Language and Speech

Article

Language *and* Speech
2015, Vol. 58(2) 131–151

DOI: 10.1177/0023830914522956
las.sagepub.com

The Contribution of Segmental and Tonal Information in Mandarin Spoken Word Processing

Joan A Sereno and Hyunjung Lee
Department of Linguistics, University of Kansas, USA

Abstract
Two priming experiments examined the separate contribution of lexical tone and segmental information in the processing of spoken words in Mandarin Chinese. Experiment 1 contrasted four types of prime–target pairs: tone-and-segment overlap (ru4-ru4), segment-only overlap (ru3-ru4), tone-only overlap (sha4-ru4) and unrelated (qin1-ru4) in an auditory lexical decision task with 48 native Mandarin listeners. Experiment 2 further investigated the minimal segmental overlap needed to trigger priming when tonal information is present. Four prime–target conditions were contrasted: tone-and-segment overlap (ru4-ru4), only onset segment overlap (re4-ru4), only rime overlap (pu4-ru4) and unrelated (qin1-ru4) in an auditory lexical decision task with 68 native Mandarin listeners. The results showed significant priming effects when both tonal and segmental information overlapped or, although to a lesser extent, when only segmental information overlapped, with no priming found when only tones matched. Moreover, any partial segmental overlap, even with matching tonal cues, resulted in significant inhibition. These data clearly indicate that lexical tones are processed differently from segments, with syllabic structure playing a critical role. These findings are discussed in terms of the overall architecture of the processing system that emerges in Mandarin lexical access.

Keywords
Auditory priming, segments and tones, Mandarin

1 Introduction

The recognition of spoken words is a key aspect of language comprehension. To recognize words, one must extract segmental and suprasegmental information from the speech signal. This is then

Corresponding author:
Joan A Sereno, Department of Linguistics, University of Kansas, 425 Blake Hall, 1541 Lilac Lane, Lawrence, KS 66045, USA.
Email: sereno@ku.edu

mapped onto internal representations in the mental lexicon. The basic issue addressed by the current study is the individual contribution of segmental and tonal information in the recognition of Mandarin words.

There are many sources of variability in word recognition, including suprasegmental factors such as stress, intonation and rate of speech. Extraction of suprasegmental information and its role in word recognition models (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Ellman, 1986; Norris, McQueen, & Cutler, 2000) have received relatively little attention since most studies have focused on a number of Indo-European languages where suprasegmental differences play a minor role in providing lexical contrasts. Although in many lexical stress languages the stress position in the word is fixed and stress is not lexically distinctive, a few languages use stress pattern to distinguish word meanings (see, for example, Gussenhoven, 2004; Hayes, 1995). In English, for example, two words that have the same segmental structure but contrast in terms of stress can differ in meaning (for example, obJECT versus OBject, with upper case indicating stress). However, relatively few English words are distinguished by this contrasting stress pattern (Fry, 1955; Lieberman, 1960; Sereno & Jongman, 1995).

Several studies have examined the contribution of segmental and suprasegmental information in word recognition. Soto-Faraco, Sebastian-Galles, and Cutler (2001), for example, investigated the contribution of suprasegmental information in auditory word recognition in Spanish. Spanish words differing in suprasegmental information (for instance, saBAna 'savannah' versus SAbana 'sheet') were used as stimuli in cross-modal (auditory-visual) fragment priming experiments. In the experiments, auditory primes (the first two syllables) were used as word fragments and presented at the end of a sentence. Visual targets either matched the auditory primes (had identical stress pattern and segmental structure) (for example, PRINci- from PRINcipe 'prince') or differed from the auditory primes in either stress pattern or segmental structure (one vowel or one consonant). Priming occurred when the prime and the target completely overlapped in stress pattern and segmental structure, while a comparable inhibition effect was found when the prime and target mismatched either suprasegmentally or segmentally. Soto-Faraco et al. (2001) found that both suprasegmental and segmental information similarly influenced word recognition, suggesting that accentual information does constrain activation and selection of word candidates.

Cutler and Otake (1999) examined Japanese words contrasting minimally in pitch accent. They found that words consisting of the same segments and contrasting only in pitch accent did not prime each other in an auditory lexical decision task. Cutler and Otake (1999) concluded that Japanese listeners used accentual information to constrain lexical activation and selection of word candidates.

The role of segmental and suprasegmental information has also been investigated in tone languages, where suprasegmental information plays a more important role. In tonal languages, tones are used to distinguish word meanings. Mandarin Chinese, for example, is a language with four distinct lexical tones: a high level pitch (Tone 1); a high rising pitch (Tone 2); a low dipping pitch (Tone 3); and a high falling pitch (Tone 4) (Jongman, Wang, Moore, & Sereno, 2006; Sereno & Wang, 2007). A syllable can combine with each of the four tones, changing its meaning. For example, the syllable 'ma' means mother with a Tone 1 (ma1), hemp with a Tone 2 (ma2), horse with a Tone 3 (ma3) and scold with a Tone 4 (ma4). Given these contrasting lexical items, tone languages provide fertile ground for examining the contribution of suprasegmental and segmental information to word recognition.

Previous studies using various experimental paradigms have demonstrated that listeners of tonal languages rely more on segmental than tonal information in spoken word recognition. Taft and Chen (1992) used homophone decision tasks to investigate participants' sensitivity to tone information in Mandarin and Cantonese. In a read aloud and a silent task, both Mandarin and Cantonese participants responded more slowly to mismatches in tone before making decisions, suggesting that syllables with the same segmental structure but different tone were more easily accepted as homophones. Overall, tones tended to be ignored when making homophone judgments in the processing of Mandarin and Cantonese isolated words.

Using a lexical decision task, Cutler and Chen (1995) examined the effect of phonological similarity of prime and target in disyllabic Cantonese words, examining mismatch in tone or rime in the first or second syllable. Although the results generally showed that tone and rime mismatch overall had similar priming effects, the effects varied as a function of mismatch position. While phonological similarity showed facilitation with matching second syllables, there was less robust facilitation when the first syllable differenced in tone than when it differed in rime.

Cutler and Chen (1997) further used speeded (lexical decision and same-different) tasks to investigate the processing of tonal and segmental information in Cantonese syllables. Disyllabic nonwords were constructed, differing by onset, vowel or tone from original disyllabic words. The results showed that the subjects had more errors when the nonwords differed in tone. That is, in lexical decision, the subjects were more likely to accept the nonwords as real words when the disyllabic items only mismatched in tone. Moreover, pairs of syllables differing in tone in a same-different task also received slower responses and more errors than other combinations. The authors suggested that tonal information was processed more slowly since tonal information becomes available later than the segmental information that bears the tone.

Yip, Leung, and Chen (1998) examined forward and backward priming in a shadowing task. For the forward priming condition (where the target of shadowing was the second syllable), results showed that shadowing was facilitated by overlapping segmental structure but inhibited by tonal overlap, consistent with the findings of Taft & Chen (1992), Cutler & Chen (1997) and Zhou (2000), showing a tone disadvantage. Yip et al. (1998) suggested that participants were more inefficient in using stored tonal information than in using stored segmental information in the processing of spoken words.

Yip (2001) further examined the influence of phonological relatedness in Cantonese using a direct priming task. Yip (2001) also found that a facilitation effect was observed only when the prime and the target shared segmental structures. Interestingly, when primes shared onset and tone with the target, an inhibition effect was found, results consistent with the findings of Slowiaczek and Hamburger (1992), Radeau et al. (1995) and McQueen and Sereno (2005), showing competition between lexical representations of phonologically related words in non-tonal languages. Yip (2001) suggested that the Cantonese speakers are more sensitive to segmental information than tonal information in the processing of Cantonese.

Ye and Connine (1999) used vowel and tone monitoring tasks to investigate the tonal information in Mandarin spoken word processing. Consistent with the findings of Cutler and Chen (1997), the results showed that subjects responded slower to tone-mismatching stimuli than to vowel-mismatching stimuli. However, the vowel advantage disappeared in constraining idiomatic contexts. Moreover, contrasting the degree of the tone mismatch (close versus far mismatch) indicated that the activation of lexical tone is not a categorical process because graded tone mismatches modulated performance.

