

## What is in the neighborhood of a tonal syllable? Evidence from auditory lexical decision in Mandarin Chinese

Yao Yao & Bhamini Sharma\*

**Abstract.** Phonological neighborhood effects have been found in spoken word recognition, word production and phonetic variation (Gahl, Yao, & Johnson, 2012; Luce & Pisoni, 1998; Vitevitch, 2002). Overall, words from dense neighborhoods are harder to recognize but easier to produce. However, most previous studies have focused on English, while evidence suggests that these effects may not generalize cross-linguistically due to language-specific configurations of the lexicon (Michael S Vitevitch & Stamer, 2006, 2009). In the current study, we investigate the effects of phonological neighborhoods in Mandarin Chinese, which has a vastly different lexicon structure from that of English. Results from an auditory lexical decision experiment showed that phonological neighborhood density and neighbor frequency (defined by the one-phoneme/tone difference rule) are predictive of the speed and accuracy of lexical decision. Homophone density also has a facilitative effect on the accuracy of lexical decision. The implications of the current findings are discussed in the framework of the lexicon model proposed by Zhou & Marslen-Wilson (1994, 2009).

**Keywords.** Phonological neighborhood effect; homophone density effect; Mandarin Chinese; spoken word recognition

**1. Background.** Words in a mental lexicon are connected via various relations, and one such relation is phonological similarity. Words that sound similar, i.e. phonological neighbors, may influence each other in lexical processing by way of co-activation. The most widely used working definition of phonological neighbors is the one-phoneme difference rule (Luce & Pisoni, 1998), according to which any two words that differ in one and only one phoneme—by addition, deletion or substitution—are considered phonological neighbors. The effects of phonological neighbors vary with the nature of the processing task. Specifically, it has been shown that the overall effect of phonological neighbors is inhibitory in spoken word recognition but facilitative in word production.

The widely-cited work by Luce & Pisoni (1998) consisted of a series of perceptual experiments (perceptual identification, auditory lexical decision, word naming), using an exhaustive set of 918 three-phoneme, monosyllabic English words. The experimental results showed that neighborhood density (i.e. the number of phonological neighbors a word has) strongly predicted recognition speed and accuracy. High-density words (i.e. words with many neighbors) had longer reaction times and lower accuracy rates in the perceptual tasks, compared with low-density words (i.e. words with fewer neighbors). These results indicated that phonological neighbors acted as competitors of the target word during the recognition process and as a result, the more neighbors (competitors) a target word has in the lexicon, the greater the

---

\* Research reported in this paper was supported by funding from the Dean's Reserve for Research, Scholarly, and Other Endeavors (Project Account Code: 1-ZVB7), Faculty of Humanities, The Hong Kong Polytechnic University. Authors: Yao Yao, The Hong Kong Polytechnic University (ctyaoyao@polyu.edu.hk) & Bhamini Sharma, The Hong Kong Polytechnic University (bhamini.sharma@polyu.connect.hk).

inhibitory effect on the recognition of the target word. Similar effects were found for the usage frequency of phonological neighbors (“neighbor frequency” hereafter). When neighborhood density was controlled for, words with higher-frequency neighbors also had slower and less accurate responses compared to words with lower-frequency neighbors, suggesting that high-frequency neighbors induced stronger competition than low-frequency neighbors. The neighbor frequency effects were more pronounced among low-density target words. Similar findings have been reported in a few other studies (Cluff & Luce, 1990; Goldinger, Luce, & Pisoni, 1989; Luce, Goldinger, Auer, & Vitevitch, 2000; Vitevitch & Luce, 1999).

By contrast, phonological neighbors play an overall facilitative role in word production tasks. Vitevitch (2002) investigated the effect of phonological neighborhoods in speech production using two speech error elicitation techniques and a series of picture naming tasks. The results showed that compared with words from sparse neighborhoods, words from dense neighborhoods elicited fewer errors and were named faster, providing evidence that phonological neighbors facilitate the production process—probably via increasing the activation of the target word. Similar findings were reported in subsequent studies (Peramunage, Blumstein, Myers, Goldrick, & Baese-berk, 2011; Vitevitch, Ambrüster, & Chu, 2004; Vitevitch & Sommers, 2003).

