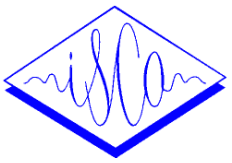

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

# Lexical Tone in the Speech Production of Chinese Words

**Xiaolin Zhou** and **Jie Zhuang**
Department of Psychology
Peking University
Beijing 100871, China
xz104pku.edu.cn

## Abstract

Picture and word naming tasks were used to investigate the activation of tonal information in the speech production of Chinese words. A picture was presented first, followed by a Chinese character superimposing on the picture. This character shared either the segmental template and the tone, or only the segmental template, or no phonological properties with the name of the picture. Subjects were asked to name the picture (Experiment 1) or the character (Experiment 2). The SOA between presentation of the picture and presentation of the character was manipulated. In Experiment 1, the naming of pictures at the SOA of 57 ms and 200 ms was facilitated when the names of the pictures shared phonological properties with the characters. But the effect for tone activation was lager at the short SOA than at the longer SOA while the effect for segmental activation was stable across SOAs, indicating different time courses for the activation of tonal information and segmental information in speech production. In Experiment 2, at the both the short and the long SOAs, character naming was facilitated when the characters shared both segmental and tonal information with the picture names. Character naming tended to be inhibitory when the characters shared only the segmental templates with the picture names. These results suggested that lexical tones are activated very early in the speech production of Chinese words and the mismatch of tonal information in primed picture naming or character naming can effectively reduce or neutralize segmental facilitation.

## Introduction

Producing a word in human speech involves several computational steps: conceptual preparation, lexical selection, morpho-phonological encoding, phonetic encoding, and articulation (Levelt, 1999). The information we want to express has to be realized as a message, which contains one or more concepts corresponding to words in our mental lexicon. The activation of a lexical concept constrains the retrieval of the phonological form of a word, which provides bases for further detailed phonetic encoding and articulation. The analyses of speech errors demonstrate that, for both Western languages and Mandarin Chinese, lexical phonological forms are not retrieved from the lexicon as entities but are constructed from smaller parts (e.g., Fromkin, 1971; Shen, 1992). These analyses, however, do not provide direct evidence about the time course of phonological encoding, which is subjected to experimental studies in recent years (e.g., Levelt, Schriefers, Vorberg, Meyer, Pechmann, & Havinga, 1991; Meyer & Schriefers, 1991; Schriefers, Meyer, & Levelt, 1990). Mandarin 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. The same segmental templates having different tones correspond to different words or morphemes. Thus the suprasegmental information carried by the tone is an indispensable part of the phonological form of a word. Phonological encoding of a Chinese word must involve the activation of tonal information. The main purpose of this study is to use a picture-word interference paradigm to investigate the time course of tone activation in producing Chinese words.

A methodological issue in the study of speech production is how to experimentally constrain and manipulate the lexical concept that is to be activated in the lexicon. The picture-naming task and the associated picture-word interference paradigm provide us with useful tools. Lexical concepts can be manipulated by presenting pictures describing different common objects. The state of activation of various processes in picture naming can be probed by presenting words that share semantic or phonological properties with the lexical concepts or their corresponding phonological forms and by examining responses to the pictures or probe words. Varying the SOA (stimulus onset asynchrony) between presentation of pictures and presentation of probes allows us to trace the time course of semantic or phonological activation in speech production.

We report two experiments that used this picture-word interference paradigm to track the time course of tone activation in speech production of Mandarin Chinese words. A picture of *pear* (li2, number indicating the tone of the syllable) was presented, followed at some point of time by a Chinese character. This character, superimposed on the picture, shared either the segmental template and the tone (e.g., 离, li2, *leave*) or only the segmental template (e.g., 历, li4, *undergo*) with the name of the picture. Subjects were asked either to name the picture (Experiment 1) or to name the character (Experiment 2). By comparing response latencies and error rates in these two conditions, we should observe the effect of tonal information on the phonological activation of picture names. Moreover, by varying the SOA between picture presentation and character presentation, we could track the time course of tonal activation in speech production of picture names.

## Experiment 1

The experimental design is presented in Table 1. A target picture had one of the three types of interference characters superimposed on it in an experimental condition. One was a homophonic

character sharing both the segmental template and the tone with the picture name, one a character sharing only the segmental template, and one an unrelated character. Compared with the unrelated control condition, the picture naming should be facilitated by the homophone character since this character activated the same phonological representation used by the picture name. The character sharing the segmental template but not the tone with the picture name should also be able to facilitate the activation of segmental information of the picture name. The activation of tonal information of this character, however, could compete with the activation of tonal information of the picture name. The differences between homophone and segment conditions could, therefore, inform us about whether and when the activation of tonal information in speech production of picture names was influenced by the tone activation of the superimposed characters. The differences between segment and control conditions, on the other hand, provided us with information about the effect of segmental activation on speech production of the picture names.

