The Roles of Segments and Tone in Mandarin Lexical Processing: An ERP Study

Dan Du¹, Minghu Jiang¹,*

¹School of Humanities, Tsinghua University, 100083 Beijing, China
*Correspondence: jiang.mh@tsinghua.edu.cn (Minghu Jiang)

Abstract

Backgrounds: Segments and tone are important sub-syllabic units that play large roles in lexical processing in tonal languages. However, their roles in lexical processing remain unclear, and the event-related potential (ERP) technique will benefit the exploration of the cognitive mechanism in lexical processing. Methods: The high temporal resolution of ERP enables the technique to interpret rapidly changing spoken language performances. The present ERP study examined the different roles of segments and tone in Mandarin Chinese lexical processing. An auditory priming experiment was designed that included five types of priming stimuli: consonant mismatch, vowel mismatch, tone mismatch, unrelated mismatch, and identity. Participants were asked to judge whether the target of the prime-target pair was a real Mandarin disyllabic word or not. Results: Behavioral results including reaction time and response accuracy and ERP results were collected. Results were different from those of previous studies that showed the dominant role of consonants in lexical access in mainly non-tonal languages like English. Our results showed that consonants and vowels play comparable roles, whereas tone plays a less important role than do consonants and vowels in lexical processing in Mandarin. Conclusions: These results have implications for understanding the brain mechanisms in lexical processing of tonal languages.

Keywords: Mandarin Chinese; segments; spoken language processing; tone

1. Introduction

Segments and tone are key components of syllables in tonal languages, and they play important roles in language processing. Existing studies on the roles of segments and tone in lexical access can be divided into two categories. On the one hand, prior studies concerned only segments in tonal languages mainly concluded that vowels play more important roles in lexical access than do consonants. Chen et al. [1] carried out a Mandarin auditory sentences recognition paradigm through noise-replacement technique and concluded that the vowel-only sentences yielded twice intelligibility as the consonant-only sentences and a small piece of vowel-consonant transition would greatly promote the intelligibility of the consonant-only sentences. Two years later, Chen et al. [2] found similar results in an auditory lexical recognition task using different proportions of segments that there is a vowel advantage over consonant in isolated lexical recognition. On the other hand, some studies concerning both segments and tone in spoken lexical processing of tonal languages. For instance, in two tasks with an auditory priming paradigm carried out by Sereno and Lee [3], participants were asked to make a lexical decision after hearing each prime-target pair in both experiments. The results presented that segments play major roles while tone play secondary role in lexical processing. In contrast, Wiener and Turnbull [4] extended the word recognition task to Mandarin—a tonal language and found that changes to vowels exhibit longer response time and lower accuracy than the changes to consonants, changes to tones and changes to any part showed the shortest response time and highest accuracy. Furthermore, in an auditory word recognition task with different contexts (including word, sentence and idiom conditions), the results showed tonal roles in Mandarin lexical access varied with task conditions with more advantages in contexts than in isolation [5].

Recently, techniques such as eye tracking [6,7] and event-related potentials (ERP) [8,9] have been applying in investigating lexical processing of Mandarin. In a spoken word recognition task [8], the roles of tones and vowels were explored by asking participants make judgement whether the last word was correct or not. The ERP results showed early negativity appeared only for vowel violation and the amplitude of N400 of vowel violation were larger than that of tone violation, indicating that vowels play more significant roles than do tones. Similarly, Zou et al. [9] explored Mandarin sentences comprehension with a tone/rime violation paradigm, in which participants were required to judge whether the sentences were comprehensible. The ERP results revealed that rime violation elicited larger N400 and smaller P600 amplitude than tone violation, indicating vowel advantage in semantic processing and tone advantage in error recovery.

To summarize, although some studies have investigated the roles of sub-syllabic units in Mandarin spoken word comprehension, congruent conclusions have not been reached. Findings such as vocalic advantages in some stud-
ies concerning only the roles of segments in lexical processing [1,2], segmental advantages in some studies concerning segments and tones [3,4] and tonal advantages in certain conditions especially within contexts rather than in isolation [5] were found. Thus, the current study was aimed at investigating the roles between segments and tones, vowels and consonants respectively while applying ERP technique. Although ERP technique has been applied in this field already, additional material types need to be explored by this technique in order to facilitate the understanding of various spoken language phenomena. The reasons why the present study was carried out were listed as follows: (1) The high temporal resolution of ERP enables the technique to interpret rapidly changing spoken language performances. (2) Different from previous studies, disyllables were selected because they are very common in Mandarin and play significant roles in spoken language comprehension. (3) Five auditory violation types (detailed information were shown in the Material session) were designed to take a close examination of consonantal, vocalic and tonal mismatch in lexical processing comprehensively.

