Effects of frequency, repetition and prosodic location on ambiguous Mandarin word production

Seth Wiener [1], *Shari R. Speer* [2], *Claire Shank* [1]

[1] Department of East Asian Languages and Literatures, The Ohio State University, USA
[2] Department of Linguistics, The Ohio State University, USA
wiener.24@osu.edu, speer@ling.osu.edu, shank.86@osu.edu

Abstract

Recent findings from large corpus studies of spontaneous English speech show that the predictability of a word is negatively correlated with the duration of its pronunciation. This predictability effect extends to word repetition in discourse, where words shorten as they are re-used, to word frequency, with shorter pronunciations for high frequency than low frequency words, and to homophones, where the higher frequency meaning has a shorter duration than its lower frequency counterpart. Standard Mandarin, due to its relatively unmarked phonology and morphology, presents the potential for a large degree of homophony and resulting lexical ambiguity. We test whether predictability factors such as word frequency, repetition, and prosodic phrasal position influence the duration of Mandarin homophones in simple sentence reading. Native monolingual Mandarin speakers read phonologically identical sentence pairs, which differed only in semantics and orthographic representation. Each pair member contained either an ambiguous high or low frequency word; these ambiguous pairs were placed in one of four different prosodic locations throughout the stimuli. Sentences were read multiple times with durations measured for each ambiguous token. The results corroborate previous findings from corpus analyses showing duration reduction with repeated mention and effects of relative meaning frequency on homophone durations. In addition, we found differences in duration due to prosodic phrasal position, but no interaction of prosodic position with other forms of word predictability. Our findings indicate that speakers produce subtle durational cues to indicate a range of information about the words they use. Such information may be useful to listeners who may use this subtle information to minimize ambiguities.

Index Terms: speech prosody, word frequency, Mandarin Chinese, lexical ambiguity, word duration

1. Introduction

Compared with English, Standard Mandarin Chinese has a considerably smaller syllable inventory. Ignoring lexical tone, there are only around 400 syllables in modern use; accounting for tone, the number increases to approximately 1,3000 syllables [1]. Given the high likelihood of homophony in the spoken language, it stands that Mandarin speakers may encounter lexical ambiguity more often than speakers of languages with larger syllable inventories. What factors in the spoken language signal are available to help resolve these ambiguities? Although listeners may believe they rely only on the semantic and syntactic content of the sentence context to determine the meaning of a homophone, speakers produce additional regular but subtle cues to the identity of words. Previous research has shown that the frequency of a word's overall occurrence in the language is correlated with its pronounced duration in spontaneous speech, such that more frequent words are shorter than less frequent ones, contain reduced vowels and overall reduced segment duration, and even segment deletion [2,3]. In addition, the pronunciation of the more-frequent meaning of a homophone pair (e.g. the English *time-thyme*) is shorter than its less frequent counterpart. These effects, found in English corpus studies, were robust even when factors such as predictability from sentence context, local speaking rate, and syntactic category were controlled for statistically [4]. Word duration is also affected by the number of times a speaker uses the word in a single conversation, with additional mentions reducing duration, giving potential cues to whether the referent of the word is the same as for its previous use (i.e. "discourse old") [e.g. 5]. All of these findings suggest that the relative duration of a spoken homophone might provide information to reduce the ambiguity of its meaning. However, prosodic aspects of the spoken signal, such as phrase-final lengthening, also have an impact on the absolute and relative duration of words, and may mitigate predictability-based duration effects.

Mandarin is an excellent choice for examining the interaction of prosodic phrasal position, lexical frequency and frequency of mention on speakers' word production. Although previous studies have shown that prosodic cues may be used to disambiguate utterances of spoken Mandarin, they have focused largely on the contribution of prosodic phrasing to the recovery of syntactic constituency [e.g. 6], or to interactions of prosodic phrase locations with the operation of tone sandhi rules that may lead to tonally-based lexical ambiguity [e.g. 7]. To our knowledge, there have been no studies of the interaction of prosodic position, repetition and word frequency on the produced duration of spoken words.