Most of the research reviewed above has focused on the processing of short, monosyllabic words in English. When the scope of investigation is extended to other languages, the patterns reported above may not generalize well cross-linguistically. Just like in English, high-density words in French are hard to perceive (Dufour & Frauenfelder, 2010; Ziegler, Muneaux, & Grainger, 2003), but they also seem to be hard to produce (Sadat, Martin, Costa, & Alario, 2014), and tended to be hyperarticulated in conversational speech (Yao & Meunier, 2014). The research on Spanish presented an even greater contrast with the findings from English: high-density words in Spanish are easy to perceive but hard to produce (Sadat et al., 2014; Vitevitch & Rodríguez, 2005; Vitevitch & Stamer, 2006, 2009)—both of which are in the opposite directions of their counterpart effects in English. There have also been studies about neighborhood effects in Japanese, but the findings are mixed (Amano & Kondo, 2000; Yoneyama, 2002).

Some cross-language discrepancies have been attributed to language-specific properties of neighborhood structure, word structure, and morphology. For instance, Vitevitch & Stamer (2006) points out that Spanish words have more word-final inflections—indicating word class, gender and tense—compared to English words. Therefore, phonological neighbors in Spanish tend to be morphologically related and sharing the onset, which makes them strong competitors in a production task (O’Seaghdha & Marin, 2000) and thus inhibit the production process.

In the current study, we focus on Mandarin Chinese, a language that is typologically distant from English. An immediate challenge for the research on Chinese phonological neighborhoods is the definition of phonological neighbors. The widely cited one-phoneme difference rule obviously has no consideration for tones, although 70% of the world’s languages are tonal languages (Yip, 2002). Malins & Joanisse (2010, 2012) studied the influence of tonal and segmental information in spoken word recognition using eye-tracking and event-related potentials. They concluded that tones and segments were processed simultaneously in Chinese word recognition, although previous studies showed that segments were processed earlier than lexical tones (H.-C. Chen & Cutler, 1997; Ye & Connine, 1999). Despite the issue of relative timing, it is widely agreed that tones do play an important role in spoken word recognition of Mandarin Chinese. A major goal of the current study is to explore the structure of phonological

neighborhood in Mandarin Chinese—in particular, the role of tone in the notion of phonological neighbor.

The second challenge is the issue with homophones. According to the one-phoneme difference rule, homophones—words with identical pronunciations (e.g. *time* and *thyme*)—are not phonological neighbors, although we may feel that homophones are phonologically more similar than one-phoneme different neighbors (see Gahl 2008 for a discussion on the phonetic realizations of homophones). The issue of homophones is negligible for the English phonological neighborhood model, due to the small number of homophones in the lexicon. But the Chinese lexicon has a vast number of homophones, as the building blocks of the Chinese lexicon are monosyllabic tonal morphemes that can be combined to make longer compound words. The set of legal monosyllables is relatively small (~1200 attested tonal syllables), leading to a high density of homophonic morphemes. According to Su & Lin (2006), a monosyllable (with tone) in Mandarin Chinese is associated with on average 8.31 morphemes (characters). The second goal of the current study is to investigate the effects of homophones in spoken word recognition, and to compare the effects of homophones with those of phonological neighbors.

A few previous studies have looked at neighborhood and homophone effects in the processing of monosyllabic items in Mandarin Chinese. Tsai (2007) found an inhibitory effect of neighborhood density in an auditory naming task, in that high-density words were responded to slower than low-density words. However, it should be noted that Tsai (2007) defined phonological neighbors using the one-phoneme difference rule, which does not include tone. More recently, Neergaard & Huang (2016) attempted to replicate Tsai's auditory word naming study and to test a number of different ways of defining Mandarin phonological neighborhoods for their effects on spoken word recognition. Neergaard and Huang claimed that the best-fitting neighborhood density measures were the ones that incorporated both tone and segments in the definition, but the validity of the claim is seriously challenged by a number of issues regarding experimental design and data analysis in their study (for example, it is not clear what criteria were used to select the set of stimuli, which contained 210 monosyllabic Mandarin words).

As for the effects of homophone, Wang, Li, Ning, & Zhang (2012) found an inhibitory effect of homophone density (i.e. the number of homophones of a target word) in an auditory lexical decision task using monosyllabic Mandarin words: words with more homophones elicited longer reaction times. Similar results are reported in Zeng & Mattys (2011).