In addition, the ERP component that is of our interest in this article is P300 elicited in the process of decision making, which refers to a spike in activity approximately 300 ms after the presentation of the target stimulus. More specifically, P300, spanning in a time window about 250–500 ms [10,11] is thought to reflect processes involved in stimulus categorization. The amplitude of the P300 is proportional to the amount of attentional resource devoted to the task and the degree of information processing required, and the latency is a measure of stimulus classification speed [12–14]. In the semantic priming paradigm, P300 has been observed in response to semantically unrelated targets, which means that the processing of semantically unrelated targets required more attention than that needed for semantically related targets [15].

2. Materials and Methods

2.1 Materials

Disyllabic words are the most common forms in Mandarin Chinese, and they bear significant roles in language comprehension [16,17]. To make clear the roles of sub-syllabic units (e.g., consonant) in lexical processing, disyllables were regarded as stimulus carriers. Sixty disyllabic words were selected from Xiandai Hanyu Pinlv Cidian (Modern Chinese Frequency Dictionary) [18] as targets with a high frequency, to avoid the influence brought about by unfamiliar words. The critical word was embedded in the second syllable of the disyllable, which covers the phonological allocations of consonants and vowels as far as possible, and they are semantically meaningful. For each target disyllable, five types of primes were created: (1) initial consonant mismatch (CM): the prime had the initial consonant of the second syllable changed, but all other components remained the same (e.g., guān qián - guān jiān); (2) vowel mismatch (VM): the prime had the vowel of the second syllable changed, but all other components remained the same (e.g., guān jiao - guān jiān); (3) tone mismatch (TM): the prime had the tone of the second syllable changed, but all other components remained the same (e.g., guān jiān - guān jiān); (4) unrelated mismatch (UM): the prime had the initial consonant, the vowel and the tone of the second syllable all changed (e.g., guān lin - guān jiān); (5) identity condition (I): the target was preceded by itself (e.g., guān jiān - guān jiān). To avoid the phenomenon of Mandarin tone sandhi (specific tone change) carried by disyllables, the tone pair of T3/T3 was excluded. In addition, second syllables that have no initial consonant (e.g., bāo ān, means “security”) were also excluded as they did not conform to the research topic. Moreover, another 60 pseudo-words served as fillers. Pseudo-words stimuli were manipulated in the same way as real disyllables to form five types of experimental conditions. The pooling of target disyllable pairs and filler pairs yielded 600 pairs of disyllables.

The stimuli were recorded by a female native speaker in a sound treated room using SONY microphone (WH-CH720N, Sony Corporation, Tokyo, Japanese); the speaker had passed the Level 1-B Putonghua Proficiency Test (Level 1-B is required for Chinese-language teachers in northern China). Disyllables were recorded at a sampling rate of 44.1 kHz and 16-bit resolution. The speaker was told to read materials at a normal speed and intensity.

2.2 Participants

A total of 30 native Mandarin speakers from Tsinghua University (18 males, 12 females) were paid to participate in the experiment. The mean age of the participants was 24.79 years (standard deviation (SD) = 4.617 years). Two of the participants were excluded from the analysis because of excessive artifacts in their electroencephalogram (EEG) data. By means of a self-filled questionnaire, all participants reported that they did not have any hearing problems and they had normal or corrected-to-normal vision. Informed consent was obtained from all the participants before the experiment. All participants were right-handed, as assessed by the Edinburgh Handedness Inventory [19]. Ethical approval of the experiment was obtained from the Human Research Ethics Committee of Tsinghua University (approval number: 202322).

2.3 Procedure

The experimental procedure was carried out in a soundproof booth. Participants were seated in a comfortable chair in front of a computer monitor. Stimuli were randomly presented using E-Prime (Psychology Software Tools, version 2.0, Psychology Software Tools Inc, Pittsburgh, PA, USA) through headphones at a comfortable listening level. To make the procedure clear, visual instructions were given on the screen and participants could consult with the experimenter if there were any questions about
the experiment procedure before the experiment. An experiment trial was started with a 100 ms warning beep, followed by the auditory presentation of a pair of disyllables while a red fixation cross appeared on the screen. Participants were asked to pay attention to the second disyllable and perform a lexical decision task by pressing the buttons, on a hand-held responder, labeled “YES” and “NO” to indicate whether the second disyllable was a real Chinese word or not when the question mark appeared on the screen. There was a maximum response time of 2000 ms for participants, and the intertrial interval was 1000 ms. For half of the participants the right button was used to signal the “YES” response. For the rest participants, the order was reversed. They were told to respond as quickly and accurately as possible.

