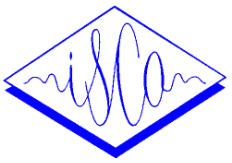

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

Lexical Tone in the Spoken Word Recognition of Chinese

Xiaolin Zhou and **Yanxuan Qu**
Department of Psychology
Peking University
Beijing 100871, China
xz104@pku.edu.cn

Abstract

Constraints of lexical tones on semantic activation in spoken word recognition of Chinese were investigated in three cross-modal priming lexical decision experiments. In Experiments 1 and 2, disyllabic compound words that shared the same segmental templates but differed in lexical tones (e.g., tiao4 yue4 vs. tiao2 yue1, *jump* vs. *treaty*, numbers indicating tone types) were used as auditory primes while words that were semantically related to one of the pairs (e.g., ben1 pao3, *run*) were visually presented for lexical decision. The semantic primes and the tone-mismatch primes differed in the tones of either the first, the second, or both syllables. In Experiment 3, nonword tone-mismatch primes were created by changing the first or the second tones of semantic primes. The similarity between the original tones and the resulting tones was also manipulated. The appearance of significant priming effects for the tone-different primes depended on lexical competition environment, the global and local goodness of fit between input tones and underlying tones, and the constituent position of tone alternation. The findings are discussed in terms of how tonal information is represented in the mental lexicon, how tonal information in speech input is mapped onto the lexicon, and how tonal constraints on lexical access and semantic activation is influenced by competition environment.

Introduction

Chinese is a tonal language in which tones, i.e., the pitches carried by the vocalic part of syllables, are used to differentiate lexical items. Because of the simplicity of syllabic structure in Mandarin Chinese, there are only about 400 different syllables in the language, representing over 5000 commonly used morphemes. However, due to the use of lexical tones, the number of syllables is increased to about 1300, reducing the extent of homophony in the language. Mandarin Chinese has four different tones: high-level (Tone 1), high-rising (Tone 2), low-dipping (Tone 3), and high-falling (Tone 4). These tones can be attached to the same segmental templates (e.g., "yi"), although not all the segmental templates are accompanied by all the four tones. Tone is a much more useful device in differentiating lexical items in Chinese than other suprasegmental devices, say, stress, in English (Cutler, 1986).

The role of tonal information in phonetic and perceptual processing e.g., Shen & Lin, 1991; Walen & Xu, 1992) and in spoken and visual word recognition of Chinese (e.g., Cutler & Chen, 1997; Ye & Connine, 1999; Zhou, 2000; Zhou & Marslen-Wilson, 1997, 2000) has become a prominent issue in recent years. However, these studies in general did not provide clear answers to the questions of how tonal information is represented in the mental lexicon, how tonal information in speech input is mapped onto underlying lexical representations, to what extent tonal information is used to constrain lexical semantic activation, and how lexical competition environment influences the processing of tonal information. The present study is to systematically address these questions by using a cross-modal semantic priming paradigm. In this paradigm, a disyllabic compound word is presented auditorily as a priming word, while a word semantically related to the prime is presented visually for lexical decision. The speech input of the prime maps onto its lexical phonological representation, whose activation passes to the semantic representation of the prime. Lexical processing of the subsequently presented visual target should be facilitated by the auditory prime, which shares many semantic properties with the target. By measuring the responses to the visual probe can we gain knowledge about how tonal information in speech input is mapped onto underlying representations and used to constrain their activation, and about how tonal information is represented in the lexicon.

