心 理 学 报 2004,36(4):379～392
Acta Psychologica Sinica

Constraints of Lexical Tone on Semantic Activation in Chinese Spoken Word Recognition

Zhou Xiaolin
(*Center for Brain and Cognitive Sciences and Department of Psychology, Peking University, Beijing* 100871, *China*)

Qu Yanxuan, Shu Hua
(*Department of Psychology, Beijing Normal University, Beijing*, 100875 *China*)

Gareth Gaskell
(*Department of Psychology, University of York, UK*)

William Marslen-Wilson
(*MRC Cognition and Brain Sciences Unit, Cambridge, UK*)

Abstract Constraints of lexical tones on semantic activation on the recognition of spoken words in 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, *jump* vs. tiao2 yue1, *treaty*; numbers denote tone types) were used as auditory primes while words that were semantically related to one of the pairs were visually presented for lexical decision. The semantic primes and the tone-mismatch primes differed in tones of 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. It was found that the appearance of significant priming effects for the tone-mismatch primes depended on lexical competition environment, the goodness of fit between input tones and underlying tones, and the constituent position of mismatching tones. The results are discussed in terms of how tonal information in speech input is mapped onto the lexicon, how tonal information is represented in the lexicon, and how tonal constraints on semantic activation are influenced by competition environment.

Key words spoken word recognition, semantic activation, lexical tone, cross-modal priming.

1 Introduction

The role of suprasegmental information carried by the pitch or amplitude contour in lexical access and semantic activation of spoken words has been the focus of attention only in recent years[1～7], yet systematic studies of spoken word recognition began more than 20 years ago. This state of affairs reflects the fact that most current theories of spoken word recognition, such as the Cohort model[8,9], the Distributed Cohort Model[10], the TRACE model[11], the Shortlist model[12], and the Neighborhood Activation Model[13], ware developed on data from experiments in English, in which suprasegmental information by itself does not play a strong role in lexical semantic activation[14]. However, there are many languages in the world that use suprasegmental information to differentiate lexical items. For example, some words in Spanish differ only in stress[15,4]. Mandarin Chinese uses tones of individual syllables, i.e., the pitches carried by the vocalic part of syllables[16], to disambiguate monosyllabic morphemes that would be otherwise extensively homophonic. In Mandarin Chinese there are 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, although not all the segmental templates are accompanied by all the four tones. Thus

收稿日期:2004－04－29
Corresponding author: Zhou Xiaolin, E-mail: xz104@pku.edu.cn

the syllable "yi1" (numbers denote the tone patterns in Mandarin Chinese) represents the morphemes 衣 (*cloth*), 医 (*medical*), 蚁 (*ant*) etc, the syllable "yi4" corresponds to 易 (*easy*, *trade*), 译 (*translate*), 义 (*righteous*) etc. The purpose of the present study is to investigate how tonal information is used to constrain semantic activation in the spoken word recognition of Chinese, how the processing of tonal information interacts with other processes, and how tonal information is represented in the Chinese mental lexicon.

Studies of segmental lexical access in spoken word recognition have established at least two general principles: multiple activation and lexical competition. Segmental speech input maps onto multiple underlying lexical representations that are activated simultaneously. This multiple lexical activation competes for the eventual recognition. Different models of spoken word recognition, however, have different suggestions for how multiple activation and how lexical competition is conducted on-line. For multiple activation, the early version of the Cohort model [9] emphasizes the sequential matching of speech input and underlying phonological representations in the lexicon. Lexical candidates (i.e., the word-initial cohorts) are firstly established on the basis of the initial input syllable. Underlying representations that match the initial syllable are activated and underlying representations that do not match the initial syllable could never be candidates for recognition in initial access. As more speech input comes in, candidates that continue to match the input will stay on and get a boost in activation level while candidates that no longer match the input will begin to de-activate and drop out of the cohort. A word is recognized only when one word is left in the cohort.

However, other models of spoken word recognition, including the later version of the Cohort model [8], emphasize the global goodness of fit between speech input and underlying representations, rather than the strict sequential matching, in determining what representations are activated as lexical candidates. A word can still be a lexical candidate if speech input does not deviate too much from its underlying phonological representation, whether the mismatch is at the beginning, the middle, or the end of speech input. A mismatch at the initial position may still exert more influence on lexical activation than a mismatch at other positions, but this is not due to sequential matching per se but due to the dynamics of activation and competition, as in the TRACE model [11]. Underlying representations that match the speech input best will have a higher activation level than underlying representations that deviates more from speech input

What then is a tolerable degree of mismatch? There is some variability in the answers that different experimental techniques provide. For example, Connie, Blasko, and Titone [17] found that nonword speech input deviating from the underlying representation in terms of a single feature could activate the semantic representation of the base word even if this mismatch was at the beginning of the word, but speech input that deviated more severely could not. On the other hand, some experiments found that even a single feature mismatch has a strong effect on lexical activations. For example, a study using French by Frauenfelder, Scholten & Content [18] examined the effect of single feature deviation on lexical activation using phoneme monitoring. If a deviation such as focabulaire for vocabulaire results in activation of the lexical item, then there should be facilitated responding when one of the phonemes in the word (e.g., the final /r/ of vocabulaire) is a target for phoneme monitoring. Frauenfelder et al. found facilitation, but only when there was a sufficient delay between mismatch position and target phoneme. Soto-Faraco et al. [4] found similar strong and immediate effects of single feature deviations using cross-modal priming.