The whole experiment was comprised of one practice session and four experimental sessions. The warm-up trial, lasting about 5 minutes, containing 40 pairs of disyllables that did not occur in the ensuing experiment sessions. To familiarize participants with the procedure, the practice session could be repeated if necessary. There was a short break after each block of 150 pairs of disyllables. There were 600 pairs of disyllables in four blocks in total. Participants were asked to refrain from blinking and making extra body movements, as much as possible, during the experiment. The whole process lasted for about 2 hours.

2.4 Behavioral Data Analysis

We first used the interquartile range (IQR) method to identify outliers, and incorrect responses and latencies that exceed “−1.5” * IQR to “+1.5” * IQR were excluded from the data analysis. Reaction times (RT) and response accuracy (RA) of both pseudo-words and words were collected and analyzed. To make clear the difference between pseudo-words and real words, two-way analysis of variance (ANOVA) with Word Type (pseudo-words/real words) × Mismatch Type (CM, VM, TM, I, UM) was applied for both RA and RT. Then, one-way repeated measures ANOVA was carried out for RA and RT of real words, using the five priming conditions (consonant mismatch, vowel mismatch, tone mismatch, unrelated mismatch and identity condition) as within-subject factors. Statistical analysis performed
with SPSS (version 26, International Business Machines Corporation, New York, NY, USA), Bonferroni correction was used to correct the problem of multiple comparisons when applicable and p values of <0.05 were considered to be statistically significant.

2.5 EEG Recording and Analyses

Scalp voltages were recorded with a sampling rate of 500 Hz from 64 cap-mounted electrodes (Brain Products GmbH, Gilching, Germany) positioned according to the extended International Standard 10–20 System. We used AFz as the ground electrode, and FCz as online reference electrodes. Eye movements were recorded with additional electrodes located below the left eye (the vertical electrooculogram, VEOG) and at the lateral canthus of the right eye (the horizontal electrooculogram, HEOG). The ERP signals were filtered with an analogue bandpass filter of 0.01–100 Hz and were sampled continuously throughout the experiment with a sampling rate of 500 Hz. Interelectrode impedance was maintained below 10 KΩ for all channels.

The EEG data were analyzed with Brain Vision Analyzer software (version 2.0, Brain Products GmbH, Gilching, Germany). The ERP signals were re-referenced to the average of VEOG and HEOG and filtered with a bandpass filter from 0.1 to 30 Hz. We used independent component analysis to deal with EEG signals contaminated with eye blinks. Epochs of the EEG were from 100 ms pre-stimulus to 800 ms post-stimulus, and time-locked to the onset of auditory presentation. Those that correspond to correct responses and were free of ocular (blinks and movements) and muscular artifacts were averaged and analyzed (more than 94% of trials). Baseline correction was performed using the average EEG activity in the 200 ms preceding the onset of the target stimulus as a reference signal value.

One time-window was selected based on the mean Global Field Power [20] as shown in Fig. 1. A set of 24 electrodes was used for analysis; and they were divided into six areas of four electrodes each (see Fig. 2). They were Left Frontal (F3, F1, FC3, FC1), Right Frontal (F2, F4, FC2, FC4), Left Central (C3, C1, CP3, CP1), Right Central (C2, C4, CP2, CP4), Left Posterior (P3, P1, PO7, PO3), and Right Posterior (P2, P4, PO4, PO8) (F means frontal lobe, C means central, P means parietal lobe, O means occipital lobe). For each area, the mean amplitude of the four electrodes was calculated and used for subsequent analyses.

3. Results

3.1 Behavioral Results

The mean RA and the mean RT for real words and pseudo-words were shown in Fig. 3 and Fig. 4. We found significant differences in RA and RT between pseudo-words and real words through statistical analysis. There was a main effect of word type, and subjects showed lower mean RA (F = 25.42, p = 0.000, η² = 0.086) and longer mean RT for pseudo-words than for real words (F = 53.77, p = 0.000, η² = 0.175). Pseudo-words were used to help listeners avoid selection inertia, so we focused on performance for real words in the following.

Repeated measures ANOVA showed a significant effect of mismatch types on RT (F (1, 28) = 28.78, p = 0.039, η² = 0.20) and RA (F (2, 33) = 19.57, p = 0.040, η² = 0.09) of real words. Post-hoc pairwise comparisons with Bonferroni adjustment of RA revealed significant differences between CM and I (p = 0.040), VM and I (p = 0.040), UM and I (p = 0.031) whereas there was no significant difference between CM and VM (p = 0.340), TM and I (p = 0.107). Post-hoc pairwise comparisons with Bonferroni adjustment of RT revealed significant differences between CM and I (p = 0.030), VM and I (p = 0.044), UM and I (p = 0.000), CM and VM (p = 0.039) whereas there was no significant difference between TM and I (p = 0.366).