Overall, these studies show that mismatching tonal information does not slow response times, with listeners able to pick out mismatching segmental stimuli faster and more accurately than mismatching tone stimuli. Not only do listeners rely more on segmental information than tonal information, but segmental processing also seems to occur earlier than the processing of tonal information.

However, there are a few studies that came to the opposite conclusion, suggesting either a tone advantage over segments or no difference in processing between the two. Schirmer, Tang, Penney, Gunter, and Chen (2005) used event-related potentials, ERPs, to examine the time course of processing in Cantonese. Stimuli mismatched either in tone, segment, or both. The results showed that while mismatched targets in Cantonese caused a larger negativity than matching targets, the time course and the amplitude of the early negativity (N400) and the late positivity (P650) were not different for segmental and tonal violations. In contrast to many earlier studies but similar to the results of Soto-Faraco et al. (2001) for Spanish, Schirmer et al. (2005) showed inhibition for both segmentally and suprasegmentally incongruous targets, suggesting that tonal and segmental semantic violations contributed equally in a sentence completion task.

Liu and Samuel (2007) examined further the role of Mandarin tones in different contextual situations: in an isolated word, in an idiomatic phrase (a condition similar to Ye & Connine, 1999), and in a sentence context. Their results showed that identification accuracy was equivalent in the word and idiom conditions; however, in a more constraining sentence context, the subjects showed a tone advantage and made more errors in the tone mismatch condition than other mismatch conditions. Similar to Schirmer et al. (2005), these data indicate that when the context provided enough information, the segmental advantage disappeared and tonal cues played a more important role.

More recently, C.-Y. Lee (2007) used priming tasks to contrast the role of Mandarin segments and tones in constraining lexical activation as well as examining the time course of this activation by presenting prime–target pairs at different ISIs. In these experiments, primes and targets either were related in both segments and tones (lou2-lou2), overlapped only in segmental structure but differed in tone (lou3-lou2), overlapped only in tone but differed in segmental structure (cang2-lou2), or shared neither segmental nor tonal structure (pan1-lou2). C.-Y. Lee found a facilitation effect only when the prime and target were identical in both segmental structure and tone. While no facilitation was found when the prime and target shared only tone, surprisingly, none also was observed with only segmental overlap, with reaction times for these conditions no different than the unrelated control condition. Since mismatching tonal information did not show facilitation, C.-Y. Lee suggested that tone information, similar to segmental information, was used to block inappropriate lexical candidates and constrain lexical activation.

However, the lack of segmental mismatching data and C.-Y. Lee's own discussion about a possible issue due to tonal overlap are of concern, given the perceptual effects of tonal similarity that have been documented especially for contour tones (Gandour, 1983; Kiriloff, 1969) as well as the data from Ye and Connine (1999) showing a modulation of monitoring effects due to degree of tonal similarity. Specifically, in C.-Y. Lee (2007), neither the number of target tones that were presented nor, more importantly, the fact that the tonal distribution in prime–target pairs was also not balanced across conditions confounded the results. For example, for the segment-only condition, where stimuli were to have no tonal overlap, more confusable prime–target pairs (for example, tone2 and tone3 combinations) were presented much more often than distinct tonal combinations (for example, tone1 and tone2 combinations) (19 versus 8 presentations, respectively). Given such an unequal design, the segmental facilitation cannot be separated from the contribution of the tonal similarity, allowing any possible segmental priming effect to be confounded by the interfering tonal interaction.

In the current study, two experiments are conducted to further clarify the nature and processing of tonal information in Mandarin spoken word recognition. The present study used a direct priming task to examine whether or not listeners are able to separately use segmental and tonal information to identify Mandarin words. Experiment 1 replicated the first experiment of C.-Y. Lee's (2007) study but with a modification that balanced the tonal information in the prime–target pairs. It is hypothesized that a facilitation effect may be found when the prime and target words are identical in both tonal and segmental structures or when prime and target overlap only in segmental structure. If, however, tonal information is immediately used to block inappropriate lexical candidates, facilitation may not be found when prime and target have segmental overlap only and a facilitation effect may be found in tone overlap pairs.

2 Experiment 1

2.1 Methods

2.1.1 Participants. Forty-eight subjects (19 females, 29 males) participated in Experiment 1. The ages of the subjects ranged from 19 to 32 years and all of the subjects were native speakers of Mandarin. All the subjects were students at the Chinese Academy of Sciences in Beijing. None of the subjects reported any speech disorders; and they were paid for their participation.

2.1.2 Materials. The experimental stimuli included 96 targets, consisting of 48 words and 48 nonwords. For Experiment 1, the prime and target words are listed in Appendix 1. Each word target was paired with four types of primes: ST (matched in all segments and tone), S (matched only in segments), T (matched only in tone) and UR (unrelated). In the ST condition, the primes and targets were identical in both segmental structure and tone (for example, ru4–ru4). In the S condition, the primes and targets were matched only in segmental structure, but differed in their tones (for example, ru3–ru4). For this condition, there was an equal distribution of prime–target tonal mismatches so comparisons could be made across conditions. In the T condition, the primes and targets were matched only in tones, but differed in their segmental content (for example, sha4–ru4). In this condition, no segments overlapped. In the UR condition, the primes shared neither segmental structure nor tone with targets (for example, qin1–ru4).

The 48 word targets were distributed equally across each of the tones. There were 12 word stimuli for each of the target tones (Tones 1, 2, 3 and 4). By balancing the number of targets for each target tone, we avoided having a particular target tone condition occur more than another, and we could observe the contribution of each of the four target tones across the prime conditions.

The frequency counts of word stimuli were obtained from Da Jun's (1998) corpus in which 45 million words were analyzed. For the present study, the log frequency of ST, S, T and UR primes were 2.11, 2.03, 2.32 and 2.15, respectively. There was no significant difference among the four types of primes in terms of frequency of occurrence (F (3, 45) = 1.959, p = 0.134, ns).

The 48 nonword targets functioned as fillers. To create similar conditions in the nonwords, each nonword target was paired with two types of word primes: T (matched in tone) and UR (unrelated). In the T condition, the primes and nonword targets were matched only in tone (for example, zen3 – sai3) but differed in their segmental structure. In the UR condition, the prime shared neither segmental nor tonal information with the nonword targets (for example, jun4 – sai3).

Since no prime–target pair was repeated for a participant, four stimulus sets were constructed for Experiment 1. For a given target item, each of the four prime conditions was assigned to a different set such that each set included an equal number of prime types without repeating any particular stimulus item in a set. Each of the four stimulus sets included all 96 targets (48 words, 48 nonwords). Within a stimulus set, twelve prime words for each of ST, S, T and UR conditions were paired with each of 48 word targets, and 24 primes from each of the T and UR conditions were paired with the 48 nonword targets, respectively. For an equal distribution of stimuli across participants, the 48 subjects were divided into four groups (12 subjects for each group), who were randomly assigned to a stimulus set. Thus, participants did not hear the same stimulus item more than once during the experiment.

2.1.3 Procedure. The stimuli were recorded by a male native speaker of Mandarin Chinese in an anechoic chamber at the University of Kansas, using a cardioid microphone (Electrovoice-N/D-767) and a digital recorder (Marantz PMD 671). The recorded stimuli were transferred to a PC at a sampling rate of 22.05 kHz, and then segmented using Praat (Boersma & Weenink, 2007). Onset and offset of each stimulus was determined from amplitude measures in the waveforms.

The extracted audio files were presented to participants using the Paradigm software package (Tagliaferri, 2008). Subjects were tested in a quiet computer room at the laboratory of the Chinese Academy of Sciences in Beijing. Subjects listened to the stimuli through headphones connected to a PC. The participants' task was to make a lexical decision regarding the target item. After hearing each prime–target pair, the participant was instructed to respond to the target as quickly and accurately as possible by pressing the buttons labeled '是' (Yes) and '不是' (No) on a computer keyboard. Both response time (ms) from the offset of the target to the response press as well as response accuracy were recorded.

The prime–target pairs within each of the four stimulus sets were randomized. The ISI between the prime and target was 250 ms, and the inter-trial interval between each pair was 3 seconds. Prior to the actual experiment, 8 trials were given to the subjects as a practice to provide familiarization with the task. The experimental session took approximately 15 minutes.