Table 1 Experiment 1: Experimental Design and Sample Stimuli

| | **Interference Character** | | | **Target Picture** |
|---|---|---|---|---|
| | Homophone | Segment | Control | |
| | **离** | **历** | **晚** | **梨** |
| Pinyin | Li2 | li4 | wan3 | li2 |
| Gloss | *Leave* | *undergo* | *evening* | *Pear* |
| Frequency | 389 | 394 | 390 | |
| Stroke | 9.3 | 8.8 | 9.5 | |

Two SOAs were used, one 57 ms and one 200 ms. That is, pictures were presented either 57 ms or 200 ms before presentation of the characters. If tonal information of picture names was activated very early, we should observe differences between the homophone condition and the segment condition at the short SOA. If phonological encoding of tonal information was carried out rapidly and was finished before the characters were presented, presenting characters at the longer SOA may not have significance influences on the encoding process and consequently we may not find significant differences between the two conditions.

There were 54 critical pictures, each accompanied by three types of characters. All these pictures had monomorphemic, monosyllabic names. These morphemes, if written down, shared no orthographic similarities with the interference characters. The three types of interference characters were matched on frequency and visual complexity (in terms of the average number of strokes, see Table 1). The critical pictures and their interference characters were assigned, in a Latin square design, to three counter-balanced test versions. Twenty-four filler pictures, each having one interference character, were added to each of the test versions. Among the filler pictures, 4 had semantic relations with the interference characters. The names of other 4 pictures, if written down, were orthographically similar to the interference characters. The remaining 16 pictures had no relations with their interference characters.

Two groups of subjects were tested, 27 for the SOA of 57 ms and 33 for the SOA of 200 ms. Before the formal test, they received a booklet showing the pictures and their names. They also had 6 practice trials. Subjects were tested individually in a quiet room. They were sit in front of a computer screen and asked to name the pictures as quickly and as accurately as possible into a microphone linked to the computer. Response latencies were recorded from the onset of the picture to the initiation of naming. Naming errors were recorded by hand on preprinted sheets.

Table 2 Experiment 1: Mean Reaction Time (ms) and Error Percentage (in parentheses)

| SOA | Homophone | Segment | Control |
|---|---|---|---|
| 57 ms | 708 | 744 | 770 |
| | (3.8) | (3.3) | (3.8) |
| 200 ms | 695 | 707 | 734 |
| | (3.4) | (3.0) | (4.1) |

Mean reaction times, based on correct responses, and error percentages are presented in Table 2. Statistical analyses of the reaction times showed significant differences between the segment and control conditions at both SOAs (26 and 27 ms respectively), suggesting that the shared segmental information provided by characters facilitated the phonological activation of picture names. This phonological facilitation took place very early and was stable over the short and long SOAs. The analyses of tonal effects, however, found a different pattern. The difference between the homophone and segment conditions at the short SOA (36 ms) was significant, but the difference at the longer SOA (12 ms) was not. The analyses of error rates did not find significant effects.

Therefore, it seems that although both segmental and tonal information in activated very early in producing Chinese words, they have different time courses. Phonological encoding of tonal information takes place rapidly and enters next stage of processing very quickly while phonological encoding of segmental information lasts a relatively longer time. We suspect that this difference in time course of activation may due to the difference in complexity of tonal and segmental information. After all, Mandarin Chinese uses only four types of tones while the number of consonants and vowels in the language is much larger. Moreover, each syllable has only one tone but usually has more than one phonemes, which probably need to be retrieved separately and assembled.

**Experiment 2**

Experiment 2 aimed to collect converging evidence for the time course of tonal and segmental activation in the speech production of Chinese words. The experimental design was similar to Experiment 1, except that subjects were required to name characters instead of pictures. A picture was presented and after either 57 ms or 200 ms a characters was superimposed on it. Thus this experiment was more like a standard primed naming experiment (e.g., Zhou & Marslen-Wilson, 2000a) than like a picture-word interference experiment as Experiment 1. Nevertheless, we hoped this experiment to provide us with additional information concerning the time course of phonological activation in picture processing.

The experimental design and sample stimuli are presented in Table 3. There were four types of characters following a picture prime: characters sharing both the segmental template or sharing only the template with the name of the picture, and characters unrelated to the picture name. The two types of unrelated characters were created by re-pairing homophone or segment target characters with the pictures. There were 52 critical picture primes. They, together with their target characters were assigned to four counter-balanced test versions. Another 55 filler pictures were added to each version. After presenting a filler picture for 200 ms, a sign "?" was superimposed on the picture and subjects were required to name the pictures. In other words, when a picture had a character superimposed on it, subjects were supposed to name the character; but when a picture was accompanied by "?", they were supposed to the name the picture. The purpose of using this procedure was to ensure that subjects conducted phonological encoding for critical picture primes. The other procedures were the same as Experiment 1.