Lexical competition may also be relevant to the goodness-of-fit issue, depending on how competition is conducted. In the Cohort model, lexical candidates passively decay and drop out of the word-initial cohort mainly because of their mismatch with speech input and the bottom-up inhibition from mismatched segments. Different words in the same cohort do not compete with or inhibit each other directly [19]. In most other models, however, there exists lateral inhibition among lexical candidates. A word having a higher activation will exert large inhibition on competing words with lower activation level. This competition takes place not only among lexical candidates in the same word-initial cohort, but also among candidates that deviate from the speech input in the word initial, middle, or final position. A word suffering from mismatch may still play a role in the ongoing lexical competition, although its activation and the competitive strength depend greatly on the degree of its match/mismatch with the speech input. A striking demonstration of the influence of lexical competition on mismatch effects involved the manipulation of voice onset time for initial consonants of monosyllabic words [20]. This manipulation can create tokens that are acoustically intermediate between two minimally differing phonemes (e.g., between (/t/ and /d/). When these ambiguous phonemes were placed in contexts for which only one alternative was a word (e.g., [t/d]ask, where task but not dask is a word), there was evidence of activation of the word alternative despite the minor mismatch. However, when both alternatives were words (e.g., [t/d]ent, where tent and dent are words) there was no evidence of lexical activation. Therefore, in some cases the mismatching effect of a deviation depends critically on the lexical neighbourhood of the word. Some uncertainties about the goodness of fit computation in speech perception remain, but in terms of segmental mismatch, the constraints of mismatching information on the activation of underlying phonological and semantic representations appear to be determined jointly by the degree of mismatch, the position of mismatch, temporal factors, plus the properties (such as the frequency and the number) of other lexical competitors.

Evidence supporting the above general principles comes mainly from studies on the effect of mismatching segmental information on lexical activation. Studies of the role of suprasegmental information in spoken word recognition concentrate mostly on its usefulness in speech segmentation [21~23] (see [24] for a review). However, a few recent studies do address the extent to which suprasegmental information constrains the lexical access and semantic activation of individual words. Using gating and repetition priming lexical decision tasks, Cutler and Otake [25] found that Japanese participants are sensitive to pitch-accent information in lexical selection and lexical activation. For Spanish, Soto-Faraco et al. [4] showed that mismatch in stress, like vowel or consonant mismatch, has strong effects on response speed in a fragment priming lexical decision task. Furthermore, Cutler and Van Donselaar [26] found that Dutch word pairs differing only in stress are not treated as homophones and are not activated in parallel when one word of the pair is heard. This contrasts with the findings of an earlier study in English [14], which showed multiple semantic activation of English words differing only in stress. This difference between languages suggests that suprasegmental stress information is used to constrain semantic activation in Dutch but not in English, and that the usefulness of suprasegmental information in constraining lexical access and semantic activation depends to a great extent on the extensiveness of this information being used in the language to differentiate lexical items. Even in English, however, suprasegmental effects on lexical activation can be found. Davis, Marslen-Wilson & Gaskell[27] showed that differences between phonemically identical syllables in mono- and polysyllabic words (e.g., the syllable "cap" in cap and captain) could be used to influence lexical activation. Their experiments suggested that syllable duration is used as a weak source of discriminating information between words of different lengths, reflecting the fact that longer words in English tend to contain shorter initial syllables.

For tonal information in Chinese, earlier studies were mostly concerned with the acoustic and perceptual aspect of speech processing. These studies address issues such as the importance of fundamental frequency (F0), amplitude, and length in the perception of tones[16,29~31]. In general, they do not provide clear and consistent 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. A study by Cutler and Chen [32] used the lexical decision task to nonwords created by altering either the consonants, vowels, or the tones of the second syllable of compound words, and found that tone-altering nonwords were much more difficult to reject than segment-altering nonwords. The authors concluded that tonal information becomes available only later than the associated segmental information. This conclusion, however, is contradicted by the finding that compound words differing only in tones (i.e., tone neighbors, such as "tiao2 yue1", *treaty* and "tiao4 yue4", *jump*) inhibit each other in cross-modal and visual-visual priming lexical tasks [6].

Other studies also examined the constraints of tonal information on lexical semantic activation. Ye and Connine [5] asked participants to monitor for either vowels or tones of spoken syllables and found that in a neutral context tone monitoring elicited slower responses than vowel monitoring, perhaps suggesting that tone information was less available than vowel information. However, the difficulty was reversed when target syllables were embedded in highly constraining semantic contexts (i.e., idioms rather than in phrases), showing that lexical information can in some circumstances make tone information more prominent than vowel information. Moreover, the similarity between target tones and mismatching tones modulated the monitoring latency, showing a graded effect. Bi [1] presented participants with nonword speech input that deviated from base words either in terms of syllable-initial consonants or in terms of tones, and participants made lexical decisions to visual probes that were either the base words themselves or were semantically related to the base words. No priming effects were found for semantic probes, suggesting that mismatch in tones can effectively block access to the base word semantic representations.