2.2 Results

The participants' reaction time and accuracy were analyzed using *lmer*() function in R (version 2.15.3) from the *lme4* package (Bates & Maechler, 2009). For both the reaction time and accuracy analyses, the regression models contain one fixed Prime variable with four levels (ST, S, T and UR) and two random effects, Subject and Item. For the Prime factor, the UR condition was always set as the baseline. The *p*-values reported in this paper were obtained through the Markov chain Monte Carlo method (MCMC), using the *pvals.fnc*() function of the *languageR* package.

For the reaction times (RTs), the data were trimmed such that reaction times above or below two standard deviations of each subject's mean (113 trials) as well as all errors (114 trials) were omitted from further analysis which resulted in the exclusion of 9.0% of the data. Mean reaction times (averaged across participants) were calculated from correct responses only. Mean reaction times for each prime condition are shown in Figure 1.

The mixed-effects model for reaction time revealed a significant effect of Prime. Table 1 summarizes the regression results; the intercept represents the baseline to which the remaining levels of the Prime factor are compared. Thus, the estimate for the intercept provides the mean reaction time (ms) for the unrelated condition.

Overall, the mean reaction time for the four prime conditions was in order from fastest to slowest ST (619 ms), S (666 ms), UR (713 ms), and T (775 ms). The comparisons relative to the unrelated baseline condition (UR) revealed that the reaction time for the ST condition (primes matched targets in both segments and tone) was significantly faster than the unrelated UR baseline condition, indicating a significant facilitation (93 ms). There was also a marginally significant effect for the S condition (primes matched targets in segments only) as compared to the unrelated UR baseline condition, indicating that segment-only overlap (with tone mismatch) also produced strong facilitation (47 ms). The reaction time difference between T (primes matched targets in tone only) and the unrelated baseline (UR) condition was also significant, indicating that tone-only overlap (with segment mismatch) produced significant inhibition (62 ms).

As a second phase of the analysis, a mixed-effect logistic regression was performed for accuracy. Table 2 shows that the logistic mixed-effects model for accuracy reports a significant effect

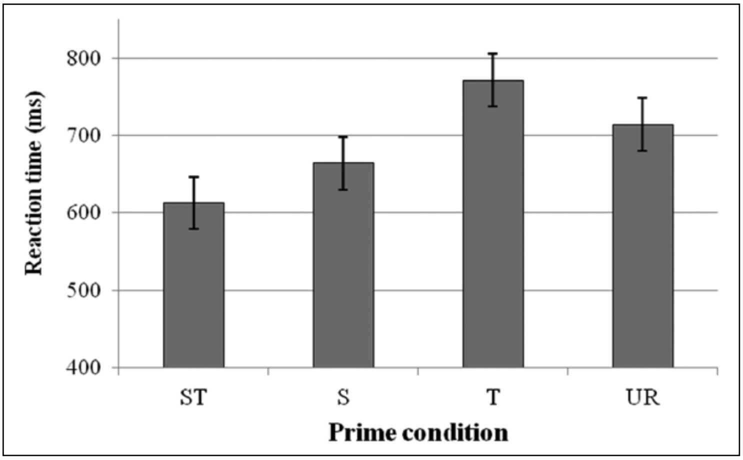

Figure 1. Mean reaction times (ms) for ST (matched in all segments and tone), S (matched in segments), T (matched in tone) and UR (unrelated) prime conditions. Error bars indicate standard errors.

Table 1. Fixed-effect coefficients in a mixed-effects model fitted to reaction times in Experiment 1.

		Estimate	Std. error	*t*-value	*p*MCMC
(Intercept: UR)		713.17	35.70	19.978	0.0001
Prime	ST	−93.16	24.51	−3.801	0.0001
	S	−46.77	24.52	−1.908	0.0538
	T	62.11	25.00	2.484	0.0132

Table 2. Fixed-effect coefficients in a logistic mixed-effects model fitted to accuracy in Experiment 1.

		Estimate	Std. error	z-value	Pr(>|z|)
(Intercept: UR)		3.6452	0.2820	12.925	0.0000
Prime	ST	0.3115	0.3142	0.991	0.3215
	S	0.5484	0.3346	1.639	0.1013
	T	−0.6565	0.2635	−2.491	0.0127

of Prime. The effect of Prime is due to the fact that while the ST and S condition did not show more errors than the unrelated condition, participants made significantly more incorrect responses for the tone only condition, with the error rates patterning ST = S = UR < T.

To further examine the contribution of mismatching tone, the reaction times and error rates for the segment-only condition were examined. While this analysis is based on a subset of the data where segments matched and tonal information mismatched, there is an equal distribution of prime–target tonal mismatches so comparisons can be made within the condition. As shown in Figure 2, each target tone is separately represented for each mismatching preceding tonal prime.

An examination of the mismatched tone condition showed differences across primes within each target tone. That is, different prime–target tone combinations resulted in different amounts of

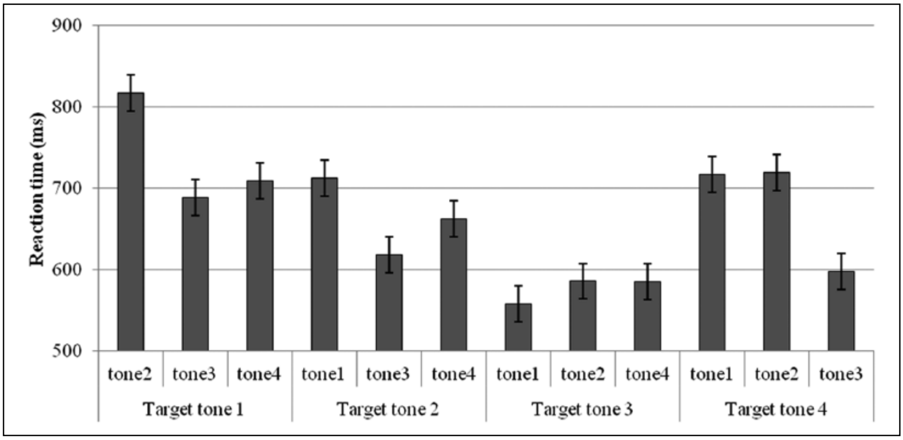

Figure 2. Mean reaction times (ms) for the S (matched in segments) condition. Reaction times are presented for each prime–target tone combination. Error bars indicate standard errors.

Table 3. Mean fundamental frequency onset and offset values (Hz) (standard deviations in parentheses) for each prime tone (Tones 1–4) and each target tone (Tones 1–4) in Experiment 1.

	Prime tone		Target tone	
	Onset	Offset	Onset	Offset
Tone 1	166 (10)	162 (9)	162 (18)	160 (4)
Tone 2	107 (7)	168 (10)	120 (15)	158 (8)
Tone 3	106 (5)	119 (9)	130 (16)	115 (7)
Tone 4	197 (10)	123 (23)	160 (20)	106 (23)

priming. Overall, when *offset* fundamental frequency values of the prime tone contrast with *onset* fundamental frequency values in the target tone (see Table 3), responses are likely to be speeded (642 ms) while when *offset* F0 values of the prime tone are similar to *onset* target tone F0 values, responses are slowed (687 ms). This pattern shows up as a marginally significant effect for Tone 1 and Tone 4 targets (RTs: $t(274)=1.684$, $p=.091$; errors $t(286)=1.773$, $p=.077$). That is, for a target Tone 1 (high onset), a preceding Tone 3 or Tone 4 prime (contrasting low offset) showed faster response times (699 ms) while a preceding Tone 2 prime (similar high offset) showed slower times (817 ms);for a target Tone 4 (high onset), a preceding Tone 3 prime (contrasting low offset) showed faster times (598 ms) while a preceding Tone 1 prime or Tone 2 prime (similar high offset) showed slower response times (719 ms). It is important to note that these analyses are based on a subset of the data only. As such, the trends observed here suggest that acoustic contrast between prime and target tone affects the amount of priming.