Crucially for the present purposes, a visual probe (e.g., "ben1 pao3", *run*) is preceded not only by a semantically related auditory prime (e.g., "tiao4 yue4", *jump*), but also by an auditory prime (e.g., "tiao2 yue1", *treaty*) that shares the same segmental template with the semantic prime but differs in tones. At the segmental level, the speech input of both "tiao4 yue4" and "tiao2 yue1" activates the two words' phonological representations in the lexicon. However, the tonal information in the speech input of one word will create mismatch for the lexical representation of another word. That is, although the speech input of "tiao2 yue1" activates the segmental phonological representation of "tiao4 yue4", the tonal information in the speech input does not activate the suprasegmental or tonal phonological representation of "tiao4 yue4". If tonal information is used immediately to constrain semantic activation, this tonal mismatch between speech input and the underlying phonological representation may cancel or even inhibit the activation of segmental information, leading to no semantic activation of "tiao4 yue4". However, if tonal information, compared with segmental information, plays only a secondary role in the processing of spoken Chinese, the mismatch of tonal information may not completely block or prevent the semantic activation of the base word (i.e., semantic prime). The response to the visual probe may still be facilitated to some extent by the tone-mismatch prime.

In Experiments 1 and 2, we ask whether mismatch of one or two tones in the speech input of disyllabic primes will lead to no or partial semantic activation of their base words. The tone-mismatch primes here are real words themselves and have their own representations in the lexicon. The activation of these representations by the speech input of mediated primes may create competition for the activation of base words, reducing semantic priming for visual probes. In Experiment 3, we use nonwords differing only in the first or the second tones from base words as tone-mismatch primes. These nonwords do not have their own representations in the lexicon and hence may not create competition for lexical activation of the base words.

We also ask whether mismatch in the first or the second tone of the disyllabic speech input will have different consequences on the semantic activation of base words. This question is related to the issue of whether the first syllable of speech input plays a fundamental role in spoken word recognition (Connine, Blasko & Titone, 1993; Connine, Blasko & Wang, 1994; Marslen-Wilson, 1987; Marslen-Wilson & Zwisterlood, 1989). The Cohort model proposed by Marslen-Wilson and his colleagues suggests that lexical representations activated by the initial speech input forms a cohort of lexical candidates. As more speech input comes in, candidates that still match with the input are kept in the cohort and candidates that no longer match the input is deactivated and dropped out of cohort. A word is recognized when only one word is left in the cohort. This view is in contrast with other views that stress the importance of global goodness of matching, rather than the strict sequential matching, between speech input and underlying representations. According to the latter views, a word can still be a lexical candidate and recognized if speech input does not deviate much from its lexical phonological representation, whether the mismatch is at the initial, the middle, or the end of speech input. The word-initial matching has not special status in spoken word recognition. These contradictory views will be examined in this study by mismatching the first or second tones in speech input with underlying representations of base words.

We also address the question of whether tonal information is represented in the lexicon in terms of categorical tone patterns or in terms of fine-grained tonal features. According to some phonological theories (e.g., Guo, 1993), tones in Mandarin Chinese can be represented in terms of tonal features, such as tone contour and their relative height of pitch. Some tones are more similar to each other than other tones because they share more features. Using a phonological task in which subjects had to judge whether a pair of consecutively presented characters were homophones (disregarding their tones), Zhou and Marslen-Wilson (2000) found that characters having similar tones were responded to faster that characters having dissimilar tones. Since characters do not carry tonal information in their visual form, this effect reflects the activation of tonal information, which is represented in terms of features in the lexicon. In this study, we manipulate the similarity between tones in the speech input of mediated primes and the corresponding tone representations of base words. We hope to collect converging evidence about whether tonal information is represented in the lexicon in terms of fine-grained features.

Experiment 1

Experiment 1 examined whether disyllabic speech input (e.g., "tiao2 yue1") would be able to activate the underlying representation of a base word deviating in both tones (e.g., "tiao4 yue4", *jump*), when the speech input itself corresponds to another word (e.g., "tiao2 yue1", *treaty*). If tones plays an important and immediate role in constraining lexical processing of Chinese, the speech input should activate its own lexical representation but should not activate the semantic representation of the tone mismatching base word. Consequently we should not observe a priming effect for the visual probe (e.g., "ben1 pao3", *run*) when it was preceded by the mismatch prime. If, on the other hand, tones have no strong or immediate effect on lexical access, words differing in tones are virtually homophones and speech input of both "tiao2 yue1" and "tiao4 yue4" should activate the semantic representation of "tiao4 yue4" (*jump*), leading to facilitation in response to the target word "ben1 pao3" (*run*).