The present study employed cross-modal semantic priming to investigate more systematically the constraints of tonal information on lexical semantic activation in speech recognition of Chinese spoken words. In this paradigm, a disyllabic compound word is presented auditorily as a prime, while a semantically related word is presented visually for lexical decision. Typically, the spoken input will activate the corresponding meaning in the lexicon, and facilitate processing of the probe word, due to the overlap in meaning. By measuring the response to the visual probe we can gain knowledge about how tonal information in speech 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 study, 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, both "tiao4 yue4" and "tiao2 yue1" should activate the two words' phonological representations in the lexicon. However, the tonal information in the speech input of one word may create mismatch with 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 fit the suprasegmental or tonal phonological representation of "tiao4 yue4". If tonal information is used immediately to constrain semantic activation, this tone mismatch between speech input and the underlying phonological representation may cancel or 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 semantic activation of the segmentally matching neighbor (i.e., the semantic prime). The response to the visual probe may still be facilitated to some extent by the tone-mismatch prime.

We conducted three experiments. In Experiments 1 and 2, we asked whether mismatch of one or two tones in the speech input of disyllabic primes would reduce or even eliminate semantic activation of their base words. The tone-mismatch primes here were real words themselves and had their own representations in the lexicon. The activation of these representations by the speech input of tone-mismatch primes may create competition for the activation of base words, reducing semantic priming for visual probes. In Experiment 3, we used nonwords differing only in the first or the second tones from base words as tone-mismatch primes. These nonwords did not have their own representations in the lexicon and hence could not create competition for lexical activation of the base words.

We also asked whether mismatch in the first or the second tone of the disyllabic speech input would 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 [8, 17, 33, 34]. We further asked whether tonal information is represented in the recognition system in terms of categorical tone patterns or in terms of fine-grained tonal features. According to some phonological theories[35], tones in Mandarin Chinese can be further decomposed into 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 judgment task in which participants had to judge whether a pair of consecutively presented characters were homophones (disregarding their tones), Zhou and Marslen-Wilson[36] 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[5]. In the present study, we manipulated the similarity between tones in the speech input of mediated primes and the corresponding tone representations of base words. Most studies of suprasegmental mismatch effects have used binary contrasts[4,26]. Here, however, we have something more akin to the segmental deviation studies, in which the strength of the tonal mismatch can be manipulated, to examine whether further parallels exist between segmental and suprasegmental match and mismatch.

2 Experiment 1

Experiment 1 examined whether disyllabic spoken words (e.g., "tiao2 yue1", *treaty*) would activate the underlying representations of words deviating in both tones (e.g., "tiao4 yue4", *jump*). If tones play an immediate role in constraining lexical processing of Chinese, the speech input should activate its own lexical representations but should not activate the semantic representation of the tone mismatching word. Consequently the processing of a visual probe (e.g., 奔跑 "ben1 pao3", *run*) should be facilitated by presenting its semantic associate "tiao4 yue4" (*jump*), which will be referred to as the "base word", but should not be influenced by the presentation of a word differing in tones from the base word (e.g., "tiao2 yue1", *treaty*), which will be referred to as the "tone mismatch" prime. On the other hand, if tones have no strong or immediate effect on lexical access, words differing in tones are virtually homophones and presenting "tiao2 yue1" should be able to activate semantic representations of both "tiao2 yue1 (*treaty*)" and "tiao4 yue4" (*jump*). Consequently lexical decision to the visual target 奔跑 (ben1 pao3, *run*) should be facilitated by "tiao2 yue1", even though it is not semantically related to the input word directly.

To examine potential competition between semantic activation of base words and tone-mismatch words, the relative frequency between the base words and tone-mismatch words was manipulated. If words differing in tones are virtual homophones, semantic activation of the lower frequency primes could be reduced due to competition and inhibition from their higher frequency neighbors. Consequently, the priming effects for lower frequency primes could also be reduced.

2.1 Method

2.1.1 Design and Materials

The experimental design and sample stimuli are presented in Table 1. Three groups of stimuli were selected according to the relative frequency of the semantic primes (the base words) and the tone-mismatch primes. Each group had 21 pairs of base words and tone-mismatch words. The average frequencies of these words for the three groups are presented in Table 1. The frequency information came from the Institute of Language Teaching and Research (1986) corpus of 1.3 million words. Frequencies listed in Table 1 (and the following Tables 3 and 5) have been transformed into occurrences per million. Visual probes were selected from semantic associates of the base words. Unrelated control primes were created by re-pairing the tone-mismatch primes and visual probes within each group. Thus in each frequency group control primes and tone-mismatch primes were the exactly same set of words. Tone-mismatch primes and control primes had no phonological or semantic relations with targets. Control primes were matched to tone-mismatch primes rather than to semantic primes because the most important empirical question here was to check whether there were priming effects for tone-mismatch primes. The direct semantic priming effects from base words should be robust enough to withstand differences in the frequency of control primes.