Overall, the analysis of reaction time and error data indicated that listeners were fastest when there was overlap in both segmental and tonal information as compared to an unrelated baseline condition and this facilitation effect was maintained, although to a lesser extent, when there was segment-only overlap. Moreover, listeners were slower and less accurate compared to the unrelated baseline when the overlap was only in tonal information, that is, with mismatching segmental cues. These data suggest important differences between mismatching segmental and mismatching tonal information. While listeners show inhibition with mismatching segmental information, suprasegmental mismatch resulted in facilitation relative to an unrelated baseline. Listeners are more sensitive to segmental mismatch than suprasegmental mismatch, with segmental mismatch being more detrimental in lexical access. Finally, in the segment-only overlap condition with tonal mismatch, there is a sizeable contribution of individual prime tones. The pattern of data that emerges suggests that the offset fundamental frequency value of the prime stimuli does affect processing of the following target tone, showing a contrastive effect.

3 Experiment 2

Experiment 2 examined further priming effects involving partial segmental overlap. Specifically, Experiment 2 investigated whether minimal segmental overlap would show facilitation when tonal information was matching and whether the facilitation would appear for both onset-only and rime-only segmental overlap.

Previous research investigating non-tonal languages has examined phonological priming effects in which prime and target overlap involves coincidence either at the beginning or at the end of words. Phonological priming has been used to explore issues such as the nature of lexical access and the lexical mechanisms involved in recognition. (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, McQueen, & Cutler, 2002). While there is some disagreement on the nature and locus of these phonological priming effects (see McQueen & Sereno, 2005), a number of consistent findings do emerge. Facilitation is consistently observed when primes and targets share an onset phoneme (for example, Goldinger, Luce, Pisoni, & Marcario, 1992; Hamburger & Slowiaczek, 1996; Radeau, Morais, & Dewier, 1989; Radeau, Morais, & Segui, 1995; Slowiaczek, Nussbaum, & Pisoni, 1987). While variation in relatedness proportion, task, and ISI all have been shown to contribute (for example, Goldinger, 1999; Pitt & Shoaf, 2002) and the amount of phonological overlap (that is, one-, two-, or three-phoneme overlap) can moderate the observed priming effect, it appears that participants use the expectation of a shared initial segment to benefit recognition of the target.

Furthermore, priming studies on offset phonological overlap have also consistently shown strong facilitation effects (for example, Burton, Jongman, & Sereno, 1996; Jakimik, Cole, & Rudnicky, 1985; Meyer, Schvaneveldt, & Ruddy, 1974; Slowiaczek, McQueen, Soltano, & Lynch, 2000). Increasing the amount of phonological overlap (one-, two-, or three-phoneme overlap) significantly increases the amount of the observed facilitation, but this increase seems to be mediated by whether the overlap constitutes a rime. Variation in relatedness proportion, task, and ISI have also been shown to influence the priming effect (Norris, McQueen, & Cutler, 2002) but to a much lesser extent than in onset priming. It appears that participants use the activation of a repeated rime to benefit recognition of the target.

While onset and offset phonological priming effects have been extensively examined in lexical stress languages, they have been less systematically investigated in tone languages. Experiment 2 extends the priming paradigm to contrast minimal segmental overlap in onset and offset position. Experiment 1 clearly showed superiority for segmental information in lexical activation, with facilitation found when segmental cues overlap even with mismatching tonal information. With no segmental overlap and only matching tonal information, no facilitation was observed relative to an unrelated control, suggesting that tonal information plays little role in lexical access. Given these results for segmental priming in Mandarin, one might expect that partial segmental overlap (for example, onset or rime phonological priming) in Mandarin would be similar to what has been found in lexical stress languages, with consistent facilitation found for both onset as well as rime offset priming. Experiment 2 specifically examined this research question.

3.1 Methods

3.1.1 Participants. Sixty-eight subjects (43 females, 25 males) participated in Experiment 2. The ages of the subjects ranged from 19 to 30 years, and all of the subjects were native speakers of Mandarin. All the subjects were students at the Chinese Academy of Sciences in Beijing. None of the subjects reported any speech disorders; and they were paid for their participation. None of them had participated in Experiment 1.

3.1.2 Materials. The target stimuli for Experiment 2 were the same 96 targets (48 words, 48 nonwords) used in Experiment 1. Each target was paired with four types of primes in Experiment 2. All prime and target word stimuli for Experiment 2 are listed in Appendix 2. Experiment 2 investigated whether minimal segmental overlap between prime and target such as onset-only or rime-only also triggers faster reaction time in a tone language. We included Onset+Tone and Rime+Tone prime conditions in addition to the segmental and tonal overlap condition (ST) and unrelated condition (UR). Each word target was thus paired with four types of primes: ST (matched in all segments and tone), Onset+Tone (matched in onset and tone), Rime+Tone (matched in rime and tone) and UR (unrelated). In the ST condition, the primes and targets were identical in both segmental structure and tone (for example, ru4–ru4). In the Onset+Tone condition, the primes and targets were matched in onset consonants and tones, but differed in their rime segments (for example, re4–ru4). In the Rime+Tone condition, the primes and targets were matched in their offset segmental rime structure and tones, where offset rimes consisted of either vowel with coda consonant or vowel only (for example, pu4–ru4). In the UR condition, the primes shared neither segmental structure nor tone with targets (for example, qin1–ru4).

The 48 word targets were distributed equally across each of the target tones. There were 12 word stimuli for each target tone (Tones 1, 2, 3 and 4). By balancing the number of targets for each target tone, we avoided having a particular target tone condition occur more than another, and we could control the contribution of each of the four target tones across the prime conditions.

As for the 48 nonword targets that functioned as fillers, T and UR primes were paired with each of the nonwords. Frequency counts of word stimuli analyzed based on Da Jun's (1998; 2004) corpus with 45 million words showed that the log frequencies of ST, Onset+Tone, Rime+Tone and UR primes were 2.11, 2.29, 2.18 and 2.15, respectively. There was no significant difference among the four types of primes in terms of frequency of occurrence (F (3, 45) = 0.850, p = 0.474, ns).

As in Experiment 1, four stimulus sets were constructed in order to avoid the repetition of prime–target pairs in Experiment 2. For a given target item, each of the four prime conditions was assigned to a different set such that each set would include an equal number of prime types without repeating any particular stimulus item in a set. Each of the four stimulus sets included all 96 targets (48 words, 48 nonwords); within each stimulus set, twelve prime words for each of ST, Onset+Tone, Rime+Tone, and UR conditions were paired with each of 48 word targets, and twelve primes from each of the conditions were paired with the 48 nonword targets. For an equal distribution of stimuli across participants, the 68 subjects were divided into four groups (17 subjects for each group), who were randomly assigned to a stimulus set. Thus, participants did not hear the same stimulus item more than once during the experiment.

3.1.3 Procedure. The procedure for Experiment 2 was identical to that used in Experiment 1. The ISI between the prime and target was 250 ms and the inter-trial interval between each pair was 3 seconds. Prior to the actual experiment, 8 trials were given to the subjects as a practice, to provide familiarization with the task. The experimental session took approximately 15 minutes.

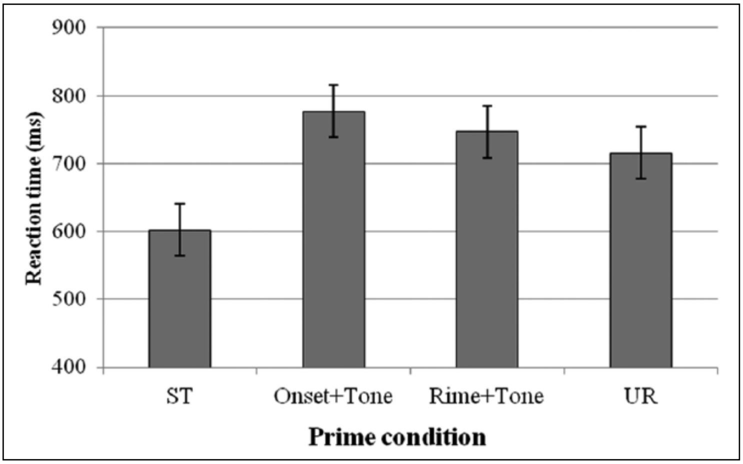

Figure 3. Mean reaction times (ms) by ST (matched in all segments and tone), Onset+Tone (matched in onset segment and tone), Rime+Tone (matched in rime segment and tone) and UR (unrelated) prime conditions. Error bars indicate standard errors.