3.2 ERP Results

We saw from the scalp maps (Fig. 5) and grand-average waveforms for auditory stimuli (Fig. 6) that ERPs displayed a positive component from 250 ms to 360 ms. After one-way repeated measures ANOVA analysis, we found that there was a significant main effect of mismatch types (F (4, 152) = 31.92, p = 0.017, η² = 0.180) of P300 amplitude for real words. Post-hoc pairwise comparisons with Bonferroni adjustment indicated that the P300 amplitude of the mismatch type of CM (p = 0.0310), VM (p = 0.028), and UM (p = 0.070), but not TM (p = 0.063), showed significantly greater amplitude than did I, and there existed significant differences between CM and TM (p = 0.000), CM and UM (p = 0.007), VM and TM (p = 0.007), VM and UM (p = 0.010), and TM and UM (p = 0.039) of the P300 amplitude, although there was no significant differences between CM and VM (p = 0.193).
Fig. 3. The mean reaction time (RT) for five different mismatch types of pairing of real words (RW) and pseudo-words (PW).

Fig. 4. The mean response accuracy (RA) for five different mismatch types of pairing of real words (RW) and pseudo-words (PW).
4. Discussion

The goal of the present study was to investigate the roles of segments and tone in Mandarin lexical processing combining both behavioral measurement and ERP method. An auditory mismatch paradigm was adopted with the target syllables embedded in the second syllable of disyllables. We tested the performances of 28 native Mandarin speakers (excluded data of two participants with excessive artifacts in their EEG data), using five types of prime stimuli that differed minimally in either a consonant, a vowel, or a tone, then collected and analyzed the performances of all participants.

Overall, behavioral results indicated that when the initial consonant or vowel of the second syllable in prime disyllables was changed, the lexical identification was delayed and the accuracy was decreased. In addition, when the initial consonant, vowel, and tone of the second syllable in prime disyllables were all different from those of the target, the lexical items were processed more slowly and less correctly. However, when only tone of the second syllable in prime disyllables was changed, the processing of lexical items was not different from the identity condition.

ERP results revealed that the prime type of the identity condition produced the lowest mean amplitude of P300 of the five mismatch types. Using the identity condition as the baseline, the consonant mismatch, the vowel mismatch, and the unrelated mismatch produced larger mean amplitudes of P300 whereas the tone mismatch was not significantly different from the identity condition. In further analysis, the consonant mismatch and the vowel mismatch did not produce significant differences in the amplitude of P300, although they produced larger amplitude than both the unrelated mismatch and the tone mismatch. In addition, the unrelated mismatch produced a larger amplitude of P300 than did the tone mismatch. Based on previous research, the P300 is related to neurophysiological mechanisms of decision making [21–23]. The amplitude of P300 becomes larger when much attention needs to be paid to decision making. As our ERP results showed, the consonant mismatch and the vowel mismatch produced larger amplitudes of P300 than did the unrelated mismatch, and the unrelated mismatch produced a larger amplitude of P300 than did the tone mismatch and the identity condition. There were no significant differences between the consonant mismatch and the vowel mismatch, or between the tone mismatch and the identity condition.