The critical primes and their visual probes were assigned, in a Latin Square design, to three counterbalanced test versions. In each version, there were 21 semantic prime-visual target pairs, 7 from each of the three frequency group. Each version also included 21 tone mismatch-target pairs and 21 control prime-target pairs. Filler word-word and word-nonword pairs were then added to the three test versions. Among the 87 word-word pairs, 15 were semantically related. The other 72 pairs of primes and targets were neither semantically nor phonologically related. The 150 word-nonword pairs had no phonological relations between primes and targets. The two-character nonword targets were created by randomly 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. Thus in each test version, there were 300 prime-target pairs, with half of the targets requiring "yes" responses in lexical decision and the other half requiring "no" responses. Among the prime-target pairs, 36 (12%) were semantically related and 21 (7%) were pairs in which targets were preceded by words that deviated in tones from their semantic primes. There were also 30 pairs of practice items that had a similar composition.

Table 1 Experiment 1: Design and Sample Stimuli

Stimulus Group		Auditory Priming Type			Visual Probe
		Base Word	Tone Mismatch	Control	
Base Word Dominant		tiao4 yue4 跳跃 *jump*	tiao2 yue1 条约 *treaty*	ban1 fa1 颁发 *issue*	奔跑 ben1 pao3 *run*
	Frequency	133	6	6	111
Equal Frequency		Jiu3 jing1 酒精 *alcohol*	jiu1 jing4 究竟 *ever*	bian1 jie4 边界 *boundary*	燃料 ran2 liao4 *fuel*
	Frequency	14	12	12	54
Mismatch Dominant		xian4 jin1 现金 *cash*	xian1 jin4 先进 *advanced*	gu1 ji4 估计 *estimate*	支票 zhi1 piao4 *cheque*
	Frequency	15	72	72	24

A pseudo-random sequence was used to arrange the stimuli in each version in such a way that, across the three test versions, the same target appeared at the same position. The only difference between versions was that the primes for a particular critical target were different. There were never more than 4 consecutive targets requiring the same responses. Prime-target pairs with the same relations were roughly evenly distributed across a test sequence. There was a break after practice and two breaks in the main test session. The first three prime-target pairs after each break were always fillers.

2.1.2 Procedures A female native speaker of Mandarin Chinese recorded all the auditory primes on DAT tapes. These stimuli were then digitized and stored as individual computer files. Visual probes were created in 48-point songti font. Words and nonwords were about 2.4×3.8 cm in size.

The presentation of stimuli to participants and recording of reaction times was controlled by the experimental software VMASTR and DMASTR, which was made available to us by Ken and Jonathan Forster of the University of Arizona. Control files were constructed to play auditory primes through headphones to participants, who were tested in groups of three or less in a session. Participants were told that all the words they heard through headphones were real words and they had to listen carefully. On about 10% of trials participants were given time, after making lexical decision, to write down what they just heard. Participants were also asked to judge as quickly and as accurately as possible whether the visual probe in each trial was a real word or not, and press the corresponding button on a response box. Probes were presented at the offset of the speech at the center of the computer screen. Participants were seated 50 to 60 cm from the screen. Response errors and reaction times, measured from the onset of target words, were automatically recorded by computers for each participant and each item. Responses slower than

2000 ms were counted as errors. There was an inter-trial interval of about 3 seconds.

2.1.3 Participants A total of 30 native speakers of Mandarin Chinese from the Beijing area were tested for the experiment, 10 in each test version. All were undergraduate students from Beijing Normal University and were paid for their participation.

2.2 Results

Participants' mean reaction times based on untrimmed correct responses, and error rates are summarized in Table 2. Analyses of variance (ANOVAs) were conducted for reaction times and error rates, with prime type as a within-participant, within-item factor and frequency group as a within-participant, between-item factor.

In the analyses of reaction time, the main effect of prime type was significant, $F1(2, 58)=7.275$, $p<0.01$, $F2(2, 116)=15.086$, $p<0.01$. Post hoc Newman-Keuls tests showed that the 28 ms semantic priming effect, compared with the control condition, was significant ($p<0.05$); but there was no priming effect for the tone-mismatch primes ($p>0.1$). Thus the processing of semantic primes facilitated lexical decision to the visual probes, but the processing of auditory primes differing in tones from the semantic primes had no effect on the processing of the visual probes. The main effect of frequency group was not significant, nor the interaction between prime type and frequency group, $F1<1$, $F2<1$. The pattern of priming effects was essentially the same whether the semantic primes had higher frequency than the tone-mismatch primes or not. The analyses of error rates found no effect at all ($F1<1$, $F2<1$).

2.3 Discussion

The present experiment found that 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. Clearly, 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 the tone-mismatch primes could be used to activate the underlying phonological representations of base words (semantic primes), the mismatch in two tones was sufficient to cancel this activation, leaving no space for the semantic activation of the base words. Moreover, the activation of a fully matching lexical competitor may have reduced activation of the base word through lexical competition.