Table 4. Fixed-effect coefficients in a mixed-effects model fitted to reaction times in Experiment 2.

		Estimate	Std. error	*t*-value	*p*MCMC
(Intercept: UR)		722.08	42.15	17.132	0.0001
Prime	ST	−106.75	22.55	−4.732	0.0001
	Onset+Tone	72.80	22.92	3.176	0.0020
	Rime+Tone	31.43	22.89	1.373	0.1742

Results

Similar to Experiment 1, both reaction time and accuracy were analyzed using *lmer*() function in R (version 2.15.3) from the *lme4* package (Bates & Maechler, 2009), including the fixed Prime variable with four levels (ST, Onset+Tone, Rime+Tone and UR) and two random effects, Subject and Item. For the Prime factor, the UR condition was always set as the baseline. The *p*-values reported in this paper were obtained through the Markov chain Monte Carlo method (MCMC), using the *pvals.fnc()* function of the *languageR* package. The RT data were trimmed such that reaction times above or below two standard deviations of each subject's mean (2 trials) as well as all errors (173 trials) were omitted from further analysis, which resulted in the exclusion of 5.4% of the data. Mean reaction times were calculated from correct responses only. Mean reaction times for each prime condition are shown in Figure 3.

The mixed-effects model for reaction time revealed a significant effect of Prime as summarized in Table 4.

Overall, the mean reaction time for the four prime conditions was from fastest to slowest in the order of ST (615 ms), UR (722 ms), Rime+Tone (754 ms), and Onset+Tone (795 ms). The comparisons relative to the unrelated baseline condition (UR) revealed that the reaction times for the ST condition (primes matched targets in both segments and tone) were 107 ms faster than for the UR baseline, indicating a significant facilitation effect for identity priming. In addition, reaction times for the Onset+Tone condition (primes matched targets only in onset segment and tone) were 73 ms slower than for the UR baseline, indicating a significant inhibition effect for partial onset

Table 5. Fixed-effect coefficients in a logistic mixed-effects model fitted to accuracy in Experiment 2.

		Estimate	Std. error	z-value	Pr(>|z|)
(Intercept: UR)		3.3222	0.2285	14.536	0.0001
Prime	ST	0.9652	0.2892	3.338	0.0008
	Onset+Tone	−0.1526	0.2186	−0.698	0.4849
	Rime+Tone	−0.2097	0.2160	−0.971	0.3315

segmental mismatch coupled with complete tonal overlap. However, such an inhibition was not observed for the comparison between the UR and the Rime+Tone condition (primes matched targets only in rime segment and tone), suggesting that the significant inhibition effect for the onset segment matched condition was compromised when the segmental overlap was in the rime.

According to the regression report in Table 4, the comparisons between ST and Onset+Tone and between ST and Rime+Tone reveals that the ST condition was 180 ms faster than Onset+Tone and 138 ms faster than Rime+Tone, respectively, suggesting that any mismatching segmental overlap is inhibitory. Lastly, to see the reaction time difference between the Onset+Tone condition and the Rime+Tone condition, another mixed-effect regression analysis in which the Rime+Tone was set as the baseline was performed. The comparison between the Onset+Tone condition and the Rime+Tone condition was near-significant by $\beta = -41$ ($p = 0.08$), indicating that primes matching in rime information were slightly more likely to trigger faster reaction time than primes with matching onset segments.

In addition to the reaction time analysis, Table 5 shows the analysis for accuracy. The mixed-effects logistic regression for accuracy showed a significant effect of Prime, which resulted from significantly less errors for the ST condition compared to the unrelated baseline condition. Accuracy rates, however, were not significantly different for the partial segmental overlap conditions, Tone+Onset and Tone+Rime, compared to the unrelated baseline condition.

Overall, the analysis of reaction time and error data indicated that while listeners were faster and more accurate when there was both complete segmental and tonal overlap, any partial mismatch in segmental information eliminated the facilitation, with listeners' reaction time most affected when there was mismatching rime segmental information.

4 Discussion

The current study investigated the nature and processing of tonal information in Mandarin spoken word recognition by directly evaluating the separate contributions of segments (consonants and vowels) and suprasegmentals (tone) to word recognition processes. In tone languages, lexical identity depends on these two types of information that seem qualitatively different. The question addressed by Experiment 1 was whether lexical tones are processed and accessed in a similar fashion to the processing of segmental information. Experiment 1 used a priming task to examine whether or not listeners were able to separately use segmental and tonal information to identify Mandarin words.

Overall, the analysis of Experiment 1 indicated that listeners were significantly faster when there was overlap in both segmental and tonal information as compared to an unrelated baseline condition. Moreover, this facilitation effect was maintained, although to a lesser extent, despite mismatch in tonal information. Experiment 1 clearly showed that segmental information functions to access the lexicon in tone languages, with a facilitation effect found when the prime and target words are identical in segmental structure. This facilitation effect was robust when both segmental and tonal information matched. When all segmental cues were still present but tonal information mismatched, the priming was still observed, but was significantly less than in the identity condition.

Interestingly, in the opposite mapping of segmental and tonal matching cues – namely, when there was only tonal overlap and segmental cues mismatched – listeners were significantly slower and less accurate. These data indicate significant differences between mismatching segmental and mismatching tonal information, suggesting that listeners are more sensitive to segmental mismatch than suprasegmental mismatch, with segmental mismatch being more detrimental to lexical access. When prime and target completely overlapped in segmental information, responses to targets showed facilitation while complete overlap in tonal information was inhibitory, showing significantly more errors than unrelated controls. Together, these data suggest that segmental information plays an important role in lexical access in Mandarin, showing facilitation of target responses when both matching consonantal and vocalic information is present. Tonal information, on the other hand, does not appear to facilitate lexical access, contributing only in a minor way by slowing responses when tonal information mismatches.

These data are compatible with previous perceptual results. Listeners of tonal languages are more sensitive to mismatch in segment than mismatch in tone. Participants, for example, were more likely to accept nonwords as real words when they only mismatched in tone (Cutler & Chen, 1997) and they responded more slowly and had more errors to tone-mismatching stimuli than to segmental-mismatching stimuli (Cutler & Chen, 1997; Ye & Connine, 1999). Moreover, in a recent study, Burnham et al. (2011) examined Thai stimuli which differed by tone, by phone, or by both in an odd-one-out design. Burnham et al. found that tonological awareness can be observed in children but develops more slowly than phonological awareness. This pattern is also observed across listener groups (Thai, Cantonese, and Australian English), suggesting a relatively early perceptual bias for phones over tones. These studies show that listeners are able to pick out mismatching segmental stimuli faster and more accurately than mismatching tone stimuli. Not only do listeners rely more on segmental information than tonal information, but segmental processing also seems to occur relatively earlier in time and in development than the processing of tonal information (Burnham et al., 2011; Ye & Connine, 1999).

In addition to comparing segmental and suprasegmental contributions to lexical access, an examination of the segment-only condition also revealed an interesting pattern of tone interactions. In Experiment 1, for the segment-only overlap condition, primes and targets were matched only in segmental structure but differed in their tones, with prime and target tones equally distributed to insure each particular prime–target tone combination occurred equally. Consequently, individual prime-tone combinations could be separately examined to evaluate the contribution to the overall facilitation when segmental cues matched while tonal information mismatched. For this condition with tonal mismatch, the observed trends suggest there is a contribution of prime tone across target tones. Namely, for each target, different prime–target tone combinations resulted in different amounts of priming. In general, the present data show that when offset fundamental frequency values of the prime tone are distinct from onset fundamental frequency values in the target tone, responses tend to be faster. Moreover, the opposite holds, that is, responses tend to be slower, when offset F0 values are similar to onset target F0 values. The pattern of data that emerges suggests that the offset fundamental frequency value of the prime stimuli does affect processing of the following target tone, showing a contrastive effect.