2. Background

Findings of subtle durational differences that depend on the predictability of words in spoken sequences have been found primarily in corpus studies, where a large body of language can be examined and factors like successive mention and prosodic position can be measured. Studies using corpus data, however, are restricted to productions available from spontaneous speech, often hindering the comparison of matched tokens (such as homophones) in identical prosodic positions. Laboratory data collection allows for controlled comparisons of tokens of interest by manipulating the number of mentions, the prosodic location as well as the order presented. Previous work on the production of homophone pairs [e.g. 4] is in a sense limited to the more-frequently occurring homophones available in a given corpus and to the prosodic position(s) where the homophones happened to have appeared. Similarly, the number of repeated mentions and the prosodic position of each repeated target word in studies of "old" versus "new" information [e.g. 5] are limited in that both the number of mentions and the prosodic positioning are serendipitous in corpus data. By designing a set of stimuli in

which these confounds are reduced, the duration of lexical ambiguities can be measured in a series of specifically controlled homophone pairs, in which prosodic positions, number of mentions and word frequency are all manipulated.

Given the relatively unmarked phonology and morphology of Mandarin Chinese – a (C)(G)V(X) syllable structure in which (G) denotes optional glide and (X) denotes optional nasal or glide [1] – the potential for lexical ambiguity is fairly large and could easily be manipulated for laboratory data. It is established that the majority of modern Standard Mandarin words are disyllabic – percentages range, but it is generally agreed that at least two-thirds are disyllabic, with each syllable carrying a lexical tone (or the second syllable being weakly stressed or 'neutral tone') [1]. Neutral toned syllables, however, cannot occur in isolation. Thus a neutral tone syllable is necessary joined together (i.e. a foot with the preceding full-toned syllable) [8]. In this study, we examined disyllabic Mandarin words with two fully-realized lexical tones. The four lexical tones are denoted as '1' high level tone, '2' rising tone, '3' low dipping tone, and '4' high falling tone. Given these parameters, an example disyllabic word like 'zheng4 shi4' has at least six different orthographic and semantic representations. Due to frequency and contextual clues, rarely is one of the meanings such as "government affairs" 政事 confused with another meaning such as "exactly" 正是. Prosodic structure in Mandarin, as characterized by the Pan-Mandarin ToBI transcription [8], contains eight tiers. For the purpose of this study, we are interested in tier 7, indicating six levels of break indices. The reduced syllable boundary (B0) is not examined in this study. We examine the remaining five: the normal syllable boundary (B1), considered the default case for prosodic words, the minor (B2) and major (B3) phrase boundaries and the prosodic group boundaries; reset of pitch (B4) and pause (B5). By limiting contextual clues and exploring varying frequency interactions, we believe Mandarin provides the ideal linguistic toolbox to examine the interaction of prosodic phrasal position, lexical frequency and frequency of mention on speakers' word production.

3. Lexical Ambiguity in Mandarin Speech

3.1. Participants

Seven native, monolingual, Standard Mandarin (Putonghua) speakers between 25 and 45 years old were recruited from the greater Columbus, Ohio area. Participants had lived in the U.S. for less than one year, and none of them were able to communicate in English beyond a few words. They came from five different provinces (all but one participant was from a province within Norman's Northern Mandarin group [9]; the one participant that came from Norman's Southern Mandarin group was from Shanghai and did not speak Wu) and therefore were exposed to different dialects and non-standard Mandarin varieties. All participants, however, were educated in Putonghua and spoke Putonghua with their family in the United States. Participants were specifically recruited for being monolingual Putonghua speakers. In order to ensure similar productions among participants (e.g. no rhotacized tokens), each speaker was first given a short oral interview by the first author (a fluent non-native Mandarin speaker).

3.2. Design and Materials

Eighteen lexically ambiguous but syntactically identical Mandarin utterance pairs were constructed to felicitously contain either meaning of an ambiguous disyllabic word. Sentences contained no contextual or syntactic clues so that when spoken alone, there were no apparent means of ambiguity resolution. The mean number of syllables per utterance was 7.6.