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

Stimulus	Priming Type		
	Base Word	Tone Mismatch	Control
Base Word Dominant	499	527	527
	(0.5)	(0.5)	(0.9)
Equal Frequency	527	564	554
	(2.4)	(2.9)	(1.9)
Mismatch Dominant	512	540	539
	(1.1)	(2.1)	(4.2)
TOTAL	512	544	540

3 Experiment 2

Experiment 2 continued to investigate constraints of tonal information in speech input on lexical access and semantic activation. While the speech input of tone-mismatch 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 a role in constraining activation of underlying representations. This was carried out on two groups of stimuli, one with tone-mismatch primes matching the semantic primes on the first but not the second tome, and another matching on the second but not the first tone.

3.1 Method

3.1.1 Design and materials The experimental design and sample stimuli are presented in Table 3. Two groups of critical stimuli were selected. In the first group (the Early Mismatch group), semantic primes and tone-mismatch primes had the same segmental templates and the same second tones. In the second group (the Later Mismatch group), semantic primes and tone-mismatch primes had the same segmental templates and the same initial tones. Semantic primes and tone-mismatch differed phonologically in the initial tones for the first group of stimuli and in the second tones for the second group of stimuli. Visual probes were selected from the semantic associates of the semantic primes (base words). Each group had 30 pairs of semantic primes and tone-mismatch primes. Unrelated control primes were created by re-pairing tone-mismatch primes with the visual probes in the same group. The mean frequencies of primes and targets are reported in Table 3.

Table 3 Experiment 2: Design and Sample Stimuli

Stimulus Group		Auditory Priming Type			Visual Probe
		Semantic	Tone Mismatch	Control	
Early Mismatch		wen1 he2 温和 *geniality*	wen3 he2 吻合 *tally with*	qu4 shi4 去世 *die*	严厉 yan2 li4 *severe*
	Frequency	83	42	42	29
Later Mismatch		qing1 xin1 清新 *clear*	qing1 xin4 轻信 *gullibility*	shi2 ji1 时机 *occasion*	污浊 wu1 zhuo2 *grime*
	Frequency	40	94	94	40

Critical primes and their corresponding visual probes were assigned, in a Latin Square design, into three counter-balanced test versions. Each version had 20 semantic primes, 20 tone mismatching primes and 20 control primes, half of them coming from the Early Mismatch group and half from the Later Mismatch group. The same 60 word-word pairs and 120 word-nonword pairs were added to each version as fillers. These fillers had no semantic or phonological relations between primes and targets. Nonwords were created in the same way as in Experiment 1.

The recording and editing of auditory primes, the presentation of stimuli to participants, the recording of reaction times and error rates were conducted in the same way as Experiment 1.

3.1.2 Participants Thirty native speakers of Mandarin Chinese from the Beijing area were tested, 10 for each version. They were undergraduate students of Beijing Normal University and were not tested for Experiment 1.

3.2 Results and Discussion

Mean reaction times, base on untrimmed, correct responses, and error percentages are listed in Table 4. ANOVAs were conducted for reaction times and error rates respectively, with prime type as a within-participant, within-item factor and stimulus group as a within-participant, between-item factor.

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

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

As in Experiment 1, there was significant main effect of prime type, $F1(2, 58)=8.156$, $p<0.01$, $F2(2, 216)=4.922$, $p<0.01$. Newman-Keuls tests found a significant effect (27ms) for semantic primes ($p<0.05$), but no effect for tone-mismatch primes. The interaction between prime type and stimulus group was not significant ($F1<1$, $F2<1$), indicating that the earliness of tone mismatch did not influence the pattern of priming. However, the analyses of error rates found significant effects for both semantic and tone-mismatch primes. The main effect of prime type was significant, $F1(2, 58)=5.822$, $p<0.01$, $F2(2, 116)=6.870$, $p<0.01$, and Newman-Keuls tests found a significant effect not only for semantic primes ($p<0.05$), but also for tone-mismatch primes ($p<0.05$). Fewer errors were committed when visual targets were preceded by tone-mismatch primes than by control primes. The interaction between prime type and stimulus group was not significant, $F1(2, 58)=1.257$, $p>0.1$, $F2(2, 116)=1.652$, $p>0.1$, suggesting that the effects in error rates were essentially the same for the two groups of stimuli.

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 representations is an important factor in constraining lexical access and semantic activation. The goodness of fit takes into account not only the match and mismatch for segmental information, but also the match and mismatch for suprasegmental tones.

4 Experiment 3

Experiment 1 showed that two mismatching tones were enough to block lexical access as measured by semantic priming. Experiment 2 showed that deviations of a single tone also eliminated priming of response speed, although evidence of some activation was present in the finding of a reduced error rate. These two experiments have begun to narrow down the range of tone deviations that are tolerated during lexical access for a tone language. Clearly tone information has an important role to play in the lexical specification of spoken words. Experiment 3 extended this approach, investigating the ineraction of several factors that could influence 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, even though this competition could be modulated by the global goodness of fit between the speech input of mismatch primes and 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 base words by altering their first or second tones. Such speech input is less likely to generate lexical competition and more likely to activate the base words.