The experimental design and sample stimuli are presented in Table 1. Three groups of stimuli were selected according to the relative frequency between semantic primes (base words) and tone-mismatch primes. Unrelated control primes were created by re-pairing mediated primes with targets. Each group had 30 triplets of primes. Both tone-mismatch primes and controls had no phonological or semantic relations with the targets. The critical stimuli were assigned to three counter-balanced test versions and unrelated filler word-word and word-nonword pairs were then added. The two-character nonword targets were created by combining characters that were not used in the critical stimuli. Although characters used in nonwords were themselves meaningful morphemes, they did not form interpretable meanings for the nonwords. Subjects were asked to listen carefully to the auditory words presented through headphones and make timed lexical decisions to visual probes that were presented at the offset of speech. Occasionally subjects were given time to write down what they just heard. Mean reaction times and error percentages of their responses are presented in Table 2.

Table 1 Experiment 1: Design and Sample Stimuli

Stimulus Group	Auditory Priming Word				
	Base Word	Tone Mismatch	Control	Visual	Target

Base Word Dominant		tiao4 yue4 **跳跃**	tiao2 yue1 **条约**	ban1 fa1 **颁发**	ben1 pao3 **奔跑**
	Frequency	133	6	6	111
Equal Frequency		Jiu3 jing1 **酒精**	jiu1 jing4 **究竟**	bian1 jie4 **边界**	ran2 liao4 **燃料**
	Frequency	14	12	12	54
Mismatch Dominant		xian4 jin1 **现金**	xian1 jin4 **先进**	ling2 huo2 **估计**	zhi1 piao4 **支票**
	Frequency	15	72	72	24

Table 2 Experiment 1: Mean Reaction Times (ms) and Error Percentages (in parentheses)

Stimulus Group	Priming Conditions		
	Base Word	Tone Mismatch	Control
Base Word Dominant	499 (0.5)	527 (0.5)	527 (1.9)
Equal Frequency	527 (2.4)	564 (2.9)	554 (1.9)
Mismatch Dominant	512 (1.1)	540 (2.1)	539 (4.2)
TOTAL	512 (1.3)	544 (1.8)	540 (2.6)

Statistical analyses revealed a significant main effect of facilitation (28 ms) for semantic primes and no effect (-4 ms) for tone-mismatch primes. There was no interaction between priming condition and stimulus group, indicating the relative frequency of the semantic primes and tone-mismatch primes had no influence on the pattern of priming effects. There were no significant effects in the analyses of error rates. Thus, the processing of visual targets was facilitated when they were preceded by their auditory semantic primes, but not by words that differed from the semantic primes in tones. The absence of a significant priming effect for tone-mismatch primes indicated that tonal information carried in their speech input was used immediately to constrain lexical access and semantic activation. Although the segmental information in the speech input of mediated primes could activate the underlying phonological representation of base words (semantic primes), the mismatch in tones was sufficient to cancel this activation. It was also possible that the activation of lexical representations of tone-mismatch primes competed with and exerted inhibition on the activation of semantic primes.

Experiment 2

Experiment 2 continued to investigate the effect of tonal information in constraining lexical access and semantic activation. While the speech input of mediated primes in Experiment 1 deviated from the underlying phonological representations of base words in both tones, the mismatch was reduced to one tone in this experiment. This should increase the global goodness of fit between speech input and underlying phonological representations of base words and hence increase the probability of the semantic representations of base words being activated. Experiment 2 also investigated whether the earliness of tone mismatch plays an important role in constraining activation of underlying representations. This was carried out by using two groups of stimuli, one with tone-mismatch primes matching the first tones with but differing in the second tones from semantic primes, and one with tone-mismatch primes differing in the first tones and matching the second tones.

