Interspeech 2018
2-6 September 2018, Hyderabad

Effects of homophone density on spoken word recognition in Mandarin Chinese

Bhamini Sharma[1]

[1]The Hong Kong Polytechnic University, Hong Kong SAR
bhamini.sharma@connect.polyu.hk

Abstract

Homophones, words that sound same, influence spoken word recognition. Whether the effects of homophone density (i.e., number of same-sounding words) on spoken word recognition are facilitatory or inhibitory or complex is a matter of ongoing debate. In addition, there are limited studies investigating the effects of homophone density, probably due to paucity of homophones in the examined languages (e.g., English). In comparison, languages such as Mandarin Chinese have abundant homophony that makes it a suitable tool to investigate the effects of homophone density. In the current study, an auditory naming task was conducted using Mandarin Chinese to investigate the effects of homophone density on spoken word recognition. Using mixed modeling, a significant inhibitory effect of homophone density ($\beta = 0.0098$, $t = 2.10$) on reaction time was found. Participants were slower in naming words with high homophone density, possibly due to competition posed by more number of homophones, as compared to the words with low homophone density. Further, an interaction between homophone density and syllable frequency was found i.e., for high syllable frequency, homophone density effects were inhibitory but for low syllable frequency, the inhibitory effect was reduced. Taken together, the effects of homophone density are not straightforward but complex.

Index Terms: spoken word recognition, homophone density, auditory word naming

1. Introduction

Mental lexicon is purportedly a complex structure where words are connected through various similarities including phonological, semantic, or orthographical similarity. Words with similar pronunciation (known as phonological neighbors) are well known for their effects on spoken word recognition [1]–[3]. However, the effect of words that sound the same (known as homophones) on spoken word recognition is less investigated.

Mostly commonly, when two or more words have same pronunciation but differ in their meaning and spelling, they are referred to as homophones. Homophone density refers to number of same sounding words that a word have. There are limited studies [4]–[15] that have investigated the effect of homophone density on spoken word recognition. This is probably due to the fact that many of the languages studied (e.g., English) are not rich in homophony to test the effects of homophone density. In comparison, languages such as Mandarin Chinese, due to plenty of homophones in the language, could provide a suitable tool in studying the role of homophone density in spoken word recognition. In the current study, Mandarin Chinese was used as a vehicle to study the effects of homophone density on spoken words recognition.

1.1. Mandarin syllables

Mandarin Chinese is a tone language. It consists of set of monosyllabic morphemes that combine to form compound words. Each monosyllabic morpheme is associated with at least one Chinese character (orthographic form). Mandarin Chinese has a relatively small syllable inventory of around 1300 monosyllabic morphemes, including all four tones, due to phonotactic constraints in the language. Due to the limited number of monosyllabic morpheme words in Mandarin Chinese, it leads to generous amount of homophones ranging from at least 2 to more than 40 homophones. Homophones in Mandarin Chinese can be defined as two or more monosyllables that share same syllable and tone but differ in meaning and spelling. Thus, Mandarin Chinese provides a good way to investigate the role of homophone density on spoken word recognition.

1.2. Effects of homophone density in Mandarin

Studies focusing on the effects of homophone density have reported mixed findings, facilitatory or inhibitory, on spoken word recognition [4], [11]–[15].

Wang et al. [4] used an auditory lexical decision task to examine the effects of homophone density in Mandarin monosyllables. They found that monosyllables with high homophone density were responded slower and less accurately compared to monosyllables with low homophone density. They found an inhibitory effect of homophone density. Based on their findings, syllables with high homophone density activate representations from more homophones resulting in stronger competition in processing compared to low homophone density monosyllables.

In comparison, Yao and Sharma [13] investigated the effects of homophone density using an auditory lexical decision task in Mandarin Chinese. Their findings revealed a facilitatory effect of homophone density on accuracy. Monosyllables with high homophone density were responded more accurately as compared to monosyllables with low homophone density.