The second factor investigated in Experiment 3 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 some phonological theories of lexical tone [35], some tones, like Tone 2 and Tone 3, share more tonal features than other tones and hence are more similar phonologically. If tonal information in speech input is represented at some level in terms of tonal features, the activation of these representations should be influenced by the degree of tonal similarity, just as phonemic similarity plays a role in the segmental matching process. Speech input is more likely to activate the relevant underlying representations if the input tone is more similar to the lexically represented tone (i.e., having fewer deviating tonal features). 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. We did not observe significant effects for this factor in Experiment 2. However, the effect of the earliness of mismatch in tones in Experiment 2 may have been overshadowed by lexical competition between tone mismatch primes and base words (semantic primes). In this experiment, we manipulated the earliness of mismatch in speech input of nonwords.

4.1 Method

4.1.1 Design and Materials The design and sample stimuli are presented in Table 5. Two sets of semantic primes and visual probes were selected respectively for the Early Mismatch group and the Later Mismatch group. Each group had 50 prime-target pairs. The first syllable of the semantic primes (base words) in the Early Mismatch group and the second syllable of the semantic primes in the Later Mismatch group had the second or third tones. Mismatch nonword primes were then created by changing the initial tone of the base words in the Early Mismatch group and the second tone of the base words in the Later Mismatch group. For the High Similarity nonword primes, Tone 2 of the critical syllables of base words was changed into Tone 3 and Tone 3 into Tone 2. For the Low Similarity nonword primes, Tone 2 of the critical syllables of base words was changed into Tone 4, and Tone 3 into either Tone 1 or Tone 4. For semantic primes, unrelated control primes were created by re-pairing semantic primes with visual probes. For nonword primes, control primes were created by re-pairing Low Similarity primes with visual probes.

A Latin Square design was used to assign the critical primes and the corresponding visual probes into five counter-balanced test versions, with 10 primes from each of the five prime types in each stimulus group. Each test version now had 20 semantic primes, 20 real words control primes, 20 nonword primes with similar tone, 20 nonword primes with dissimilar tones, and 20 nonword control primes. Three types of filler prime-target pairs were then added to each of the test versions. The first type of fillers were 20 word-word pairs. The second type of fillers were 60 word-nonword pairs and the third were 60 nonword-nonword fillers. Thus in each test version, there 120 visual targets that required "yes" responses and 120 required "no" responses. Among the auditory primes, 120 were real words and 120 were nonwords. There were also 24 practice prime-target pairs with similar compositions as the formal test. The recording and editing of auditory primes, the presentation of stimuli to participants, the recording of reaction times and error rates were conducted as for previous experiments.

4.1.2 Participants

A total of 50 participants were tested, 10 for each test version. They were native speakers of Mandarin Chinese from the Beijing area. They were undergraduate students in Beijing Normal University and were not tested for Experiments 1 and 2.

4.2 Results

Mean reaction times and error percentages are presented in Table 6. ANOVAs were conducted separately for word and nonword primes.

Table 5 Experiment 3: Design and Sample Stimuli

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

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	555	548	567	569
	(3.1)	(3.9)	(2.0)	(1.8)	(2.3)
Later Mismatch	520	555	551	548	566
	(1.8)	(2.2)	(1.2)	(2.6)	(3.8)

For word primes, it is clear from Table 6 that there was no difference in the semantic priming effects for the Early Mismatch and Later Mismatch groups. The combined effect (35ms) was highly significant, $F1(1, 49) = 43.359$, $p < 0.01$, $F2(1, 98) = 33.019$, $p < 0.01$. More importantly, in the analyses of priming effects for nonword primes, the main effect of prime type was significant, $F1(2, 98) = 3.868$, $p < 0.05$, $F2(2, 194) = 5.463$, $p < 0.05$. Newman-Keuls tests showed that the main priming effect for "high similarity" nonword primes (18ms) was significant ($p < 0.05$), although the main priming effect for "low similarity" nonword primes (10ms) did not reach significance. Although the interaction between prime type and stimulus group did not reach significance, $F1(2, 98) = 2.135$, $p > 0.1$, $F2(2, 194) = 1.822$, $p > 0.1$, there was some suggestion that the patterns of priming effects were different for Early Mismatch group and Later Mismatch group for the "low similarity" primes. The 2-ms difference between "low similarity" primes and control primes in the Early Mismatch group did not reach significance in separate tests, but the 18-ms difference in the Later Mismatch group did ($p < 0.05$). The analyses of error rates showed no significant results.

4.3 Discussion

The response time priming effects for tone-mismatch nonword primes, in comparison with the null or weak effects for tone-mismatch real word primes in Experiments 1 and 2, demonstrates the importance of lexical competition in the semantic activation of spoken words. In the present experiment, nonword primes were used to examine the mismatching effects of tone difference. These stimuli should not strongly activate other words competing with the base words, and the lack of lexical competition in this experiment gave more opportunity for the mismatching speech input to activate the lexical representations of the base words.

The degree of mismatch between speech input and underlying phonological representations played a role in determining whether the semantic representations of base words could be activated. Input tones sharing many features with the underlying tones activated the base words and led to semantic facilitation of the visual probes. Input tones sharing fewer features with underlying tones, however, did not activate significantly the underlying tonal representations.

Moreover, there was weak evidence that the tone similarity effect was modulated by the earliness of mismatch. 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.