The experimental design and sample stimuli are presented in Table 3. The experimental procedures were the same as Experiment 1. Mean reaction times and error percentages are shown in Table 4. For reaction time, statistical analyses revealed a significant facilitatory effect (27 ms) for semantic primes and no effect for tone-mismatch primes (2 ms). However, the analyses of error rates found significant effects for both semantic and tone-mismatch primes. There was no interaction between priming condition and stimulus group. Thus, although not shown in reaction times, the processing of visual targets was facilitated by tone-mismatch primes, indicating that lexical representations of base words were activated to some extent by the speech input of tone-mismatch primes differing in one tone. Given the absence of such an effect in Experiment 1 in which both tones were mismatched, this experiment demonstrated that the global goodness of fit between speech input and underlying representation is an important factor in constraining lexical access and semantic activation.

Table 3 Experiment 2: Design and Sample Stimuli

Stimulus Group		Auditory Priming Word			Visual Probe
		Semantic	Mediated	Control	
Early Mismatch		wen1 he2 **温和**	wen3 he2 **吻合**	qu4 shi4 **去世**	yan2 li4 **严厉**
	Frequency	83	42	42	29

Later Mismatch		qing1 xin1 **清新**	qing1 xin4 **轻信**	shi2 ji1 **时机**	wu1 zhuo2 **污浊**
	Frequency	40	94	94	40

Table 4 Experiment 2: Mean Reaction Time (ms) and Error Percentage (in parentheses)

	Auditory Priming Condition		
	Base Word	Tone Mismatch	Control
Early Mismatch	594 (1.7)	616 (1.7)	619 (3.3)
Later Mismatch	599 (1.7)	628 (3.7)	628 (6.7)
TOTAL	597 (1.7)	622 (2.7)	624 (5.0)

Experiment 3

Experiment 3 investigated several factors and their interactions that could have important influences on the efficiency of tonal information in constraining lexical access and semantic activation. The first factor was competition environment. In the previous two experiments, speech input of tone-mismatch primes mapped not only onto lexical representations of their base words (semantic primes), but also onto lexical representations of tone-mismatch primes themselves. The activation of mismatch primes created competition with the activation of base words. In this experiment, we used speech input that did not correspond to any lexical representations in the lexicon. That is, we used nonwords that were created from semantic primes by altering their first or second tones. The second factor was the similarity between tonal information in speech input and tones in underlying representations. The priming effects for tone-mismatch primes in Experiments 1 and 2 demonstrated the importance of the goodness of fit between speech input and underlying representations. In this experiment we went further to manipulate the degree of similarity between mismatching tones and underlying tones. According to phonological theories of lexical tone (e.g., Guo, 1993), some tones, like Tone 2 and Tone 3, shares more tonal features than other tones. If tonal information in speech input is mapped onto lexical representations in terms of tonal features and if lexical phonological representations are also in terms of tonal features, the activation of these representations should be influenced by the degree of tonal similarity. Consequently the priming effect for tone-mismatch primes should also vary as a function of this similarity. The third factor we examined was, as in Experiment 2, the constituent position of tone mismatch. The speech input of mismatch primes differed from their base words (semantic primes) either in the first tones or the second tones.

Table 5 Experiment 3: Design and Sample Stimuli

Stimulus Group	Auditory Priming Word					Visual Probe
	Word Prime		Nonword Prime			
	Semantic	Semantic Control	High Similarity	Low Similarity	Nonword Control	
Early Mismatch	yi2 han4 **遗憾**	hu2 die2 **蝴蝶**	yi3 han4 **椅憾**	yi4 han4 **易憾**	hu4 die2 **护蝶**	hou4 hui3 **后悔**
Later Mismatch	xin1 qin2 **辛勤**	jian4 quan2 **健全**	xin1 qin3 **辛寝**	xin1 qin4 **辛沁**	jian4 quan4 **健劝**	lan3 duo 4 **懒惰**

