# Perception of Isolated Tone2 words in Mandarin Chinese

Lei Xu & Shari R. Speer

Department of Linguistics Ohio State University, USA {xu; speer}@ling.ohio-state.edu

# ABSTRACT

Many tone3 words in Mandarin undergo "third tone sandhi"-a phonological rule that changes the first tone3 word in a tone3+tone3 sequence to a tone2 word. Thus, spoken tone2 words that have a tone3 counterpart are lexically ambiguous between a tone2 and a sandhi tone2 (underlyingly tone3). Thus while "yu2" in isolation means fish, in the sequence "yu2 hen3" it might mean either fish or rain. A cross modal priming experiment examined the processing of such potentially ambiguous words in isolation. Visual targets were Chinese characters of four kinds: identical to the auditory word, differentonly-in-tone, irrelevant to the auditory word or nonword. They were presented immediately after auditory primes of four types: tone2 word with tone3 counterpart, tone2 word without tone3 counterpart, tone3 word with tone2 counterpart, or tone3 word without tone3 counterpart. Listeners' response times were measured as they made word/non-word judgments. RTs were slower for potentially ambiguous words (tone2 words with tone3 counterparts) than for unambiguous words (tone2 words without tone3 counterparts). RTs to tone3 words with or without tone2 counterparts did not differ. The result suggests integration of tone and segmental information during word recognition, without recourse to a "tonal level", which predicts comparable RTs for all tone2s.

#### 1. Introduction

It is not clear how prosodic factors interact with segmental factors during the process of lexical access. Lexical tone, such as that found in Mandarin Chinese, is an excellent example of a prosodic, yet lexical feature that is actively involved in word recognition. Previous studies have suggested separable components of the access processes for lexical tone vs. vowel and consonant segments in Chinese. Some research has suggested that because tonal information is assigned at the syllable level, it must be distributed across multiple segments, and thus becomes available at a later point in the recognition process than segmental information [3]. Another proposal is that lexical tone is stored at a "toneme level", separated from the "phoneme level" [9]. This study investigates whether native speakers of Mandarin Chinese access an integrated tone/segment unit during word recognition or if they access the tonal/segmental information of a word separately.

# 2. Background

As a tonal language, Mandarin Chinese has four lexical tones and one neutral tone. Almost every syllable in Chinese utterance receives one tone, and segmentally identical syllables with different tones have very different meanings. Mandarin tones are usually referred to as Tone 1, Tone 2, Tone 3, and Tone 4. These four tones correspond to the tone shapes "high-level", "high-rising", "low-dipping" and "high-falling" respectively. [1]. For example, the same syllable "yu", when associated with the high-rising tone2, can mean "fish" (as in Figure 1), but when associated with the dipping tone3 (as in Figure 2) can mean "rain".



Figure 1. Wave form and fundamental frequency contour for the syllable Yu2 (fish)



Figure 2. Wave form and fundamental frequency contour for the syllable Yu3 雨 (rain)

In addition, the lexical tones are often changed by tone sandhi, a non obligatory, but highly productive phonological process that often drastically alters the phonetic shape of adjacent tones in connected speech [2]. The most frequently applied tone sandhi rule is third tone sandhi, which changes the first word in two tone3-word sequences into a tone2 word. For example, the word "yu3" is spoken in isolation with the dipping tone3, but is changed to a word with a high-rising tone2 "yu2" when followed by another tone3 word (as seen in Figure3).



Previously, we conducted two experiments that used a cross modal naming task to compare the processing of tonally ambiguous and unambiguous words in sentence contexts. Ambiguity was constructed by the operation of the above mentioned 3<sup>rd</sup> tone sandhi process, whereby the first word of a tone2-tone3 sequence may be underlyingly either a tone2 or tone3. Listeners heard sentence fragments that ended in a word with real tone2, sandhi tone2 (a tone2 that is resulted from sandhi, and has slight phonetic differences from a 'real' tone2), a sandhi sequence tone 2-3, or tone 3 word. Results showed that tone2 words, even when without a following tone3 sandhiinducing context, were ambiguous in sentence context. That is, naming times for visual semantic associates of both the tone2 and tone3 items were shorter for visual associates of the tone3 meaning than for associates of the tone2 meaning. However, all the tone2 stimuli used in these experiments had tone3 counterparts in Mandarin. For example, the word "yu3" (rain) is the tone3 counterpart of the tone2 word "yu2" (fish). However, there are many tone2 words in Mandarin that do not have a tone3 counterpart, and many tone3 words that do not have tone2 counterpart. Thus, there are two possible interpretations for these results. On the one hand, all words pronounced with a tone2 may be indeterminate in lexical tone, with either a tone2 or tone3 interpretation possible. On the other hand, words pronounced with a tone2 may be temporarily ambiguous only when they have a tone3 counterpart.