The ambiguous targets were placed in four different prosodic positions: phrase initial (either ToBI break indices B1 or B2), mid-phrase (either B2 or B3), phrase-final (either B4 or B5), and before a neutral tone (qingsheng) phrase-final particle such as 'ma' (interrogative particle) or the verbal suffix 'le' (either B3 or B4). Prosodic location was manipulated across tokens so that these differed for each location group. It is important to note that even though questions were used in the stimuli, ambiguous tokens were not placed in the phrase final position in question utterances (thus potential rising intonation was not used by speakers). All frequency counts were taken from the SUBTLEX-CH Chinese word and character frequency corpus [10]. Token frequencies for each disyllabic word rather than character frequencies were used. Each homophonous pair consisted of one high frequency and one low frequency word in which "high" (H) and "low" (L) were classified as relative to one another rather than an arbitrary high/low cutoff point. All frequency pairings were crosschecked with the Xiandai Hanyu pinlü cidian ("The Contemporary Chinese Word Frequency Dictionary") [11].

For example, participants were presented with the ambiguous sentence, 'nan2 shi4 dou1 zhe4 yang4.' This utterance contains the ambiguous word 'nan2 shi4' which can mean either "men" 男士 (high frequency token with 345 mentions) or "difficult things" 难事 (low frequency token with 72 mentions). Therefore these two homophonous sentences are processed as either "all men are like this" or "all difficult things are like this."

Ambiguous pairs were presented first as a randomized high/low pair in which each utterance was read aloud and then followed immediately by its high or low frequency counterpart. This sequence is denoted as AB; HL vs. LH order was counterbalanced across two lists. After all eighteen pairs were read through once, the participants read through the randomized ambiguous pairs again, this time repeating the same frequency target two times before reading the alternate high or low frequency word. This sequence is denoted as AABB; again HHLL and LLHH sequences were counterbalanced across lists. Each homophonous target was produced as three sequences for a total of six renditions: three as the high frequency and three as the low frequency. All presentation sequences were counterbalanced so that each ambiguous pair was presented to half of the participants with the high frequency word coming first and to half of the participants with the low frequency word coming first.

3.3. Procedure

Participants were given spoken instructions in Mandarin by the experimenter, who asked them to read at a normal speaking rate. A written list of all eighteen sentences (characters only, no pinyin) and one practice sentence was given to each participant, who read through the list at his or her pace while being recorded with a handheld digital recorder. All digital recordings were sampled at 44,100 Hz. Participation was voluntary and lasted approximately eight minutes.

All recordings were analyzed using the sound editing software "Praat" [12]. Using Praat text grids, each ambiguous token was annotated and measured by hand and extracted using a Praat script. In total, 756 tokens were measured from the seven participants.

3.4. Predictions

If the relative frequency of a homophone pair member influences its spoken duration in laboratory speech, we expect to find shorter durations for the more-frequent meaning of the homophone than for its less frequent counterpart. This predication is based on findings for English (e.g. *time-thyme*) [4], which we anticipate our Mandarin examples will duplicate. Similarly, we expect repetition to reduce duration. Across the list, durations should be the longest for the initial mention of a word, and shorten with successive repetition. Repetition effects may have different sources in the language production system. For example, word durations may shorten due to repetition of motor routines used to pronounce the phones of a word, but they may also shorten due to reduction of time for word recognition and lexical access. If repetition effects have a source primarily in motor-routine aspects of syllable pronunciation, we expect each repetition of the same homophone to be shorter than the one that preceded it, regardless of meaning. If factors associated with repetition of a particular word control duration, we expect pronunciation durations to shorten for each repetition of the same homophone meaning. If duration effects of prosodic position, such as phrase-final lengthening, overwhelm the subtle effects of word frequency, we might expect to find that duration differences in some prosodic positions will be neutralized

3.5. Results

Measured durations were included in repeated measures ANOVAs with tokens and subjects as random variables. Factors included pair member (Low or High frequency), sequence type (LH, HL), rendition number (1-6), and prosodic location (initial, mid, pre-particle and final). A total of four word tokens from two subjects were removed as outliers; all were beyond 3.5 sds from the mean for both subject and token. In general, results indicate that the repeated mention of a word shortens its duration, and that lower frequency pair members tend to be longer than their higher frequency counterparts; the frequency effect is somewhat conditioned by sequencing of high and low frequency members. Although we found effects of prosodic location, frequency and number of renditions did not interact with prosody.

We found a numerically small but robust effect of pair member frequency on duration, such that higher-frequency homophone pair members were shorter than their low frequency counterparts. In addition, the sequence in which a pair was pronounced (whether H preceded L or the reverse) influenced duration differences. When low frequency pair members were pronounced first in AB pairs (LH), their durations were longer than when they were pronounced second (HL) (Fig 1). These effects produced a significant two-way interaction of pair member and sequence type, ($F_1(1,6) = 9.68$, $p < .05$; $F_2(1,16) = 5.11$, $p < .05$).