The design and sample stimuli are presented in Table 5. Two groups of stimuli were selected and created according to whether the first or the second tones of base words were altered. Target words were preceded by their semantic primes and unrelated control primes. The control primes were created by re-pairing semantic primes with the targets. Crucially, the targets were also preceded by nonword primes that were created by changing the tones of semantic primes or semantic controls. In the Early Mismatch group, the two sets of tone-mismatch nonword primes differed in the first tones from with semantic primes. What differed them was that the first tones in the "high similarity" set were either Tone 2 or Tone 3, which shared many features with the first tones of semantic primes, which were either Tone 3 or Tone 2. The first tones in the "low similarity" set were either Tone 1 or Tone 4, which shared fewer features with the first tones of semantic primes. Nonword control primes were created by changing the first tones of semantic controls. Nonword primes in the Later Mismatch group were created in the same way, except that tone alternations were all on the second tones.

Mean reaction times and error percentages are presented in Table 6. Statistical analyses revealed significant semantic priming effects (35 ms) for both groups of stimuli. More importantly, nonword tone-mismatch primes of both "high similarity" and "low similarity" showed significant priming effects (15 and 18 ms) when mismatch was on the second tones. Nonword primes of "high similarity" also showed a significant priming effect (21 ms) when mismatch was on the first tones. But nonword primes of "low similarity" in the

Early Mismatch group did not have a significant effect (2 ms). The analyses of error rates did not find any significant effects or interactions.

Table 6 Experiment 3: Mean Reaction Time (ms) and Error Percentage (in parentheses)

Stimulus Group	Auditory Priming Condition				
	Word Prime		Nonword Prime		
	Semantic	Semantic Control	High Similarity	Low Similarity	Nonword Control
Early Mismatch	520 (3.1)	555 (3.9)	548 (2.0)	567 (1.8)	569 (2.3)
Later Mismatch	520 (1.8)	555 (2.2)	551 (1.2)	548 (2.6)	566 (3.8)

Therefore, it seems that both competition environment, the degree of tone mismatch, and the earliness of tone mismatch had effects on lexical access and activation of lexical representations. The influence of competition environment was clear when priming effects for tone-mismatch primes in this experiment were compared with those in Experiments 1 and 2. Lack of lexical competition in this experiment gave more opportunities for the mismatching speech input to activate phonological representations of base words. The effect of tone similarity was clear when the priming effects for mismatching primes of high and low similarity in the Early Mismatch group were compared. Input tones sharing many features with the underlying tones activates the latter and leads to semantic facilitation while input tones sharing fewer features do not activate significantly the underlying tonal representations. The effect of the earliness of tone mismatching was clear when the patterns of priming effects for the two groups of stimuli were compared. Matching the first tone will guarantee the base word as a lexical candidate and further speech input will support its activation, no matter the degree of deviation of the second tone. Mismatching the first tone, however, will the leave the candidacy of the base word to the goodness of fit between the input tone and the underlying representation. Clearly, competition environment, tone similarity, and constituent position of tone deviation interact in constraining lexical access and semantic activation of the base word.

General Discussion

The main purposes of this study were to investigate whether tonal information has immediate effect on constraining semantic activation in spoken word recognition of Chinese, how lexical access is influenced by competition environment and by global and local goodness of fit between speech input and underlying phonological representation, and how tonal information is represented in the mental lexicon. Three cross-modal semantic priming lexical decision experiments found complex interactions between competition environment, degree of mismatch of tonal information, and the earliness of mismatch in constraining lexical access and semantic activation. Experiments 1 and 2 used real words that differed from base words (semantic primes) in one or two tones as mediated primes. While Experiment 1 found no priming effect for mismatch primes differing in both tones, Experiment 2 did reveal a significant effect, in error rates, for primes differing in one tones. The constituent position of mismatching tones had no influence on the priming effects. Experiment 3 used nonwords differing from semantic primes in one tone as mediated primes, minimising competition between lexical candidates. Significant priming effects were found for primes mismatching the second tones, whether the resulting tones share many or few features with the underlying tones. However, a significant effect was found only for primes whose first mismatching tones share many features with the underlying tones.