In this paper, we report an auditory lexical decision experiment in Mandarin Chinese using an exhaustive set of more than 1200 monosyllabic morphemes. Similar to Neergaard & Huang (2016), we tested multiple neighborhood density measures for their predictive power for the speed and accuracy of lexical processing. To preview the results, we found an inhibitory effect of neighborhood density and a facilitative effect of neighbor frequency on processing speed, as well as a facilitative effect of homophone density on processing accuracy. The best-fitting neighborhood metrics were defined by the one-phoneme/tone difference rule, which recognizes any two syllables that only differ in one phoneme or tone as phonological neighbors. The findings are discussed in the lexicon model proposed by Zhou & Marslen-Wilson, 1994, 2009).

## **2. Method.**

2.1. PARTICIPANTS. Seventy-eight right-handed native Mandarin speakers (49F, 29M; mean age = 23.4 years, SD = 4.26) born and raised in Mainland China participated in the experiment. None of the participants reported any speech or hearing problem.

2.2. STIMULI. The stimuli consisted of 1258 real monosyllables (critical) and 761 pseudo-syllables (filler). The set of real syllables was an exhaustive set of all possible monosyllables in Mandarin Chinese, i.e. all the syllables that can be associated with at least one morpheme (Chinese character). The set of filler pseudo-syllables consisted of two types of unattested but phonotactically possible monosyllables: syllables whose segmental composition is attested in a different tone (“tonal gaps”, N = 353 items, e.g. [pan 2] *ban2*) and syllables whose segmental composition is not attested in any tone (“segmental gaps”, N = 408 items, e.g. [lun1] *lun1*).

All the stimuli were recorded by a female native Mandarin speaker in an acoustically treated room using a uni-directional microphone, routed to Digi design. The stimuli had a mean duration of 626 ms (SD = 100), and were normalized for intensity at 70 dB using Praat (Boersma & Weenink, 2010). To keep the experimental sessions under reasonable durations, the set of real syllable stimuli were evenly divided into six blocks; each block also contained an equal number of randomly drawn pseudo-syllables (half tonal gaps and half segmental gaps). Each participant only worked on one block, and each block was presented to a total of 13 participants. In order to evaluate cross-block consistency, 12 items were shared among all the blocks.

2.3. PROCEDURE. The auditory lexical decision experiment was carried out with E-Prime version 2.0 (Schneider, Eschman, & Zuccolotto, 2007) on a Lenovo laptop that was connected to a set of headphones and a Chronos response box. During each trial, initially a fixation-cross appeared at the center of the computer screen for 500 ms, and then an auditory stimulus was presented. Participants were instructed to make a judgment as fast and accurately as possible on whether the spoken stimulus could be associated with a legal Chinese character or not, and to press the appropriate key in the Chronos response box to indicate their judgment. If no response was recorded within 4000 ms after the onset of the stimulus, the experimental session would proceed to the next trial.

All experimental sessions were conducted in a sound-attenuated room. Each session began with 30 practice items (with feedback), followed by a block of test items presented in a randomized order (without feedback). Response times (RT) were recorded as the duration between the onset of the stimulus and the participant’s button-press response. Most experimental sessions lasted less than 30 minutes.

2.4. ANALYSIS. Lexical measures (syllable frequency, neighborhood density, neighbor frequency, homophone density) of the stimuli were obtained from the SUBTLEX-CH corpus (Cai & Brysbaert, 2010) and the neighborhood metrics database based on the SUBTLEX-CH corpus (Neergaard, Xu, & Huang, 2016).<sup>1</sup> Neergaard et al. calculated the neighborhood metrics for a number of different schemas. All the schemas defined phonological neighbors with the one-difference rule, but differed in how the syllables were segmented and whether tone was considered when counting the number of differences. In the current study, we chose four sets of neighborhood metrics from Neergaard et al.’s database, which were contrasted by whether syllables were segmented into phonemes or larger components (onset+rime) and whether tone was included when counting the differences. Each set consisted of a neighborhood density measure (i.e. number of neighbors) and a neighbor frequency measure (i.e. sum of all neighbors’ usage frequencies). Table 1 lists the names of the neighborhood measures. Table 2 summarizes the distribution of all the lexical measures in the current set of syllable stimuli.

---

<sup>1</sup> The neighborhood metrics used in the current study have been slightly adjusted based on those in Neergaard et al.’s database. For example, all frequency-related measures were converted to the scale of number of occurrences per million words.

