, 31(25), 9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011

Tong, X., Choi, W., & Man, Y. (2018). Tone language experience modulates the effect of long-term musical training on musical pitch perception. *The Journal of the Acoustical Society of America*, 144(2), 690–697. https://doi.org/10.1121/1.5049365

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception Performance*, 21(1), 98–108. https://doi.org/10.1037/0096-1523.21.1.98

Wang, Z., & Feng, S. (2006). Tonal contrast and disyllabic stress patterns in Beijing Mandarin (in Chinese). *Linguistic Sciences*, 1, 3–22. https://doi.org/10.3969/j.issn.1671-9484.2006.01.001

Xu, S. (1956). Outline of basic knowledge of Mandarin phonetics (in Chinese). *Chinese Language Learning*, 7, 19–22.

Xu, S. (1982). Volume analysis of bisyllabic words (in Chinese). *Language Teaching and Linguistic Studies*, 2, 4–19.

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

Yin, Z. (1982). A preliminary analysis on the unstress and stress of Mandarin bilabial common words (in Chinese). *Chinese Language*, 2, 169–173.

Yin, Z. (2021). A study of Chinese word stress (in Chinese). *Chinese Journal of Phonetics*, 2021(2), 95–109.

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

Yu, W., Fan, P., Yu, J., & Liang, D. (2019). The study of the rhythmic grouping perception on Chinese Mandarin (in Chinese). *Journal of Psychological Science*, 42(2), 293–298.

Zhang, Q., Li, X., Gold, B. T., & Jiang, Y. (2010). Neural correlates of cross-domain affective priming. *Brain Research*, 1329, 142–151. https://doi.org/10.1016/j.brainres.2010.03.021

Zhang, Z., Zhang, H., Wang, B., Zheng, Z., Zhao, L., & Li, W. (2021). The processing of the tone and the vowel in poems under different tasks: evidence from ERPs (in Chinese). *Studies of Psychology and Behavior*, 19 (6), 728–735.

Zheng, Y., & Samuel, A. G. (2018). The effects of ethnicity, musicianship, and tone language experience on pitch perception. *Quarterly Journal of Experimental Psychology*, 71(12), 2627–2642. https://doi.org/10.1177/1747021818757435

Zhou, R. (2018). A review on Mandarin word stress study in 60 years (in Chinese). *Language Teaching and Linguistic Studies*, 6, 102–112. https://doi.org/10.3969/j.issn.0257-9448.2018.06.013

Cite this article: Yu, W., Chien, Y. -F., Wang, B., Zhao, J., & Li, W. (2024). The effects of word and beat priming on Mandarin lexical stress recognition: an event-related potential study, *Language and Cognition*, 1–23. https://doi.org/10.1017/langcog.2023.75