To account for these findings, it is important to recognise that in spoken word recognition of Chinese, speech input maps onto lexical representations both in terms of segmental information and suprasegmental (tonal) information. While match between segmental input information and underlying representations can lead to the activation of corresponding semantic representations, a mismatch between tonal information and underlying representations will counter-act or cancel the effect of segmental matching. This "bottom-up" inhibitory effect (Marslen-Wilson, 1987) depends on the degree of mismatch. Mismatching in two tones, as in Experiment 1, is sufficient to cancel the facilitatory effect from segmental matching, resulting in no semantic activation of base words and no priming for visual probes. Mismatching in one tone, as Experiment 2, however, has less inhibitory effect on the activation of underlying phonological representations, leaving a window for semantic activation. Mismatching in some features of a tone, as in Experiment 3, gives even more opportunities for the underlying representations to be activated.

The activation of underlying representations is constrained not only by the match or mismatch from speech input, but also by lexical competition in the lexicon. For real word mismatching primes (Experiment 2), the speech input maps not only onto lexical representations of semantic primes, but also onto representations of themselves. The activation of semantic primes is influenced not only by mismatch of tonal information from speech input, but also by competition or inhibition from the activation of tone-mismatch primes. For nonword primes (Experiment 3), such lexical competition is reduced because such primes do not have their own representations in the lexicon. Nonword speech input is more likely to be "attracted" to lexical representations of base words and activate their semantic representations. This difference in

lexical competition accounts for the more prominent priming effects for tone-mismatch primes in Experiment 3.

The present data also speaks to the controversy about the importance of speech onset in lexical access. While Experiment 2 did not find a difference for real-word mediated primes mismatching the first or second tones, Experiment 3 found that the earliness of tone mismatch and the similarity between input tones and underlying representations can jointly influence semantic activation of base words. Matching the initial tone, the semantic primes (base words) are in the activated initial cohort. Their phonological representations can be sufficiently activated by further speech input whether the mismatching second tones share more or less features with the underlying tones. Mismatching the initial tone, however, whether the semantic primes will be lexical candidates depends on the extent of fit between the input tone and underlying representations. Therefore, both sequential processing and global goodness of fit constrains lexical access of spoken words.

The similarity effect also suggests that Mandarin Chinese has at least a level of underlying representation in terms of fine-grained tonal features, although the present data does not exclude the possibility of a level of representation in terms of categorical tone patterns. This proposal is consistent with Zhou and Marslen-Wilson (2000) who also found a similarity effect in phonological judgement to pairs of Chinese characters. In spoken word recognition, tonal features in speech input are subtracted and mapped onto phonological representations in the lexicon. Tone patterns are activated according to the extent of underlying tonal features being activated. Similar tones sharing tonal features are co-activated by speech input.

Reference

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29 (3),* 201-219.

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

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of words have a special status in auditory word recognition? *Journal of Memory and Language, 32,* 193-210.

Connine, C. M., Blasko, D. G., & Wang, J. (1994), Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics, 59,* 624-636.

Guo, J. F. (1993). *Exploring tone and intonation in Chinese*. Beijing: Beijing Language Institute Press.

Marslen-Wilson, W. D. (1987), Functional parallelism in spoken language understanding, *Cognition, 25,* 71-102.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken word: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 576-585.

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34,* 145-156.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49,* 25-47.

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

Zhou, X. (2000). Phonology in lexical processing of Chinese: Priming tone neighbors. *Psychological Science, 23,* 133-140.

Zhou, X., & Marslen-Wilson, W. (1997). The abstractness of phonological representation in the Chinese mental lexicon. In H.-C. Chen (Ed). *Cognitive Processing of Chinese and other Asian languages* (pp. 3-26). Hong Kong: The Chinese University Press.

Zhou, X., & Marslen-Wilson, W. (2000). *The representation of tonal information in the Chinese mental lexicon*. Manuscript submitted for publication.