<b>Neighborhood definition</b>	<b>Neighborhood density</b>	<b>Neighbor frequency</b>
Syllable segmented into phonemes. One-segment/tone difference rule.	ND_SegT	NF_SegT
Syllable segmented into phonemes. One-segment difference rule.	ND_Seg	NF_Seg
Syllable segmented into components (onset + rime). One-component/tone difference rule.	ND_CompT	NF_CompT
Syllable segmented into components (onset + rime). One-component difference rule.	ND_Comp	NF_Comp

Table 1. Neighborhood density (ND\_) and neighbor frequency (NF\_) measures under four different schemas.

	<b>Range (min, max)</b>	<b>Mean</b>	<b>Median</b>	<b>SD</b>	
<b>Neighborhood measures</b>	<b>ND_SegT</b>	[3, 32]	14.89	14.00	5.09
	<b>NF_SegT</b>	[77.39, 64152.06]	8131.92	3974.42	10795.31
	<b>ND_Seg</b>	[8, 49]	23.01	23	5.79
	<b>NF_Seg</b>	[1832.11, 122056.61]	30204.29	22403.19	24789.54
	<b>ND_CompT</b>	[7, 48]	24.15	24	6.35
	<b>NF_CompT</b>	[294.73, 115602.20]	14147.23	8318.15	15380.39
	<b>ND_Comp</b>	[23, 84]	44.79	43	10.60
	<b>NF_Comp</b>	[5719.10, 174509.12]	54941.09	47808.51	34473.06
<b>Homophone density</b>	[1, 42]	4.26	3	3.98	
<b>Syllable frequency</b>	[0.03, 50147.83]	610.21	46.71	3297.58	
<b>Syllable length (# of phonemes)</b>	[1, 4]	2.93	3	0.71	

Table 2. Summary of lexical measures for all the real syllable items. All frequency-related measures (syllable frequency, and all the NF\_ measures) are number of occurrences per million words.

Figure 1 shows the mean neighborhood density—under different schemas—for each syllable length in real syllables. Similar to the findings for English, shorter syllables tend to have more neighbors than longer words, across all four schemas. To control for the correlation between syllable length and neighborhood density, in the current study we focus the analysis on 3-phoneme consonant-initial (CXX) syllables, which comprise about half of the items (625 real syllables, 378 pseudo-syllables).

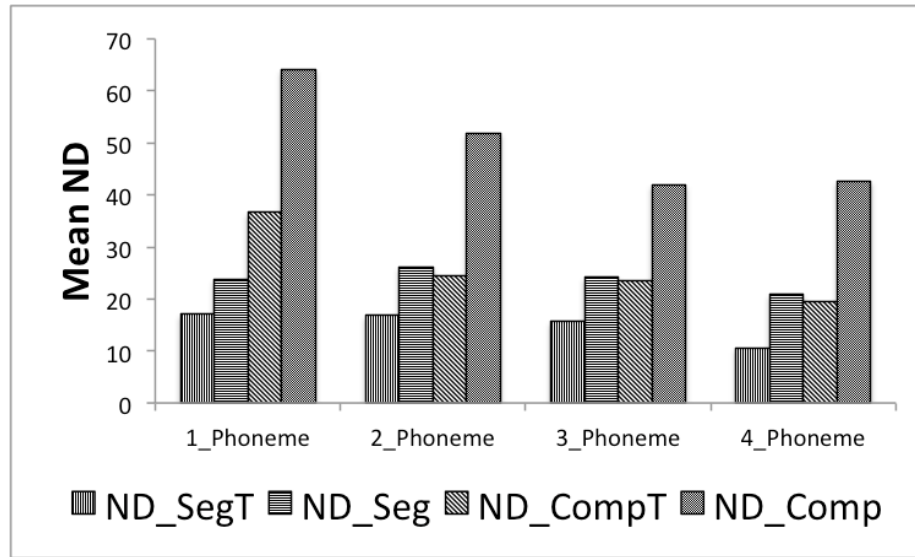


Figure 1. Mean neighborhood density by different neighborhood definitions and by syllable length.

Two types of mixed-effects models were built: linear regression for modeling (log) response times and logistic regression for modeling accuracy. In the models of critical real-syllable trials, fixed effects included neighborhood density, (log) neighbor frequency, (log) homophone density (HomoD), (log) syllable frequency (SyllFreq), (log) syllable duration (SyllDur), and lexical tone (Tone); the models of filler pseudo-syllable trials excluded homophone density and syllable frequency—both of which were not available for pseudo-syllables—and added pseudo item type (ItemType) as a fixed-effect predictor. Each model was built in four versions, with each version using only one of the four sets of neighborhood measures. Effectiveness of neighborhood measures was evaluated by comparing across different versions of the same model. All the models had random intercepts for participant and item. Numerical variables were log-transformed (if necessary for a more normal distribution) and centered before being entered into the models. All the analysis was completed with the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R (R Team, 2014).