Two outcomes from this tonal analysis are noteworthy. First, variation in the amount of priming across distinct mismatching prime–target combinations for this segment–only condition may provide a possible explanation for the differences between the results of the current study as compared to those of C.-Y. Lee (2007). Using unequally distributed prime–target pairings, C.-Y. Lee (2007) did not find a significant facilitation for his overlapping segment condition. The present study suggests that C.-Y. Lee's inclusion of a greater number of *similar* prime–target tonal pairs could have resulted in less overall priming, which could account for C.-Y. Lee's earlier non-significant effect. With balanced prime–target pairing, the current study did find a marginally significant effect (p=.0538) for overlapping segmental cues in the context of mismatching tonal information. The present data clearly show that balanced prime–target tone pairing is critical to evaluating successfully the major contribution of segments and the reduced importance of tonal information to the priming effect in Mandarin word recognition.

An additional implication of the analysis of the mismatched tonal pairings is that there seems to be a lower-level acoustic-phonetic contribution to the segmental priming effect. Specifically, there was a trend for the occurrence of priming for prime–target pairs that contrasted in F0 offset and onset, respectively. Contrast effects in speech have been shown in the perception of a variety of distinctions, including vowel quality (Ladefoged & Broadbent, 1957) and speaker identity (Johnson, 1990). Contrast effects have also been reported in the perception of tone. For example, Moore and Jongman (1997) and Sereno, Lee, and Jongman (2011) tested the identification of stimuli along a Mandarin Tone 2 to Tone 3 continuum and found that the nature of the precursor sentence affected listeners' perception. Specifically, when hearing a precursor with a relatively small change in F0 and a relatively fast speaking rate, listeners were more likely to identify an ambiguous target stimulus as Tone 3. This is a contrast effect in that the small F0 range and fast speaking rate of the context make the turning point (the duration between tone onset and the lowest F0 point) and the ΔF0 (change in F0) of the tone seem relatively large and late, respectively, in the ambiguous target. A similar effect may occur in the segment-only (mismatching tones) condition in Experiment 1: a low F0 of the prime may make the onset F0 of the target seem higher. As a result, a Tone 1 target would be even more Tone 1-like when preceded by a low F0 offset. While segmental overlap between prime and targets resulted in facilitation, this effect was moderated by the nature of the tonal mismatch. A contrastive mapping of prime and targets, especially the fundamental frequency difference between the offset of the prime and the onset of the target, resulted in the fastest response times while a greater similarity of prime offset and target onset F0 values produced a slower response time. When offset fundamental frequency values of the prime tone contrasted with onset fundamental frequency values in the target tone (for example, a low offset prime tone 3 at 119 Hz precedes a high onset target tone 1 at 162 Hz), responses are speeded while the opposite pattern (for example, a high offset prime tone 2 at 168 Hz precedes a high onset target tone 1 at 162 Hz) results in slower response times. The nature of the tonal mismatch moderated the segmental priming, with distinct tonal information aiding when segmental cues overlapped.

Experiment 2 investigated further the contribution of segments and tones by examining the specific contribution of onset and offset segmental information to the priming effect. The relative scarcity of such processing studies in tone languages makes an investigation of such phonological effects in an on-line speeded priming paradigm in spoken Mandarin particularly significant. Most interesting for Mandarin is that neither onset overlap (single consonant) nor offset overlap (vowel alone or vowel plus coda consonant) produced any facilitation in Experiment 2. While complete overlap in both segmental and tonal cues results in facilitation, primes matching targets only in onset segment plus tone as well as primes matching targets only in offset segments plus tone (the entire rime) were significantly slower, with mismatched offsets more destructive to access processes than mismatched onset information. These data clearly demonstrate a surprising inhibition effect for any partial segmental mismatch even when there is complete tonal overlap. For Mandarin, any mismatching segmental overlap is inhibitory.

This pattern of data is in stark contrast to experiments examining non-tonal languages. Most previous research investigating phonological priming in non-tonal languages has consistently found facilitation effects with rime segmental overlap (for example, Burton et al. 1996; Jakimik et al. 1985; Meyer et al. 1974; Slowiaczek et al. 2000), with increasing amount of observed facilitation with increasing amount of overlap. Facilitation effects when primes and targets share an onset phoneme, though less consistent, have also been widely reported (for example, Goldinger et al. 1992; Hamburger & Slowiaczek, 1996; Hamburger & Slowiaczek, 1999; Radeau et al. 1995; Slowiaczek et al. 1987). While these effects are less robust, moderated by relatedness proportion, task, and ISI (for example, Goldinger, 1999; Pitt & Shoaf, 2002), listeners in these non-tonal languages do use the expectation of a shared initial segment to benefit recognition of the target. While previous phonological priming studies have been extensively conducted in languages that do not have phonemic tone, the current data suggest that the observed patterns may not necessarily extend to other languages, and specifically not when tone is lexically distinctive in the language. Our data suggest that for Mandarin, any partial mismatch in segmental information eliminated the facilitation, with listeners most affected (slowest and least accurate) when there was mismatching offset segmental information. Even with a complete tonal match between primes and targets, neither onset segmental overlap (without matching offset segments) nor offset segmental overlap (without matching onset segments) is effective in producing robust facilitation. Unlike non-tonal languages which consistently show robust rime priming, Mandarin shows no facilitation for partially overlapping prime–target pairs, even when matching tonal information is also present.

In conclusion, the current direct priming experiments clearly demonstrate that the contribution of tone is weaker than the contribution of segments in constraining word recognition in Mandarin. This evidence for primacy of segments over tones is in agreement with a majority of prior evidence using a range of tasks (for example, nonword identification, same-different judgments, as well as segmental/tone monitoring) showing that lexical tones seem to be processed separately from segments (Cutler & Chen, 1997; Taft & Chen, 1992; Ye & Connine,1999). While these studies have also found that these effects do vary with changing tasks and contextual conditions (Liu & Samuel, 2007; Mattys et al., 2005; Soto-Faraco et al., 2001; Xu, 1997; Xu, 1999; Ye & Connine, 1999), the greater importance of segmental information in comparison to suprasegmental cues is clear.

A second important implication of the present results is that the entire syllable plays an essential role in the perception of Mandarin. While a rime constituent or even a single onset segment can produce a benefit in processing in many non-tonal languages, neither type of overlap produces facilitation in a language such as Mandarin. In the data presented here, listeners seem to only benefit from advance knowledge of the full syllable. Our data show that syllabic overlap (both segments and tone) rather than partial segmental information alone was effective in producing facilitation. These findings seem to suggest that the syllable rather than the segment is the critical unit in Mandarin Chinese word recognition.

An account of Mandarin that includes the fact that tones are weaker than segments in constraining word recognition as well as the fact that rime segmental priming, even when coupled with tonal match, did not produce facilitation, must embrace the unique characteristics of tonal languages in general and Mandarin specifically. Languages exploit contrasting vowels and consonants to distinguish meaning while tonal languages additionally use tone to distinguish word meaning. Tone languages constitute an estimated 70% of the world's languages (Yip, 2002) (Maddieson's 2011 survey shows 43% are tone languages, but concedes this underrepresents the number of tonal languages in the world due to sampling bias). In tone languages, tone consists primarily of level or contour pitch variation used to distinguish meaning. Due to a restricted number of tones, each tone is often associated with many words, significantly more than words are associated with specific consonants and vowels. As a consequence, segmental cues are more critical than tonal cues in constraining word identity. Moreover, Mandarin, as contrasted to other tonal languages, brings additional unique characteristics. In his analysis of 207 tone languages, Maddieson (1978) found that Mandarin is among the 19% which has a complex tonal system, with use of four or more tones (most use only two or three tones) and the use of complex tones (for example, Tone 3 which is a falling then rising tone). As Maddieson (2011) notes, languages with tone, specifically complex tone, are highly likely to have more simple syllable structures.

The syllable as an important unit makes sense for Mandarin considering its other characteristics. Mandarin syllables are not complex but quite simple in structure, with Mandarin syllables less numerous and variable than those of other languages. All syllables have single-consonant onsets and end in a vowel or a nasal consonant. Consequently, there is a relatively restricted number of syllable types. Mandarin also has clear syllable boundaries, little resyllabification, and no ambisyllabicity (Duanmu, 2000). Interestingly, Mandarin orthography is also based on syllable-size characters. These properties of Mandarin make the syllable an ideal unit for lexical processing.