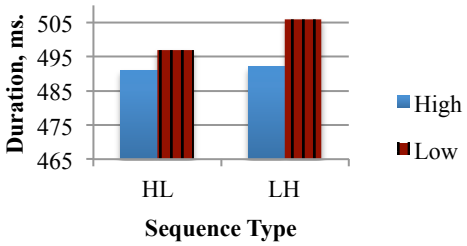

Figure 1: *Mean durations for High and Low frequency members of homophone pairs in HL and LH sequences (renditions 1 and 2, AB trials).*

When high frequency pair members were pronounced first in AABB pairs (HHLL), their durations were shorter than their low frequency counterparts, and the immediate second pronunciation of the same pair member was speeded as compared to the first (Fig 2). In contrast, when low frequency pair members were pronounced first in AABB pairs (LLHH), although the immediate second pronunciation of the same pair member was again speeded as compared to the first, there were no differences between high vs. low frequency pair members (Fig 2). These effects produced a three-way interaction of pair member frequency, sequence type and repetition number ($F_1(1,6) = 12.09$, $p < .05$; $F_2(1,17) = 6.51$, $p < .05$); the two-way interaction of pair member frequency and repetition number approached significance by tokens $F_2(1,16) = 4.38$, $p = .053$).

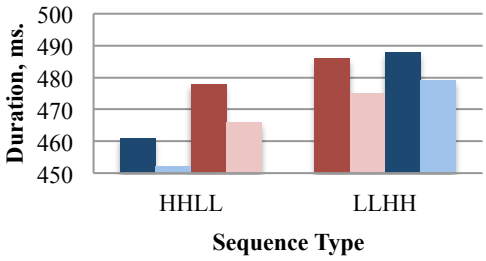

Figure 2: *Mean durations for High and Low frequency members of homophone pairs in HHLL and LLHH sequences (renditions 3, 4, 5, and 6, AABB trials).*

Although we found a main effect of prosodic position as a within-subjects variable, there was no effect in the items analysis, where prosodic position was manipulated between tokens ($F_1(1,6) = 12.07$, $p < .05$; $F_2 < .01$). There were no significant interactions of prosodic position with other factors (all $Fs<1$). Figure 3 shows that words pronounced in pre-particle position were shorter than those at other positions, and that the durational differences between the high and low frequency homophones do not differ significantly at any prosodic position. Because we were primarily interested in whether frequency differences between homophone pair members would be modulated by prosodic position, we did not manipulate prosodic position factorially across tokens. Thus it is less of a concern that phrase-final tokens were not longer than other prosodic positions, as this may indicate length differences across token groups.

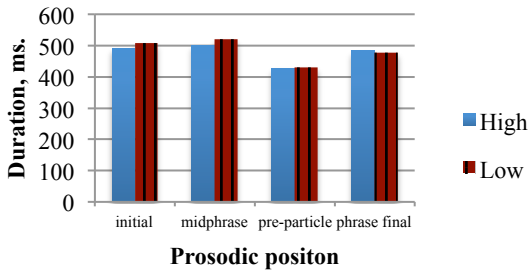

Figure 3: *Mean duration of High and Low frequency homophone pairs in all four prosodic positions.*

4. Discussion and Conclusion

We observed durational differences across all homophone frequency pairs in which the high frequency token was consistently shorter in duration than its low frequency counterpart. This finding corroborates results of high and low frequency paired English homophones in corpus data [4]. We also observed a repetition number effect similar to discourse data from [5] in which durational reduction occurs after each repetition of the token. In addition, we found that the order in which high and low frequency pair members were produced had an effect on durational indications of frequency. This effect has not been noted in corpus studies, and points out the advantage of the control available in laboratory studies. In the AB trials, we observed that while high frequency pair members were consistently shorter than low frequency members, producing the high-frequency token first (HL sequence) allowed shorter productions of the low-frequency pair member (as compared to the LH sequence). These results suggest that the mechanism that underlies shorter productions is more than just a motor-routine pronunciation effect; it may also reflect cognitive retrieval of lemma-level information about word meaning. Assembly of this information and pronunciation routines would thus be slower for the low frequency counterpart when it is encountered first; when it is encountered second, common pronunciation routines may be available from the homophonous high-frequency pair. Data for AABB trials (in Figure 2) suggests that as the low frequency meaning is repeated in initial position for LLHH trials, any pronunciation advantage from a lexical level is reduced for high frequency pair members, so that durational differences due to frequency disappear. We can only speculate about how this effect might function during discourse. It may reflect the availability of both meanings in short term memory, perhaps competing for resources needed for speech production. The effect may also reflect evidence of frequency re-ordering of the meanings of the ambiguous words similar to proposed re-ordered access models [13].