However Li, Wang and Li [14] used adapted dictation paradigm and cross modality matching to explore the role of word frequency and homophone family. They found that words with higher word frequency were activated strongly while the words with lower word frequency were suppressed within the homophone families. Their study confirms the existence of word frequency effects within each homophone family. Also, they found that the activation of homophone representations is different between high, mid or low frequency words within the family.

Further, Li et al. [5] using cross-modal matching task and visual lexical decision found significant interactions between word frequency and homophone family size. High word

 10.21437/Interspeech.2018-2114

frequency within a homophone family facilitates the word processing whereas low word frequency within a homophone family suppresses the word processing. In addition to this they also found that homophone family size facilitated the recognition of low-frequency homophone words but not high-frequency homophone words. In a big homophone family a word with low word frequency will be recognized faster than the word with low word frequency from a small homophone family.

Sharma and Yao (to appear) [15] further explored the interactions between homophone density and maximum homophone frequency (highest frequency monosyllable among all the homophones in a homophone family) using an auditory lexical decision task. They found an inhibitory effect of homophone density when maximum homophone frequency is high whereas a facilitatory effect of homophone density was seen when maximum homophone frequency was low.

In order to further understand the effects of homophone density on spoken word recognition and to contribute towards settling the debate on whether the effects of homophone density are facilitatory or inhibitory or more complex on spoken word recognition. Also, to confirm the findings we [15] got from auditory lexical decision task, the current study was conducted using an auditory naming paradigm. If the effects of homophone density on spoken word recognition turn out to be inhibitory, it can be predicted that the monosyllables with high homophone density would be named slower as compared to those with low homophone density, probably due to the increased competition among more number of homophone in the spoken word recognition than the less number of homophones. Alternatively, if the effect of homophone density on spoken word recognition turns out to be facilitatory, it can be predicted that the monosyllables with high homophone density would be named faster than those with low homophone density, probably due to enhancement of spoken word recognition by the existence of more number of homophones than less number of homophones. On the other hand, if the effects of homophone density on spoken word recognition turn out to be complex, an interaction of homophone density with the syllable frequency is predicted.

2. Method

2.1. Participants

Thirty native Mandarin speakers (16M, 14F, mean age = 19.16; SD = 1.02) participated in the present study. All the participants were born and raised in Mainland China and reported no speech and hearing problems.

2.2. Stimuli

The stimuli consisted of 1259 real monosyllables. This set of Mandarin monosyllables was extracted from an online Chinese dictionary. This set of stimuli consisted of all the possible Mandarin monosyllables that exist in Mandarin Chinese. Each Mandarin monosyllable can be associated with at least one character (orthographic form) in Mandarin Chinese. A female native Mandarin speaker recorded all the stimuli in an acoustically treated room using a uni-directional microphone, routed to Digi design recording system. All the stimuli were normalized for intensity at 70 dB using Praat. Mean duration of stimuli was 624.5 ms (SD = 103.8 ms). The stimuli were evenly divided into five blocks and each block was presented to 6 participants. Items within each block were randomly presented across the participants. Five items were shared across blocks to evaluate for cross-block consistency.

2.3. Procedure

Auditory word naming experiment was conducted using OpenSesame version 3.1 [16] on a Philips desktop connected to M-audio interface and a multichannel recorder for headphone output and dual channel recording. Channel 1 was dedicated to headphone output (stimuli) while channel 2 recorded the participant utterances. Each trial started with a button-press followed by a fixation-cross for 500 ms followed by the auditory stimulus. Participants were instructed to repeat the stimuli as quick and as accurately as possible.

Each experiment session consisted of 15 practice trials followed by a block of test items presented in a random order. None of the practice trials were repeated in the main experiment. No feedback was provided to the participants. Responses from the participants were time locked to 2000 ms after the presentation of the stimulus. All experimental sessions were conducted in a quiet room. Reaction time was recorded as duration between the onset of the stimulus and the onset of the response from the participant. Each session lasted for not more than 30 minutes, including breaks in between.