The current data emphasizing syllabic priority over segmental cues are in line with three other distinct types of data examining Mandarin perception and production. First, Tong et al. (2008), using a Garner speeded classification task where listeners attend to one dimension while ignoring the variation along another, found an asymmetric interference effect for consonantal, vocalic and tonal dimensions. Their data indicated that the vowel is pivotal in Chinese speech processing in comparison to consonants and tones. Tong et al. (2008) concluded that variation in segmental dimensions generates greater interference to a suprasegmental dimension than the reverse. Second, Mok (2009) also recently provided data supporting a syllabic basis for Mandarin processing using the recently developed acoustic measures of speech that typically classify languages based on their suprasegmental and rhythmic characteristics into stress-timed languages, syllable-timed languages, and mora-timed languages (Ramus et al., 1999). Mok (2009) investigated the consonantal and vocalic variability in Beijing Mandarin speech. Her data suggest that Mandarin falls into the syllable-timed category. Mok (2009) concluded that these rhythmic and phonological characteristics shape the nature of processing at the word level and, for Mandarin, that processing is syllable-based. Finally, recent data examining word production makes similar claims. Using an implicit priming task, Chen and colleagues (Chen, Chen, & Dell, 2002; Chen, Lin, & Ferrand, 2003; O'Seaghdha, Chen, & Chen, 2010) demonstrated that the syllable is the proximate unit (O'Seaghdha, Chen, & Chen, 2010), the first unit selected for word form encoding in Mandarin word production (Chen & Chen, 2013). They also showed that the syllable (onset plus rime) by itself produced implicit priming effects, whereas a tone-alone prime did not. These data also suggest that in Mandarin more emphasis is placed on the syllable rather than the segment. These data, using quite distinct methodologies in both production and perception tasks, demonstrate a syllabic-level processing preference in Mandarin.

Taken together, an important conclusion that arises from the current study is that lexical tones are processed differently from segments, with segments playing a major role and tones a secondary one. We can further conclude that the overall architecture of the processing system that emerges from these priming studies suggests that the syllable is an important unit of lexical access in Mandarin.

Acknowledgements

The authors would like to express thanks to Tim Bunnell for acting as editor for this article. We would also like to thank Allard Jongman, Jie Zhang, Anne Cutler, Tim Bunnell and an anonymous reviewer for their helpful comments and critique. We are grateful to Bin Liu and Jiang Liu who helped us collect the data in Beijing, China. An earlier version of this work was presented at the 164th meeting of the Acoustical Society of America in Kansas City, Missouri. Part of the manuscript was written while the first author was on University of Kansas sabbatical leave at the MARCS Institute (Sydney, Australia).

Funding

This research is partially supported by the National Science Foundation [grant number 084680].

References

Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.99375-42.

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.38). [Computer program] http://www.praat.org/.

Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknect, C., Kasisopa, B., & Luksaneeyanawin, S. (2011). Are tones phones? *Journal of Experimental Child Psychology*, *108*, 693–712.

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (Language Learning and Language Teaching series) (pp. 259–280). Amsterdam, The Netherlands: John Benjamins.

Burton, M., Jongman, A., & Sereno, J. (1996). *Phonological and orthographic priming effects in auditory and visual word recognition*, In W. Ham & L. Lavoie (Eds.), *Working papers of the Cornell Phonetics Laboratory*, 11, 17–41.

Chen, T.-M., & Chen, J.-Y. (2013). The syllable as the proximate unit in Mandarin Chinese word production: An intrinsic or accidental property of the production system? *Psychonomic Bulletin & Review*, *20*, 154–162.

Chen, H.-C., & Cutler, A. (1997). *Auditory priming in spoken and printed word recognition*. In H.-C. Chen (Ed.), *Cognitive processing of Chinese and related Asian languages* (pp. 77–81). Hong Kong, China: Chinese University Press.

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language*, *46*, 751–781.

Chen, J.-Y., Lin, W.-C., & Ferrand, L. (2003). Masked priming of the syllable in Mandarin Chinese syllable production. *Chinese Journal of Psychology*, *45*, 107–120.

Cutler, A., & Chen, H.-C. (1995). Phonological similarity effects in Cantonese word recognition. In K. Elenius, & P. Branderud (Eds.), *Proceedings of the Thirteenth International Congress of Phonetic Sciences ICPhS 1995*, Vol. 1 (pp. 106–109). Stockholm, Sweden: Stockholm University Press.

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics*, *59*, 165–179.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken word recognition in Japanese. *Journal of the Acoustical Society of America*, *105*, 1877–1888.

Da, J. (1998). Chinese text computing. Online version. *Retrieved from:* <http://lingua.mtsu.edu/Chinese-computing/index.html>.

Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction: The studies on the theory and methodology of the digitized Chinese teaching to foreigners. In Zhang Pu, Tianwei Xie & Juan Xu (Eds.), *Proceedings of the 4th International Conference on New Technologies in Teaching and Learning Chinese* (pp. 501–511). Beijing, China: The Tsinghua University Press.

Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford, UK: Oxford University Press.

Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, *27*, 765–768.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–175.

Goldinger, S. D. (1999). Only the shadower knows: Comment on Hamburger and Slowiaczek (1996). *Psychonomic Bulletin & Review*, *6*, 347–351.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1211–1238.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge, UK: University Press.

Hamburger, M. B., & Slowiaczek, L. M. (1999). On the role of bias in dissociated phonological priming effects: A reply to Goldinger (1999). *Psychonomic Bulletin & Review*, *6*, 352–355.

Hamburger, M., & Slowiaczek, L. M. (1996). Phonological priming reflects lexical competition. *Psychonomic Bulletin & Review*, *3*, 520–525.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago, IL: The University of Chicago Press.

Johnson, K. (1990). The role of perceived speaker identity in F0 normalization of vowels. *Journal of the Acoustical Society of America*, *88*, 642–654.

Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and spelling in spoken word recognition. *Journal of Memory and Language*, *24*, 165–178.

Jongman, A., Wang, Y., Moore, C., & Sereno, J. A. (2006). Perception and production of Mandarin tone. In E. Bates, L. Tan, & O. Tzeng (Eds.), *Handbook of East Asian psycholinguistics* (pp. 209–217). Cambridge, UK: Cambridge University Press,

Kiriloff, C. (1969). On the auditory discrimination of tones in Mandarin. *Phonetica*, *20*, 63–67.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, *29*, 98–104.

Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, *50*, 101–123.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, *32*, 451–454.

Liu, S., & Samuel, A. (2007). The role of Mandarin tones in lexical access under different contextual conditions. *Language and Cognitive Processes*, *22*, 566–594.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*, 1–36.

Maddieson, Ian. (1978). Universals of tone. In J. H Greenberg, et al. (Eds.), *Universals of Human Language; Volume 2: Phonology* (pp. 335–363). Stanford, CA: Stanford University Press.

Maddieson, I. (2011). Tone. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online* (Chapter 13), available online at http://wals.info/chapter/13. Munich, Germany: Max Planck Digital Library.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 1–86). Cambridge, MA: MIT Press.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*, 477–500.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics*, *2*, 148–154.

McQueen, J. M., & Sereno, J. (2005). Cleaving automatic processes from strategic biases in phonological priming. *Memory & Cognition*, *33*, 1185–1209.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. *Memory & Cognition*, *2*, 309–321.

Moore, C.B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, *102*, 1864–1877.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences*, *23*, 299–370.

Norris, D., McQueen, J. M., & Cutler, A. (2002). Bias effects in facilitatory phonological priming. *Memory & Cognition*, *30*, 399–411.

O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*, 282–302.

Pitt, M. A., & Shoaf, L. (2002). Revisiting bias effects in word-initial phonological priming. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1120–1130.

Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition*, *17*, 525–535.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.

Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1297–1311.

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

Sereno, J. A. (1994). *Phonosyntactics*. In L. Hinton, J. Nichols, & J. Ohala (Eds.), *Sound Symbolism*. Cambridge, UK: Cambridge University Press, 263–275.