In the current lexical decision experiment, we looked to see if all tone2 words are potentially ambiguous. That is, when hearing an isolated tone2, do listeners activate both tone3 and tone2 representations? Or does this happen only for tone2 words that have a tone3 counterpart? If all tone2 words are ambiguous, this means that the locus of the ambiguity is at the level of tone processing and the listeners process the tonal information separately from the segmental information. If only tone2 with tone3 counterparts is ambiguous, this suggests that the locus of the effect is at the lexical level, and that the listeners process the lexical tone and the segments together in lexical lookup.

### 3. A Cross Modal Lexical Decision Experiment

# **3.1 Participants**

48 native speakers of Putonghua (Standard Mandarin) between 22 and 35 years old were recruited from the Ohio State University community and paid \$8 each for their participation. They were mostly graduate students from different departments of this University, but some participants were friends or family of the graduate students.

Ideally, the participants in an experiment that investigates tone sandhi should all speak the standard Beijing Mandarin. However, it was impossible to find so many native Beijing Mandarin speakers in Columbus, Ohio. Therefore, all available Mandarin speakers were included, and the participants came from almost all parts of China (e.g. Shan Dong Province, Zhe Jiang Province, Shan Xi Province, Hu Nan Province, Guang Xi Province). Thus they spoke different dialects of Chinese, which have very different set of lexical tones and Sandhi rules [6][2]. However, judging from their ages and educational backgrounds, they all had had plenty of exposure to the standard form of Mandarin-Putonghua. In addition, this same participant population was tested and gave reliable results for the previous cross-modal naming experiments. Therefore, the current experiment also used participants from all parts of China. A preliminary study on the production of Putonghua (Mandarin) by people from different dialect backgrounds showed that even those people whose Putonghua was heavily dialect accented produced a majority of correct sandhi tones (Xu, 2005).

#### **3.2 Design and Materials**

32 tone2 words with tone3 counterparts (e.g. yu2 "fish"), 32 tone2 words without tone3 counterparts (e.g. cha2 "investigate"), 32 tone3 words with tone2 counterparts (e.g. yu3 "rain"), and 32 tone3 words without tone2 counterparts (e.g. tie3 "iron") were selected to form the auditory stimuli. For the two ambiguous conditions, Tone2 counterparts served as Tone3 words and vice versa. Frequencies of all the words were counterbalanced across the four types of auditory stimuli. The average frequency of these words, both alone and as they appear word-initially in multi-syllabic Mandarin words, was calculated on the basis of Xiandai Hanyu pinlü cidian ("The Contemporary Chinese Word Frequency Dictionary" [5]. The average for tone2 words with tone3 counterparts in each list was 0.78, and the average for tone3 words with tone2 counterparts in each list was 0.79. The average of all homophones of tone2 words with tone 3 counterpart was 1.66, and the average of all homophones of tone3 words with tone2 counterparts was also 1.66.

Four types of visual targets were also constructed.: a character that represented the identical word as the auditory stimulus (e.g. yu2 "fish"), a character that represented a word

different from the auditory stimulus only in terms of tone (e.g. yu3 "rain"), a character that represented a word irrelevant to the auditory stimulus (e.g. ling2 "spirit"), or a non-word character. There were 16 different combinations of the auditory and visual stimulus types (as shown in Table 1).

		Visual conditions			
Word Type	Auditory Tone	Same Syllable Same Tone	Same Syllable Different Tone	Different Syllable & Tone	Nonword character
Sandhi Associate	T2 w/ t3	T2	T3 (sandhi assoc.)	T1/T4	No assoc. tone
	T3 w/ t2	Т3	T2	T1/T4	No assoc. tone
No Sandhi associate	T2 w/ot 3	T2	T1/T4	T1/T4	No assoc. tone
	T3 w/o t2	T3	T1/T4	T1/T4	No assoc. tone

Table 1. Experimental design with sixteen conditions.