2.4. Analysis

SUBTLEX-CH corpus [17] was used to obtain lexical measures like syllable frequency and homophone density. Table 1 summarizes the lexical measures. Mixed-effect model was used to do analysis. To control for phoneme length, only 3-phoneme syllables were analyzed. Mixed-effect model was built using (log) reaction time as the outcome variable and (log) homophone density as the predictor variables with (log) syllable frequency, (log) stimulus duration and (lexical) tone as controlled variables. Variables were log transformed for a more normal distribution and centered before entering into the models. Random effects of subject and item were included in the model. The models were tested for significance using the likelihood ratio test. In the likelihood ratio test, models were constructed with and without single critical variable to test for significance. All the analyses were done in R [18] using lme4 package [19].

Table 1: *Summary of lexical measures (syllable frequency is number of occurrences per million words).*

Variables	Range (min, max)	Mean	Median	SD
Homophone density	(1, 19)	4.50	4.00	2.99
Syllable frequency	(0.15, 7997.64)	608.10	190.66	1049.98

3. Results

Mean reaction time for 3-phoneme syllables was 715.69 ms (SD = 136.57 ms). Mixed modeling with effects of homophone density, controlled for syllable frequency, stimulus duration and lexical tone on the reaction time revealed significant main effects of homophone density based on likelihood ratio test ($\beta = 0.0098$, $t = 2.10$, $\chi^2 = 4.6614$, $p = 0.03$). Homophone density showed a significant inhibitory effect on the reaction time for naming. Monosyllables with high homophone density were named slower compared to monosyllables with low homophone density. Among the controlled variables, syllable frequency showed a significant facilitatory effect on the reaction time for naming. Monosyllables with high syllable frequency were named faster than monosyllables with low syllable frequency. Syllables with longer stimulus duration were produced slower than shorter stimulus duration stimuli. Syllables with Tones 3 and 4 were named slower on an average. See Table 2 for summary of model results with homophone density effects.

Table 2: *Summary of the model with homophone density effects.*

Variables	Estimate (β)	t-value
(log)Homophone density	0.0098	2.10
(log)Syllable frequency	-0.0032	-2.17
(log)Stimulus duration	0.3522	16.20
Tone 2	-0.0057	-0.78
Tone 3	0.0262	3.29
Tone 4	0.0197	2.60

To probe further, another model was constructed similar to the earlier model but focusing on the interactions of homophone density and syllable frequency on spoken word recognition. In this model, interaction between homophone density and syllable frequency were tested with stimulus duration and lexical tone as control variables on reaction time. Model results revealed a significant interaction between homophone density and syllable frequency ($\beta = 0.0047$, $t = 2.24$). Similar effects were seen for control variables as seen in the aforementioned model. Table 3 provides a summary of model with interactions between homophone density and syllable frequency. Further, the two models (one without interactions and other with homophone density and syllable frequency interactions) were compared using likelihood ratio test. The likelihood ratio test revealed a significant (($\chi^2 = 4.9571$, $p = 0.02$) difference between these two models. This confirmed the significance of interactions between homophone density and syllable frequency. The interaction between syllable frequency and homophone density along with the main effects can be interpreted as, when syllable frequency is high, homophone density shows inhibitory effects. However, for low syllable frequency, homophone density exhibits facilitatory effects.