Sereno, J. A., & Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech*, *38*, 57–76.

Sereno, J. A., Lee, H., & Jongman, A. (2011). Perceptual accommodation to rate, talker, and context in Mandarin. *Journal of the Acoustical Society of America*, *130*, 2445.

Sereno, J. A., & Wang, Y. (2007). Behavioral and cortical effects of learning a second language: The acquisition of tone. In O. Bohn. & M. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 239–258). New York, NY: John Benjamins Publishing Company.

Slowiaczek, L. M., McQueen, J. M., Soltano, E. G., & Lynch, M. (2000). Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language*, *43*, 530–560.

Slowiaczek, L. M., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition*, *14*, 230–237.

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *13*, 64–75.

Slowiaczek, L. M., & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology, Learning, Memory & Cognition*, *18*, 1239–1250.

Soto-Faraco, S., Sebastian-Galles, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.

Taft, M., & Chen, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H.-C. Chen & O.J.L. Tzeng (Eds.), *Language processing in Chinese* (pp.151–172). Amsterdam, The Netherlands: Elsevier.

Tagliaferri, B. (2008). 'Paradigm: Perception Research Systems [Computer Program]'.Retrieved from http://www.paradigmexperiments.com/.

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, *23*, 698–708.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, *27*, 55–105.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: the role of tone information. *Language and Cognitive Processes*, *14*, 609–630.

Yip, M. (2001). Phonological priming in Cantonese spoken-word processing. *Psychologia*, *44*, 223–229.

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

Yip, M., Leung, P.-Y., & Chen, H.-C. (1998). Phonological similarity effects on Cantonese spoken-word processing. In R. A. Mannell & J. Robert-Ribes (Eds.), *Proceedings of the International Conference on Spoken Language Processing, ICSLP 1998* (pp. 2139–2142). Sydney, Australia: International Speech Communication Association (ISCA).

Appendix 1. Experiment 1 stimuli.

Targets		Primes							
word		ST		S		T		UR	
bi	1	bi	1	bi	2	suan	1	han	3
bo	1	bo	1	bo	2	zhua	1	man	3
guo	1	guo	1	guo	2	can	1	si	3
sui	1	sui	1	sui	2	ca	1	lan	3
huan	1	huan	1	huan	3	jie	1	fo	2
tao	1	tao	1	tao	3	jiu	1	run	4
tie	1	tie	1	tie	3	san	1	qun	2
tui	1	tui	1	tui	3	sang	1	fen	2
chan	1	chan	1	chan	4	gu	1	cuo	4
heng	1	heng	1	heng	4	xiao	1	liao	2
zang	1	zang	1	zang	4	bei	1	kuo	4
zeng	1	zeng	1	zeng	4	gua	1	ruo	4
hong	2	hong	2	hong	1	ze	2	lue:	4
kang	2	kang	2	kang	1	su	2	chui	1
lu	2	lu	2	lu	1	pang	2	cang	1
pa	2	pa	2	pa	1	xun	2	kui	4
lou	2	lou	2	lou	3	xiang	2	jiang	1
cheng	2	cheng	2	cheng	3	xi	2	dai	1
rao	2	rao	2	rao	3	xu	2	geng	4
zhe	2	zhe	2	zhe	3	nin	2	shan	4
niang	2	niang	2	niang	4	mou	2	duo	3
pi	2	pi	2	pi	4	chong	2	kuan	3
tu	2	tu	2	tu	4	meng	2	xing	3
xia	2	xia	2	xia	4	hou	2	zhen	3
biao	3	biao	3	biao	1	ren	3	cong	1
pao	3	pao	3	pao	1	zen	3	jun4	4
qian	3	qian	3	qian	1	lu:	3	le	4
zao	3	zao	3	zao	1	qing	3	die	1
cao	3	cao	3	cao	2	leng	3	sun	1
chuang	3	chuang	3	chuang	2	di	3	shi	1
du	3	du	3	du	2	lao	3	hen	4
zuo	3	zuo	3	zuo	2	gai	3	te	4
guang	3	guang	3	guang	4	ji	3	lei	2
kan	3	kan	3	kan	4	niu	3	liu	2
lian	3	lian	3	lian	4	fou	3	de	2
shuai	3	shuai	3	shuai	4	ken	3	ceng	2
cha	4	cha	4	cha	1	nong	4	hui	3
dun	4	dun	4	dun	1	ta	4	ma	3
fan	4	fan	4	fan	1	shou	4	nu:	3
shai	4	shai	4	shai	1	hun	4	long	3
cun	4	cun	4	cun	2	gao	4	fa	2
mai	4	mai	4	mai	2	ku	4	lun	2
mi	4	mi	4	mi	2	fang	4	kua	1
nao	4	nao	4	nao	2	pei	4	ling	2
nu	4	nu	4	nu	3	zha	4	mang	2
quan	4	quan	4	quan	3	lie	4	xie	1
ru	4	ru	4	ru	3	sha	4	qin	1
tong	4	tong	4	tong	3	se	4	diu	1

Appendix 2. Experiment 2 stimuli.

Targets		Primes							
word		ST		T_Onset		T_Rime		UR	
bi	1	bi	1	bao	1	qi	1	han	3
bo	1	bo	1	bin	1	po	1	man	3
guo	1	guo	1	gan	1	tuo	1	si	3
sui	1	sui	1	sen	1	zhui	1	lan	3
huan	1	huan	1	hei	1	chuan	1	fo	2
tao	1	tao	1	ting	1	sao	1	run	4
tie	1	tie	1	tun	1	bie	1	qun	2
tui	1	tui	1	tan	1	gui	1	fen	2
chan	1	chan	1	che	1	ban	1	cuo	4
heng	1	heng	1	hu	1	feng	1	liao	2
zang	1	zang	1	zu	1	gang	1	kuo	4
zeng	1	zeng	1	zai	1	deng	1	ruo	4
hong	2	hong	2	huai	2	rong	2	lue:	4
kang	2	kang	2	ke	2	tang	2	chui	1
lu	2	lu	2	lin	2	fu	2	cang	1
pa	2	pa	2	ping	2	na	2	kui	4
lou	2	lou	2	lang	2	chou	2	jiang	1
cheng	2	cheng	2	chai	2	beng	2	dai	1
rao	2	rao	2	reng	2	mai	2	geng	4
zhe	2	zhe	2	zhu	2	ge	2	shan	4
niang	2	niang	2	nuo	2	liang	2	duo	3
pi	2	pi	2	peng	2	li	2	kuan	3
tu	2	tu	2	tai	2	chu	2	xing	3
xia	2	xia	2	xue	2	jia	2	zhen	3
biao	3	biao	3	bu	3	miao	3	cong	1
pao	3	pao	3	pin	3	chao	3	jun4	4
qian	3	qian	3	qu	3	xian	3	le	4
zao	3	zao	3	zui	3	kao	3	die	1
cao	3	cao	3	ci	3	dao	3	sun	1
chuang	3	chuang	3	chi	3	huang	3	shi	1
du	3	du	3	dang	3	mu	3	hen	4
zuo	3	zuo	3	zi	3	huo	3	te	4
guang	3	guang	3	gei	3	shuang	3	lei	2
kan	3	kan	3	kou	3	dan	3	liu	2
lian	3	lian	3	luo	3	nian	3	de	2
shuai	3	shuai	3	sheng	3	guai	3	ceng	2
cha	4	cha	4	chen	4	la	4	hui	3
dun	4	dun	4	diao	4	gun	4	ma	3
fan	4	fan	4	fei	4	pan	4	nu:	3
shai	4	shai	4	shun	4	pai	4	long	3
cun	4	cun	4	cai	4	kun	4	fa	2
mai	4	mai	4	mo	4	sai	4	lun	2
mi	4	mi	4	men	4	ni	4	kua	1
nao	4	nao	4	nie	4	hao	4	ling	2
nu	4	nu	4	nai	4	cu	4	mang	2
quan	4	quan	4	qie	4	xuan	4	xie	1
ru	4	ru	4	re	4	pu	4	qin	1
tong	4	tong	4	ti	4	zong	4	diu	1