There were also 52 fillers in the experiment. 18 were tone1 words, 18 were tone4 words, 8 were tone2 words, and another 8 were tone3 words. Also, there were 30 checking questions interspersed pseudo-randomly throughout the experiment that asked the subjects if they had heard a word (e.g. yu2) in the previous 3 or 4 trials. The purpose of these questions was to ensure that the subjects paid attention to the auditory stimuli.

Eight lists were generated. Each list contained 180 trials, with 128 experimental and 52 filler items, and 135 'word' vs. 45 'non-word' trials. The order of all auditory items was the same across the lists. The 128 experimental items rotated in a Latin square design through the 16 conditions across the 8 lists so that in different lists, the experimental items were in different conditions. Each list was seen by 6 participants.

A male native Beijing Mandarin speaker who was trained to read at a constant speed read the complete auditory stimulus set twice. The more natural sounding and clearer version of each item was chosen for the experiment. Digital recordings were made in a sound-proof booth, sampling at 22,050 Hz with 16-bit resolution using the sound editing computer software "Praat" (Boersma and Weenink, http://www.fon.hum.uva.nl/praat/).

## **3.3 Procedure and apparatus**

Before the experiment, the participants were asked to read a detailed description of the experimental procedure in Chinese from a piece of paper. In this information sheet, they were told that after 3 practice trials, they would hear 180 isolated words in Chinese. Immediately after the auditory word, a character in simplified Chinese would appear on the computer screen in front of them. If the character was a word of Chinese, the participant should press the "yes" button on the response box as soon as possible. If the visual target was a nonword, they should press the "no" button. Subjects were seated a comfortable distance from a 15 inch monitor, and put their index fingers over the yes button on the response box. Stimulus presentation and data collection were controlled by the E-prime program

(http://www.pstnet.com/products/e-prime/) The response time and correctness of the word/non-word answer were shown on the screen after each trial. Participants pressed the "space" bar on the keyboard to proceed to the next item. A cross symbol at the center of the screen appeared briefly each time after the participants pressed the "space" bar, indicating the coming of the next trial.

# **3.4 Predictions**

Results typical for a cross-modal priming experiment include a word-over-non-word response time advantage, as well as an advantage for identity between visual targets and the auditory primes. Therefore, when the visual target matches the auditory prime (is either identical or different only in tone) lexical decision times should be shorter than those for irrelevant or nonword visual targets. That is to say, the participants' response times to nonword or irrelevant visual targets should be longer than those for related targets.

In addition, since tone2 words in Mandarin are different from tone3 words in terms of length and tone shape, one should also be able to find an acoustic form effect. Tone2 words are usually shorter in absolute duration than tone3 words. Tone2 is the only tone in Mandarin that has initial rising shape. In contrast, the shape of tone3 falls at the beginning, and may be temporarily confused with tone4. Thus we predict that in isolation tone2 should be processed faster than tone3, leading to shorter response times for tone2 than tone3 words.

Finally, we predict an ambiguity effect if tone2 words with tone3 counterparts are more ambiguous than tone2 words without tone3 counterparts. That is, if the listeners process tonal and segmental information separately, there should be no difference in processing time between the two different types of tone2 words. However, if the tonal information and segmental information are integrated during the word recognition process, tone2 words with tone3 counterparts should be ambiguous, leading to two potential lexical entries. In contrast, tone2 words without tone3 counterparts should be unambiguous, and consistent with a single lexical entry. In this case, the listeners response times to tone2 words with tone3 counterparts should be longer than their response times to tone2 words without tone3 counterparts.

# **3.5 Results**

Mean response times for all 48 participants for each of the 16 different conditions are shown in Figure 4. A repeated measures analysis of variance (ANOVA) with subjects as the random variable was conducted on the mean response times for each of the eight subject groups. The factor 'audio stimulus' had four levels, corresponding to the spoken words: tone2 with tone3 counterpart, tone2 without tone3 counterpart, tone3 with tone2 counterpart, tone3 without tone2 counterpart. The factor 'visual target' also had four levels, corresponding to the character displayed on the screen: identical with the auditory stimuli, different with the auditory stimuli only in tone, irrelevant word and nonword. The analysis yielded a main effect of visual type (F[3,21] = 23.6, p < 0.01) and a main effect of auditory type (F[3,21] = 73, p < 0.01); and no significant interactions ( all Fs < 1).



Figure 4. Mean reaction time of the 16 conditions

A series of planned comparisons explored the main effects for visual and auditory conditions. For visual targets, planned comparison showed a clear priming effect. That is, identical and different-only-in-tone visual targets were processed more quickly than irrelevant ones and nonwords. For both auditory tone2 and tone3, response times to both visual associates of tone2 and tone3 were significantly shorter that those for irrelevant targets: (F[63,1] > 10).