Table 3: *Summary of the model with interactions between homophone density and syllable frequency.*

Variables	Estimate (β)	t-value
(log)Homophone density	0.0081	1.73
(log)Syllable frequency	-0.0023	-1.49
(log)Stimulus duration	0.3484	16.09
Tone 2	-0.0040	-0.56
Tone 3	0.0281	3.54
Tone 4	0.0190	2.53
(log)Homophone density *(log)Syllable frequency	0.0047	2.24

In addition, to confirm the above-mentioned findings, data were split (using quantile-split) into high syllable frequency (top 10%) and low syllable frequency (bottom 10%). The effect of homophone density was tested separately for items with high syllable frequency and those with low syllable frequency. Again, mixed models were constructed with homophone density as the main effect and stimulus duration and lexical tone as control variables for high and low syllable frequency groups, separately. The model with high syllable frequency items revealed a significant inhibitory homophone density effect ($\beta = 0.0139$, $t = 2.17$, $\chi^2 = 4.2297$, $p = 0.03$). The effect was not significant for the low frequency items however, a reversal in the direction of coefficient was noted. Tables 4 and 5 summarize the results of high syllable frequency group and low syllable frequency group, respectively.

Table 4: *Summary of the model with high syllable frequency syllables.*

Variables	Estimate(β)	t-value
(log)Homophone density	0.0261	2.13
(log)Stimulus duration	0.3917	6.37
Tone 2	-0.0340	-1.49
Tone 3	0.0447	1.35
Tone 4	-0.0025	-0.12

Table 5: *Summary of the model with low syllable frequency syllables.*

Variables	Estimate (β)	t-value
(log) Homophone density	-0.0018	-1.24
(log)Stimulus duration	0.3489	4.84
Tone 2	-0.0209	-0.63
Tone 3	0.0187	0.89
Tone 4	0.0123	-0.54

Overall, monosyllables with high syllable frequency and high number of homophones were named slower as compared to those with low syllable frequency and high number of homophones.

4. Discussion

The current study investigated the effects of homophone density on spoken word recognition in Mandarin Chinese using an auditory naming task. A significant inhibitory effect of homophone density was found. Monosyllables with more homophonic mates were named slower compared to monosyllables with less homophonic mates. The current findings are in agreement with Wang et al. [4] who found an inhibitory effect of homophone density. Further, Li et al. [5] found significant interactions between word frequency and homophone family size. They found that high frequency words within a homophone family size facilitates the word recognition process while low frequency words within a homophone family size inhibits word recognition process. Also, they found that homophone family size has a facilitatory effect for low frequency words but not for high frequency words. This discrepancy with the current study could be due to the differences in task employed. Li et al. [5] used a cross-modal task where participants were asked to report whether the visual and auditory stimuli match or not. Since their task was bimodal, homophone density effects could have varied across different modalities. However, within the same modality, the results of the present study were consistent with the previous findings in the auditory lexical decision task [15] that looked at the interactions between homophone density and maximum homophone frequency within the homophone family. They found that the effect of homophone density reverses when the maximum homophone frequency changes from high to low. However, Wang et al. [4] couldn't find an interaction because their stimuli only contained high syllable frequency stimuli for testing the effects of homophone density. As seen in the current study, the effects are inhibitory for high syllable frequency that again conform with the findings of Wang et al. [4] with high syllable frequency.

Taken together, the results from the current study indicate that the effects of homophone density are not merely inhibitory or facilitatory in nature. In fact, these effects are more complex to understand as they also depend on the syllable frequency when processing monosyllabic Mandarin spoken words. The findings of the current study can be discussed as a trade off between syllable frequency and homophone density where syllable frequency and homophone density compensate for each other towards the process of spoken word recognition.

5. Conclusion

From the current findings, it can be concluded that homophone density affects spoken word recognition, not in a straightforward but in a complex manner. In other words, the effects of homophone density are modulated by syllable frequency such that there is an interaction between the two.

6. Acknowledgements

This research was supported by the associate money fund provided by the Hong Kong Polytechnic University. I would like to acknowledge Dr. Yao Yao for her suggestions and Ms. Liu Chang for helping with recording the stimuli.

7. References

[1] P. A. Luce and D. B. Pisoni, "Recognizing spoken words: the neighborhood activation model," *Ear Hear.*, vol. 19, no. 1, pp. 1–36, Feb. 1998.

