Nguyen, Anh Kim *Email: anngu104@hhu.de, Faculty of Arts and Humanities, Department of General Linguistics, HHU Düsseldorf (ror.org/024z2rq82)*

Tang, Kevin *Email: kevin.tang@hhu.de, Faculty of Arts and Humanities, English and American Studies, HHU Düsseldorf; Department of Linguistics, University of Florida. (ror.org/02y3ad647)*

Title: The effect of syllable segmentation and phonological neighbours on spoken word recognition in Mandarin Chinese. **Motivation:** In Indo-European languages, the phonological access points, also known as 'proximate units,' are segmental, whereas in Mandarin Chinese, they are entire syllables [1]. There is no uniform approach to how syllables are phonologically segmented for Mandarin Chinese [2]. Words may be deemed phonological neighbours, depending on what part of the syllable is regarded contrastive. Whether the entire rime instead of every phoneme is considered as a contrastive element, and whether or not tone is included can have consequences for the way phonological neighbourhood density (PND), neighbourhood frequency (PNF), and homophone density (HD) are calculated. Neergaard et al. [3] created a database of Mandarin Chinese words, encoded more than 16 different syllable segmentation patterns, using different combinations of phonemes, syllable structure, and tonemes. Sharma [4] examined four such segmentation schemas: onset and rime with tone (C_GVX_T), and without tone (C_GVX); phonemes with tone (C_G_V_X_T) and without tone (C_G_V_X). They compared the effects of phonological neighborhood based on each of these schemas on the processing of Mandarin Chinese syllables in an auditory lexical decision task (Table 1). While they find certain neighborhood measures to have stronger predictive power when using an Onset-Rime-Tone segmentation pattern, they also point out that their findings might depend on the nature of their task. Our study continues this line of research and aims to assess further insight on phonological representation and neighbourhood measures, examining the mentioned four segmentation schemas but on a task that requires a different type of spoken word recognition. **Data:** An open-set speech-in-noise spoken word recognition task was used to elicit identification accuracy of Mandarin Chinese syllables. 30 normal-hearing participants were tested and are native speakers of Mandarin from mainland China. An exhaustive syllable set with 1,310 real syllables was tested and 1,164 of which were enriched with lexical statistics (PND, PNF, HD and syllable frequency) based on the four segmentation schemas from the database by Neergaard et al. [5]. The syllables were recorded by native speakers of both genders, intensity normalised, high-pass filtered and silence padded. Each syllable was randomly assigned one of the three Signal-to-Noise-Ratio (SNR) (-5, 0 and +5 dB) with speech-shaped noise. Each participant was tested on a random sample of 655 target syllables. Participants used a keyboard to enter their responses. 16,490 recognition trials over 1,164 real syllables were examined. **Methods:** Logistic mixed effect regressions (*glmer*) were used to predict accuracy with four lexical measures (PND, PNF, HD, Syllable Frequency) and other control variables (trial position, SNR, volume, syllable duration, speaker's gender). Four models were fitted with four different sets of lexical measures (one per schema). Participants and items are random intercepts. Continuous variables were log10-transformed (if needed), then z-transformed. The model with the lowest AIC/BIC was determined to be the best model. **Results and Discussion:** PND has an inhibitory effect (negative coefficient); HD has a facilitatory effect (positive coefficient); syllable frequency has a facilitatory effect only under the C_G_V_X schema. Two schemas, C_GVX_T and C_G_V_X, resulted in similarly low AIC and BIC scores (Δ AIC/BIC < 2 are not significant) (Table 2), suggesting that the tone needs to be considered only when the syllable is segmented into onsets and rimes. **Conclusion:** The findings of our study adds to the findings on neighbourhood effects in relation to word representation and processing in Mandarin Chinese. Our use of speech-in-noise word recognition requires a different, arguably more naturalistic, type of speech processing than what is required by auditory lexical decision. Compared to Sharma [4], we also established an inhibitory effect of neighbourhood density on the spoken word recognition and C_GVX_T was also found to be one of the best schemas, albeit with a different task. The surprising finding with C_G_V_X might be due to the role of task types and requires a more indepth investigation across spoken word recognition tasks.

Study	Task type	# and type of Items	PND	PNF	HD	Freq	Best fit schema
Sharma (2020)	Auditory Lexical Decision (Reaction Time)	598 real words 3-phoneme syllables	n.s	-	-	N/A	C_GVX_T
	Auditory Lexical Decision (Accuracy)		n.s	n.s	+	+	No best
This study	Open-set spoken word recognition (Accuracy)	1,164 real words	-	n.s.	+	n.s. or +	C_GVX_T or C_G_V_X

Table1: Comparison of previous findings on predictive power of metrics. Greyed out cells indicate reported non-significant effects (n.s.). + signs indicate that the increased presence of a respective metric correlated with correct answers. – signs indicate increased absence correlated with correct answers. N/A indicates effects that were not tested. PND: phonological neighbourhood density – number of phonological neighbours. PNF: phonological neighbourhood frequency – average frequency of the neighbours. HD: homophone density – number of homophones. Freq: syllable frequency.

Segmentation*	PND	PNF	HD	Freq	SNR	Dur	Gender**	AIC	BIC
C_GVX_T	-	n.s.	+	n.s.	+	+	+	**17123**	**17216**
C_GVX	n.s.	n.s.	(+)	+	+	+	+	17128	17221
C_G_V_X_T	-	n.s.	+	n.s.	+	+	+	17126	17218
C_G_V_X	-	n.s.	+	+	+	+	+	**17122**	**17214**

Table2: Summary of significant predictors on spoken word recognition accuracy
* based on Neergaard & Huang's schemas for "Segments" (Onset + Rime) and "Component" (Phonemes) [2]; ** Items read by a male voice were understood significantly better

[1] O'Seaghdha, P. G., Chen, J.-Y., & Chen, T.-M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition*, *115*(2), 282–302. https://doi.org/10.1016/j.cognition.2010.01.001

[2] Neergaard, K. D., & Huang, C.-R. (2019). Constructing the Mandarin Phonological Network: Novel Syllable Inventory Used to Identify Schematic Segmentation. *Complexity*, *2019*, 1–21. https://doi.org/10.1155/2019/6979830

[3] Neergaard, K., Xu, H., & Huang, C.-R. (2016). Database of Mandarin Neighborhood Statistics. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 4032–4036). European Language Resources Association (ELRA). https://aclanthology.org/L16-1636

[4] Sharma, B. (2020). *The effect of phonological neighbors and homophones on spoken word recognition in Mandarin Chinese.* Doctoral Dissertation. Hong Kong Polytechnic University. https://theses.lib.polyu.edu.hk/handle/200/10458

[5] Neergaard, K. D., Xu, H., German, J. S., & Huang, C.-R. (2021). Database of word-level statistics for Mandarin Chinese (DoWLS-MAN). *Behavior Research Methods*, *54*(2), 987–1009. https://doi.org/10.3758/s13428-021-01620-7