For auditory primes, planned comparison showed a clear acoustic form effect. Consistent with our predictions, tone2 words, which had shorter durations, were responded to more quickly than tone3 words (F[63,1]>30, p < 0.01).

Finally, additional planned comparisons showed a robust ambiguity effect. When listeners heard tone2 words that had tone3 counterparts, their response times were longer than those for tone2 words that did not have tone3 counterparts (F[63,1] > 20, p < 0.01). However, when listeners heard tone3 words that had tone2 counterparts, their response times were not different from those for tone3 words that did not have tone2 counterparts (F[63,1] = 4.17, p = 0.95). That is, the unambiguous auditory tone3 word did not prime its tone2 counterpart.

More detailed comparisons showed that, if the visual target character matched the auditory word (the identity priming condition), listeners' response times were longer for potentially ambiguous words (tone2 words with tone3 counterparts) than for unambiguous words (tone2 words without tone3 counterparts) (F[63,1] >100, p < 0.01). In contrast, and consistent with the previous effect, no difference was found for different types of tone3 words (F[63,1] < 1, p > 0.5). When the listeners saw a visual word that was different from the auditory word only in tone, their response times were also remarkably slower for the potentially ambiguous tone2 with tone3 counterpart words than for the tone2 without tone3 counterpart words than for the tone2 without tone3 counterpart words than for the tone2 without tone3 counterpart words (F[63,1] > 50, p < 0.01). Again, there is no such difference for the two different types of tone3 words, (F[63,1] < 1).

#### 4. Discussion and conclusion

This experiment demonstrated a clear difference in native Mandarin listeners' processing of tone2 words with tone3 counterparts, words that are potentially ambiguous when pronounced with tone2, vs. tone2 words without tone3 counterparts, words that are not potentially ambiguous. Overall, listeners' response times were longer when they heard tone2 with tone3 counterpart words than when they heard tone2 words that had no tone3 counterpart. This suggests that tone2 words with tone3 counterparts are ambiguous when spoken in isolation, while tone2 words without tone3 counterparts are not. The ambiguity stems from the fact that certain tone3 words frequently participate in the third tone sandhi, and where they are pronounced with a tone2. Thus, when listeners heard a tone2 word with tone3 counterpart, we argue that they accessed both the underlying a tone2 word, and the underlyingly tone3 word. Importantly, we found that tone2 words without tone3 counterparts were not ambiguous to the listeners. This suggests that when listeners heard these words, they accessed only the underlyingly tone3 lexical item. These findings are consistent with a word processing model where prosodic tonal information is not processed at a separate "toneme level" from the "phoneme level" during lexical access. Instead, lexical tone is an integral component of the auditory signal used in Mandarin word recognition.

#### REFERENCES

- [1] Chao, Y.-R. (1968). *A Grammar of Spoken Chinese*. (Berkeley: University of California Press).
- [2] Chen, M-Y. (2000). Tone Sandhi: Patterns Across Chinese Dialects (Cambridge, UK:Cambridge University Press).
- [3] Cutler, A., & Chen, H. C. (1997). Lexical tone in cantonese spoken-word processing. *Perception & Psychophysics*, 59 (2), 165-179.
- [4] Speer, S. R. & Xu. L. (2005) Ambiguous lexical tone process during mandarin sentence comprehension – Evidence from eye-movement experiment, Architectures and Mechanisms for Language Processing (AMLAP), Ghent, Belgium.
- [5] Wang, H. (1990). Xiandai Hanyu Pinlu Cidian (Contemporary Chinese Word Frequency Dictionary). Beijing: Beijing yu yan xua yuan chu ban she (Beijing Language Institute Publishing House).
- [6] Wang, J. (2003). Rhythmic grouping, Tone sandhi and Stress in Beijing Mandarin Ph.D. Dissertation, Macquarie University, Sydney, Australia.
- [7] Xu. L. (2005). A pilot Study on dialect-accented Putonghua. Manuscript, Ohio State University.
- [8] Xu, L., & Speer, S. R. (2004) Lexical tone and lexical access in Mandarin, *Psychonomics*.
- [9] Ye, Y. & Connine, C. M. (1999). Processing vowels and lexical tone in Mandarin Chinese. *Language and Cognitive Processes*, 14, 609-630.