[2] M. S. Vitevitch and P. A. Luce, "When Words Compete: Levels of Processing in Perception of Spoken Words," *Psychol. Sci.*, vol. 9, no. 4, pp. 325–329, Jul. 1998.

[3] M. S. Vitevitch and E. Rodríguez, "Neighborhood density effects in spoken word recognition in Spanish," *J. Multiling. Commun. Disord.*, vol. 3, no. 1, pp. 64–73, 2005.

[4] W. Wang, X. Li, N. Ning, and J. X. Zhang, "The nature of the homophone density effect: An ERP study with Chinese spoken monosyllable homophones," *Neurosci. Lett.*, vol. 516, no. 1, pp. 67–71, May 2012.

[5] X. J. Li, J. Fang, and J. Lou, "Interaction effect of homophone family size and specific-word frequency on the auditory lexical access of Chinese characters," *J. Psychol. Sci.*, vol. 34, pp. 43–47, 2011.

[6] W.-F. Chen, P.-C. Chao, Y.-N. Chang, C.-H. Hsu, and C.-Y. Lee, "Effects of orthographic consistency and homophone density on Chinese spoken word recognition," *Brain Lang.*, vol. 157, pp. 51–62, 2016.

[7] Y. Hino, Y. Kusunose, S. J. Lupker, and D. Jared, "The processing advantage and disadvantage for homophones in lexical decision tasks.," *J. Exp. Psychol. Learn. Mem. Cogn.*, vol. 39, no. 2, p. 529, 2013.

[8] W. Zhou, "The Homophone Effect in Mandarin Word Recognition," PhD Thesis, The Ohio State University, 2015.

[9] P. Li and M. C. Yip, "Context effects and the processing of spoken homophones," in *Cognitive Processing of the Chinese and the Japanese Languages*, Springer, 1998, pp. 69–89.

[10] H.-C. Chen, J. Vaid, and J.-T. Wu, "Homophone density and phonological frequency in Chinese word recognition," *Lang. Cogn. Process.*, vol. 24, no. 7–8, pp. 967–982, 2009.

[11] J. FANG, X. LI, and W. LUO, "Influence of Cumulative Frequencies of Chinese Syllables on the Activation of Homophone Representation in Auditory Lexical Access," *Acta Psychol. Sin.*, vol. 46, no. 4, pp. 467–480, 2014.

[12] M. C. Yip, "Access to homophonic meanings during spoken language comprehension: effects of context and neighborhood density.," in *INTERSPEECH*, 2002.

[13] Y. Yao and B. Sharma, "What is in the neighborhood of a tonal syllable? Evidence from auditory lexical decision in Mandarin Chinese," *Proc. Linguist. Soc. Am.*, vol. 2, pp. 45–1, 2017.

[14] X. J. Li, W. N. Wang, and X. Q. Li, "Auditory word frequency effect within homophone families and the activation of homophone representations," *Acta Psychol. Sin.*, vol. 43, pp. 749–762, 2011.

[15] B. Sharma and Y. Yao, "Effect of homophone density in spoken word recognition: Evidence from Mandarin," (to appear, June 2018) *J. Lab Phon*.

[16] S. Mathôt, D. Schreij, and J. Theeuwes, "OpenSesame: An open-source, graphical experiment builder for the social sciences," *Behav. Res. Methods*, vol. 44, no. 2, pp. 314–324, Nov. 2011.

[17] Q. Cai and M. Brysbaert, "SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles," *PLoS ONE*, vol. 5, no. 6, Jun. 2010.

[18] R. C. Team, *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013*. 2014.

[19] D. Bates, M. Maechler, B. Bolker, S. Walker, and others, "lme4: Linear mixed-effects models using Eigen and S4," *R Package Version*, vol. 1, no. 7, 2014.