5 General Discussion

There were three main aims of this study. First, we wished to investigate whether tonal information has an immediate effect on semantic activation in spoken word recognition of Chinese. Second, we were interested in how lexical access is influenced by competition environment and by global and local goodness of fit at the suprasegmental level between speech input and underlying phonological representations. Finally, we wanted to explore 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 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, but only in error rates, for primes differing in one tone. The constituent position of mismatching tones had no influence on the priming effects. Experiment 3 used nonwords differing from semantic primes in one tone, minimising competition between lexical candidates. When the tonal deviation was featurally minimal there were semantic priming effects, whereas more maximal deviations produced no significant priming overall. However, there was a suggestion that the maximal deviations could generate lexical access if they occurred in the second syllable, but not the first syllable.

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 counteract or cancel the effect of segmental matching. This "bottom-up" inhibitory effect[8] depends on the degree of mismatch. Mismatch 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. Mismatch in one tone, as in Experiment 2, however, has less inhibitory effect on the activation of underlying phonological representations, leaving a window for semantic activation. Mismatch in some features of a tone, as in Experiment 3, gives even more opportunities for the underlying representations to be activated. The parallels between the effects of suprasegmental and segmental mismatch on lexical activation are extremely strong. Across the experiments, mismatches in tone had severe and swift effects on lexical activation, despite complete segmental overlap. In cases where mismatch effects were not complete, the lexical competition environment and the featural similarity of the deviation dictated the extent of activation. Soto-Faraco et al.[4] argued that for the case of stress in Spanish segmental and suprasegmental information should be treated as equivalent in terms of their effect word recognition process. The current results support this finding, and extend it to a nonbinary suprasegmental dimension.

The activation of underlying representations is constrained not only by match or mismatch from speech input, but also by lexical competition in the lexicon. For real word mismatching primes (Experiment 2), the speech input is likely to map onto at least two lexical representations, one of which matches the speech only segmentally, and the other both segmentally and suprasegmentally. The activation of semantic primes is influenced not only by mismatch of tonal information from speech input, but also by competition or inhibition from activation of tone-mismatch primes. This experiment showed that the combination of the presence of a fully matching lexical competitor and a mismatch in one tone had strong effects on lexical activation, and priming effects were only found in terms of error rates. For nonword primes (Experiment 3), such lexical competition is reduced because such primes do not have their own representations in the lexicon, meaning that no lexical item will match the nonword primes both segmentally and suprasegmentally. 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, although no firm conclusions can be reached on the basis of the current results. While Experiment 2 found no difference between real-word mediated primes mismatching the first and second tones, Experiment 3 suggested that the earliness of tone mismatch and the similarity between input tones and underlying representations can jointly influence semantic activation of base words. For items in which the mismatching tone occurs only in the second syllable, the semantic primes (base words) will be activated in the word-initial cohort. In this case there is some evidence of facilitation irrespective of whether the tonal mismatch is minimal or maximal (there was a significant 18 ms priming effect for the "low similarity" mismatch condition). For items with a mismatch in the initial tone, however, there may not be sufficient activation of the base word at any point to generate semantic priming unless the tone mismatch is minimal (resulting in no priming for the "low similarity" mismatch condition). Therefore, it is possible that both sequential processing and global goodness of fit constrain lexical access of spoken words.

The similarity effect also suggests that at some level Mandarin Chinese represents tones in terms of fine-grained features, although the present data do not exclude the possibility of a more categorical representation elsewhere in the system. This proposal is consistent with Ye and Connine[5] who found a similarity effect in tone monitoring and with Zhou and Marslen-Wilson[36] who 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. In any case, it is clear that tone information is a critical component of the lexical representation of a spoken word in languages like Chinese. In all languages, the identification of a spoken word requires a close match between segments in the input and lexical representation of the word, and the success of the matching process depends on factors such as featural similarity, lexical competition environment, and the time course of match and mismatch as the speech signal unfolds. In tone languages it seems that exactly the same factors apply, strengthening the suspicion that any information-segmental or suprasegmental-that can be used to distinguish between words will be used by the recognition system in much the same way during lexical access.

Reference

1 Bi Y. Spoken word recognition in Chinese: Segment, tone, and their processing time course. Unpublished Master's thesis, Beijing Normal University, 1998

2 Cutler A, Otake T. Pitch accent in spoken-word recognition in Japanese. Journal of Acoustic Society of America, 1999, 106: 1191~1191

3 Cutler A, Otake T. Rhythmic categories in spoken-word recognition. Journal of Memory and Language, 2002, 46: 296~322

4 Soto-Faraco S, Sebastian-Galles N, Cutler A. Segmental and suprasegmental mismatch in lexical access. Journal of Memory and Language, 2001, 45: 412~432

5 Ye Y, Connine C M. Processing spoken Chinese: The role of tone information. Language and Cognitive Processes, 1999, 14: 609~630

6 Zhou X. Phonology in lexical processing of Chinese: Priming tone neighbors. Psychological Science, 2000, 23: 133~140

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

8 Marslen-Wilson W D. Functional parallelism in spoken language understanding, Cognition, 1987, 25: 71~102

9 Marslen-Wilson W D, Welsh A. Processing interactions during word recognition in continuous speech. Cognitive Psychology, 1978, 10: 29~63