Table 3 Experiment 2: Experimental Design and Sample Stimuli

| | **Picture Prime** | **Target Character** | | | |
|---|---|---|---|---|---|
| | | Homophone | Homophone Control | Segment | Segment Control |
| | **熊** | **雄** | **叔** | **胸** | **译** |
| Pinyin | xiong2 | xiong2 | shu1 | xiong1 | Yi4 |
| Gloss | *Bear* | *grand* | *Uncle* | *chest* | *Translate* |
| Frequency | | 399 | 399 | 401 | 401 |
| Stroke | | 9.3 | 9.3 | 8.8 | 8.8 |

The experimental design and sample stimuli are presented in Table 3. There were four types of characters following a picture prime: characters sharing both the segmental template or sharing only the template with the name of the picture, and characters unrelated to the picture name. The two types of unrelated characters were created by re-pairing homophone or segment target characters with the pictures. There were 52 critical picture primes. They, together with their target characters were assigned to four counter-balanced test versions. Another 55 filler pictures were added to each version. After presenting a filler picture for 200 ms, a sign "?" was superimposed on the picture and subjects were required to name the pictures. In other words, when a picture had a character superimposed on it, subjects were supposed to name the character; but when a picture was accompanied by "?", they were supposed to the name the picture. The purpose of using this procedure was to ensure that subjects conducted phonological encoding for critical picture primes. The other procedures were the same as Experiment 1.

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

| SOA | Homophone | Segment | Control |
|---|---|---|---|
| 57 ms | 661 | 699 | 693 |
| | (1.9) | (1.8) | (5.7) |
| 200 ms | 638 | 695 | 688 |
| | (3.7) | (15.4) | (9.7) |

Because mean reaction times and error rates for the homophone controls and segment controls did not differ significantly from each other, these two conditions were combined. Their average reaction time and error rate, together with the mean reaction times and error rates in the two other conditions, are presented in Table 4.

Statistical analyses of reaction times and error rates revealed significant facilitatory effects for homophone targets at both the short and long SOAs, compared with either the control condition or the segment condition. Apparently the tone facilitatory effect, when the homophone and segment conditions were compared, became stronger as the SOA increased. The effects for segment targets, compared with the control condition, were different from those for homophone targets. At the short SOA, the reaction times to segment target showed a non-significant inhibitory effect. However, this effect was balanced by a facilitatory effect in error rates, suggesting a speed-accuracy tradeoff. At the longer SOA, both reaction times and error rates showed inhibition and this inhibitory effect reached significance in error rates.

The pattern of priming effects here was similar to Zhou and Marslen-Wilson (2000a), which used a character-character primed naming task and found facilitation for targets preceded by homophone characters and inhibition for the same targets preceded by semi-homophones differing in tones. Clearly, the presentation of picture primes here activated both the segmental and tonal

representations shared between the names of the pictures and the targets, facilitating phonological processing of the subsequently presented target characters. The phonological processing of segment targets, however, had to overcome the tone activation of picture names, creating competition between the two tone patterns. Apparently, this competition could cancel the facilitatory effect from the repeated activation of segmental representations.

### General Discussion

This study was conducted to examine the time course of the activation of tonal information in the speech production of Chinese words. Using a picture-word interference paradigm, Experiment 1 found that phonological encoding of tonal information is not only very early, but also completed faster than phonological encoding of segmental information. We suggest that the difference between phonological encoding of tonal and segmental information is due to the difference in their phonological complexity. The number of tones is much smaller than the number of phonemes in Mandarin Chinese. It is easier to encode or access tonal information and segmental information.

Clearly, this suggestion implies that tonal information is represented independently from the segmental information in the lexicon. It is possible that in speech production, tonal information and phonemes are retrieved independently but in parallel. Suprasegmental and segmental information are then combined to form syllables for further processing. Indeed, there is evidence from the studies of visual and spoken word recognition for the independent representation of tonal and segmental representations in the lexicon (Zhou & Marslen-Wilson, 2000b, 2000c), although further research is need to provide direct evidence for this independence in speech production.

Experiment 2 suggested that the tonal effect in the primed naming task increases as the SOA between picture prime and character target is increased. This seems contradictory to what was found in Experiment 1, i.e., the minimize of tonal effect over time. However, this may simply reflect the demand of experimental task. Phonological encoding of tonal information has to pass on to the next stage of processing in picture naming, i.e., phonetic encoding and articulation, while phonological activation of picture names in the primed naming task need not. In the later case the processing system may simply allow the phonological activation of picture names to accrue over time while waiting for the input of target characters. It was not a coincidence that both the facilitatory effect of tone-sharing and the inhibitory effect of tone competition increased over SOA in Experiment 2.

## Reference


Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechman, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review*, *98*, 122-142.

Meyer A S, & Schriefers H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Congnition*, *6*, 1146-1160.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in production: Picture-word interference studies. *Journal of Memory & Language*, *29*, 86-102.

Shen, J. X. (1992). The types of speech errors. *Zhongguo Yuwen*, *4*, 306-316.

Zhou, X., & Marslen-Wilson, W. (2000a). The relative time course of phonological and semantic activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Congnition*.

Zhou, X., & Marslen-Wilson, W. (2000b). *Lexical tone in the spoken word recognition of Chinese*. This volume.

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