Our focus on whether frequency differences between homophone pair members would interact with prosodic position led us to manipulate prosodic position between items. A comparison of absolute length of the words at the four prosodic positions, therefore, is not appropriate as each position reflects a different set of items. Based on our data, there is no apparent interaction of frequency-based durational differences with prosodic position. Neither expected sentence final lengthening nor the reduction associated with foot-creation by the addition of a neutral-toned particle was able to overwhelm or even modulate frequency effects on duration. However, additional experimentation with a design that tests matched lexical items at multiple prosodic locations is needed to explore possible interactions of duration effects due to lexical frequency with those due to prosodic phrasal position.

We argue that speakers produce subtle durational cues to indicate a range of information about the words they use, and that these effects, though small, are robust enough to appear in a simple 8-minute laboratory experiment. If listeners are sensitive to such small differences as they accrue over spoken language processing, they may be able to use the information to minimize ambiguities. If it is the case that these durational cues are gradient across different languages, then in a language like Mandarin with a fairly reduced syllable inventory, these durational cues could serve as vital components of the spoken signal.

5. Acknowledgements

The authors would like to thank to Shih Ya-ting for helping pilot the stimuli, Zeng Zhini for helping find recruits and assisting in the recording, and Kodi Weatherholtz for sharing R code and providing feedback on the project. Any errors that remain are those of the authors'.

6. References

[1] S. Duanmu, *The Phonology of Standard Chinese*. New York: Oxford University Press, 2000.

[2] D. Jurafsky, A. Bell, M. Gregory, W.D. Raymond, "Probabilistic relations between words: Evidence from reduction in lexical production," *Frequency and the emergence of linguistic structure. 2000* Sec. *3,* pp. *229-254.*

[3] A. Bell, J.M. Brenier, M. Gregory, "Predictability effects on durations of content and function words in conversational English," *Journal of Memory and Language,* vol. *60,* pp. *92-111,* 2009.

[4] S. Gahl, " Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech," *Language*, Vol. *82*, No. *3*, pp. *474-496*, Sept, 2008.

[5] C.A. Fowler, J. Hossum, "Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction." *Journal of Memory and Language, 26*, pp. 489-504, 1987.

[6] S. S. Xiaonan, "The Use of Prosody in Disambiguation in Mandarin," *Phonetica*, Vol. *4*, No. *50*, pp. 261-271, 1993.

[7] S.R. Speer, C.-L. Shih, M. L. Slowiaczek, "Prosodic structure in language comprehension: Evidence from tone sandhi in Mandarin." *Language and Speech, 32,* 1989.

[8] S-H. Peng, M.K.M. Chan, C-Y. Tseng, T. Huang, O.J. Lee, M.E. Beckman, "Towards a Pan-Mandarin System for Prosodic Transcription," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.A. Jun (editor). Oxford, UK: Oxford University Press, pp. 230-270, 2005.

[9] J. Norman, *Chinese*. Cambridge, UK: Cambridge University Press, 1998.

[10] Q. Cai and M. Brysbaert, SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos ONE, 5(6), e10729,* 2010.

[11] H. Wang, *Xiandai Hanyu Pinlu Cidian* (Contemporary Chinese Word Frequency Dictionary). Beijing: Beijing yu yan xua yuan chu ban she (Beijing Language Institute Publishing House), 1990.

[12] P. Boersma and D. Weenink, *Praat: doing phonetics by computer,* 2011, software available at http://www.praat.org

[13] S.A. Duffy, R.K. Morris, K. Rayner, "Lexical ambiguity and fixation times in reading." *Journal of Memory and Language, 27*, pp. 429-446, 1988.