10 Gaskell M G, Marslen-Wilson W D. Integrating form and meaning: a distributed model of speech perception. Language and Cognitive Processes, 1997, 12: 613~656

11 McClelland J L, Elman J L. The TRACE model of speech perception. Cognitive Psychology, 1986, 18: 1~86

12 Norris D. Shortlist: A connectionist model of continuous speech recognition. Cognition, 1994, 52: 189~234

13 Luce P A, Pisoni D B, Goldinger S D. Similarity neighborhoods of spoken words. In: G T M Altmann ed. Cognitive Models of Speech Processing. Cambridge, MA: MIT Press, 1990. 122~147

14 Cutler A. Forbear is a homophone: Lexical prosody does not constrain lexical access. Language and Speech, 1986, 29: 201~219

15 Dupoux E, Pallier C, Sebastian N, Mehler J. A destressing "deafness" in French? Journal of Memory and Language, 1997, 36: 406~421

16 Gandour J. Perceptual dimensions of tone: evidence from Cantonese. Journal of Chinese Linguistics, 1981, 9: 20~36

17 Connine C M, Blasko D G, Titone D. Do the beginnings of words have a special status in auditory word recognition? Journal of Memory and Language, 1993, 32: 193~210

18 Frauenfelder U H, Scholten M, Content A. Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. Language and Cognitive Processes, 2001, 16: 583~607

19 Marslen-Wilson W D. Activation, competition, and frequency in lexical access. In: G T M Altmann ed. Cognitive Models of Speech Processing. Cambridge, MA: MIT Press, 1990. 148~172

20 Marslen-Wilson W D, Moss H E, Halen S. Perceptual distance and competition in lexical access. Journal of Experimental Psychology: Human Perception and Performance, 1996, 22: 1376~1392

21 Cutler A, Norris D. The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance, 1988, 14: 113~121

22 Norris D, McQueen J M, Cutler A. Competition and segmentation in spoken-word recognition. Journal of Experimental Psychology: Learning, Memory, & Cognition, 1995, 21: 1209~1228

23 Mattys S L, Samuel A G. How lexical stress affects speech segmentation and interactivity: evidence from the migration paradigm. Journal of Memory and Language, 1997, 36: 87~116

24 Cutler A, Dahan D, van Donselaar W. Prosody in the comprehension of spoken language: A literature review. Language and Speech, 1997, 40: 141~201

25 Cutler A, Otake T. Pitch accent in spoken-word recognition in Japanese. Journal of Acoustic Society of America, 1999, 106: 1191~1191

26 Cutler A, Van Donselaar W. Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. Language and Speech, 2001, 44: 171~195

27 Davis M H, Marslen-Wilson W D, Gaskell M G. Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. Journal of Experimental Psychology: Learning, Memory, & Cognition, 2002, 28 (1): 218~244

28 Fox R A, Unkefer J. The effect of lexical status on the perception of tone. Journal of Chinese Linguistics, 1985, 13: 69~90

29 Lin H B, Repp B H. Cues to the perception of taiwanese tones. Language and Speech, 1989, 30: 187~499

30 Shen X S, Lin M. A perceptual study of Mandarin tones 2 and 3. Language and Speech, 1991, 34: 145~156

31 Whalen D H, Xu Y. Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica, 1992, 49: 25~47

32 Cutler A, Chen H C. Lexical tone in Cantonese spoken-word processing. Perception & Psychophysics, 1997, 59: 165~179

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

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

35 Guo J F. Exploring tone and intonation in Chinese. Beijing: Bei-

jing Language Institute Press, 1993

36 Zhou X, Marslen-Wilson W. The representation of tonal information in the Chinese mental lexicon. Unpublished manuscript, 2000

汉语听觉词汇加工中声调信息对语义激活的制约作用

周晓林
(北京大学脑科学与认知科学中心，北京大学心理学系，北京 100871)

曲延轩 舒 华
(北京师范大学心理学院，北京 100875)

Gareth Gaskell
(英国约克大学心理学系)

William Marslen-Wilson
(英国医学研究基金会剑桥认知与脑科学研究所)

摘 要 进行了三个跨通道启动词汇判断实验，探讨汉语听觉词汇加工中声调信息对语义激活的制约作用。实验一和二使用具有相同音段信息、但不同声调信息的双音节合成词(如“条约”和“跳跃”)作为听觉启动词，与其中一个词有语义关系的双字词作为视觉探测词。语义启动词与其配对的声调不匹配词在第一音节的声调、第二音节的声调、或两个声调上有所不同。实验三改变语义启动词的第一或第二音节的声调，以产生声调不匹配的假词启动项；实验三还同时变化了原来的声调与产生的声调之间的相似性。实验结果表明，声调不匹配的启动项目是否产生显著的启动效应取决于词汇竞争的环境、声调输入与深层声调表征之间匹配的程度，以及不匹配的声调所在的位置。文章从语音输入中的声调信息如何激活词汇表征、声调信息如何存储在心理词典中、声调对语义激活的制约作用如何受竞争环境的影响这三个方面讨论了研究的发现。

关键词 听觉词汇加工，语义激活，声调，跨通道启动。

分类号 B842