In sum, three main results can be found: (1) When the initial consonant or vowel of the second syllable in the prime disyllable is changed, much more effort is needed to process lexical identification. Furthermore, when the initial consonant, vowel, and tone of the second syllable in the prime disyllable are all different from those in the target disyllable, the amplitude of P300 is smaller than that produced by the consonant mismatch and the vowel mismatch. Generally, it is thought that when consonant, vowel, and tone are all changed, much more attention is needed to deal with the lexical processing than when only the initial consonant or vowel is changed. However, more obvious differences produce quicker discrimination in the real process of lexical identification. It is more difficult to detect subtle differences than obvious differences between prime and target [24]. (2) When only the tone of the second syllable in prime disyllables is changed, the P300 amplitude is not significantly different from that for the baseline situation. Behavioral results showed that tone mismatch condition reached no significant difference with identity condition, which was consistent with ERP results. Although tone is used to distinguish word meanings in Mandarin Chinese, it bears the lowest information load. This would be compatible with findings suggesting insensitivity of tone mismatch by both children and adults. In a preferential looking task that included trial types of vowel mispronunciation, consonant mispronunciation, tone mispronunciation, and correct pronunciation, 6-year-old Mandarin monolingual children showed no significant difference in distinguishing correct pronunciations from tone mispronunciations [25]. Similarly, Tong et al. (2008) [26] concluded, using a speeded classification paradigm, that tone is the least informative mode when compared with the rime and the consonant. Another study extended the word reconstruction paradigm to Mandarin
and found that participants responded faster and more accurately to tone change conditions than to segmental change conditions [4]. The information load hypothesis also states that tone contributes less than consonants and vowels in lexical access. That is similar to the results in Indo-European languages that stress carries less information than consonant or vowel in pitch-accent languages plays less important role in lexical access [27,28]. (3) Finally, consonants and vowels in Mandarin Chinese play comparable roles in lexical processing, which is different from existing conclusions in which consonant bias (consonants are more important than vowels in lexical processing (when learning and recognizing words), and are supported by evidences ranging from different experimental paradigms) occupies a main role in some other languages [29]. For instance, when asked to turn word-like pseudowords into words by changing phonemes in a word reconstruction task, participants made faster responses with more accuracy when changing vowels rather than consonants [30]. The reason may lie in the phonological systems. For example, the essential part of a word in English includes vowel (V), consonant (C), or a cluster of consonants (CC or CCC), and some example words are shown as CV, CCV, CCCV, CVCCC, etc. [31]. The most common syllable structure of Mandarin Chinese is the monosyllable, in which the segment part is composed of a consonant (C) as the initial and a vowel (V)

**Fig. 6. Grand average event-related potential (ERP) waveforms.** The ERPs for consonant mismatch (CM), vowel mismatch (VM), tone mismatch (TM), unrelated mismatch (UM), identity (I) are shown in purple, green, blue, red and black, respectively.
as the final [32]. It seems that consonant and vowel in Mandarin Chinese contribute almost the same in respect of syllable structure whereas consonants contribute more than vowels in English. However, some research on Mandarin, using syllables of CV structure, showed different results. Chen et al. [1] determined, by using a noise-replacement paradigm, that vowel-only sentence conditions were rather more intelligible than consonant-only ones. Later, Chen et al. [2] designed a lexical recognition task using synthesized stimuli and found that participants recognized vowel-only isolated Mandarin words faster and more accurately than consonant-only words. Those two studies on Mandarin implied that vowels contribute more than consonants in the recognition of isolated Mandarin monosyllables. The present study suggested that the relative contributions of consonants and vowels in Mandarin lexical processing were influenced by contexts, i.e., consonants and vowels might contribute nearly equal in disyllables. Comparing studies carried out in English and Mandarin, it is clear that consonants are valued more than vowels in lexical processing in English for there are many more multisyllables in English [33] that make consonants seem like the skeletons of words. On the other hand, vowels contribute more than consonants in isolated syllable recognition in Mandarin [2]. Note that materials used in the present study were disyllables in which the structure is more like the syllabic structure of English multisyllables, such a phonological structure is located between English words and Mandarin monosyllable words, resulting in comparable roles of consonants and vowels in Mandarin disyllables processing.

There are also two limitations in the present study. On the one hand, more syllable types need to be considered although disyllables are very common in Mandarin. On the other hand, the sample size used in each prime condition is not large enough. Future studies should consider various types of syllables as stimuli and enlarge the sample size in experimental design.

5. Conclusions

This study is an ERP investigation into the roles of segments and tone in Mandarin auditory word recognition. The main findings are listed as follows: on the one hand, consonants and vowels play more important role in lexical access than does tone in Mandarin Chinese, mainly because tone carries a lower information load which brings down its role in distinguishing word meanings. On the other hand, consonants and vowels play similar roles in Mandarin lexical processing due to syllable structure. These findings help to understand different cognitive mechanisms of lexical processing in a tonal language like Mandarin Chinese.

Availability of Data and Materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author Contributions

MHJ and DD designed the research study. MHJ and DD performed the research. DD analyzed the data. DD wrote the manuscript. Both authors contributed to editorial changes in the manuscript. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Ethical approval of the experiment was obtained from the Human Research Ethics Committee of Tsinghua University (approval number: 202322). Informed consent was obtained from all the participants before the experiment.

Acknowledgment

We thank two anonymous reviewers for their comments and suggestions which helped improve the manuscript.

Funding

This work was supported by the Major Program of the National Natural Science Foundation of China (grant number: 62036001) and the project funded by China Postdoctoral Science Foundation (grant number: 2023M732015).

Conflict of Interest

The authors declare no conflict of interest.

References