**3. Results.** As shown in Table 3, the overall performance in the lexical decision task was slower (mean = 1054 ms) and less accurate (mean = 84.8%) than what we usually expect from a lexical decision task in other languages (e.g. English). This is probably due to two facts: (1) the high number of phonological neighbors and homophones in Chinese may heighten the level of competition in lexical processing, (2) the Chinese participants may be less familiar with processing monosyllabic morphemes as the majority of Chinese words are di- or tri-syllabic compounds.

		Mean RT (ms)	Mean Accuracy (%)
<b>Real syllables</b>		1025 (SD=344)	83.9
<b>Pseudo-syllables</b>	Tonal gaps	1135 (SD=466)	84.5
	Segmental gaps	995 (SD=377)	91.5
<b>Overall</b>		1054 (SD=412)	84.8

Table 3. Summary of response times (RT) and accuracy for real and pseudo CXX syllables.

We removed two types of outliers from data analysis: (1) items with extremely low accuracy (<20%), and (2) trials with extremely short or long RTs (>2 SD away from the mean RT). Altogether about 5% of the trials were excluded. Furthermore, RT analysis was conducted only with data from correctly-answered trials, whereas accuracy analysis included all valid trials.

3.1. RT IN REAL-SYLLABLE TRIALS. The models on RT in real-syllable trials included 12868 trials of 554 syllable types. When neighborhood measures based on the one-segment/tone difference rule (ND\_SegT, NF\_SegT) were used, the model showed a positive effect of neighborhood density ( $\beta = 0.0023$ ,  $t = 2.5$ , 95% CI=[0.0005, 0.004]; see Table 4 for model summary) and a negative effect of neighbor frequency ( $\beta = -0.012$ ,  $t = -3.0$ , 95% CI = [-0.02, -0.004]). In other words, everything else being equal, real syllables with more neighbors are slower to respond to, but real syllables with more frequent neighbors are faster to respond to. There is no effect of homophone density ( $\beta = 0.0013$ ,  $t = 0.3$ , 95% CI = [-0.009, 0.01]) in the model. In alternative neighborhood measures were used, neighbor frequency had a similar, significant negative effect ( $|t| > 2.5$ ) in two of the three alternative models, but neighborhood density was not significant in any of the alternative models ( $|t| < 2$ ; see Table 5). Homophone density was not significant in any models (all  $|t| < 1$ ). Other control effects are as predicted: syllables with high usage frequency are responded to faster; syllables with longer stimuli were responded to slower; T3 syllables were responded to faster than average, probably due to an earlier recognition point.

<b>Fixed effects</b>	<b><math>\beta</math></b>	<b><math>t</math></b>	<b>95% CI</b>
Intercept	6.84	654.6	[6.8, 6.9]
ND_SegT	0.0023	2.5	[0.0005, 0.004]
NF_SegT	-0.012	-3.0	[-0.02, -0.004]
HomoD	0.0014	0.3	[-0.009, 0.01]
SyllFrq	-0.013	-8.0	[-0.02, -0.01]
SyllDur	0.54	18.1	[0.5, 0.6]
Tone=T1	-0.011	-1.8	[-0.02, 0.001]
Tone=T2	-0.003	-0.4	[-0.02, 0.01]
Tone=T3	-0.034	-4.5	[-0.05, -0.02]

Table 4. Summary of fixed effects in the RT model for real-syllable trials, with ND\_SegT and NF\_SegT as neighborhood metrics.

<b>Models with alternative neighborhood metrics</b>	<b>Critical fixed effects</b>	<b><math>\beta</math></b>	<b><math>t</math></b>	<b>95% CI</b>
Using ND_Seg and NF_Seg	ND_Seg	0.0015	1.9	[-0.00007, 0.003]
	NF_Seg	-0.013	-2.5	[-0.02, -0.003]
	HomoD	0.00037	0.1	[-0.009, 0.01]
Using ND_CompT and NF_CompT	ND_CompT	-0.00014	-0.2	[-0.002, 0.001]
	NF_CompT	-0.013	-2.9	[-0.02, -0.004]
	HomoD	0.0034	0.7	[-0.006, 0.01]
Using ND_Comp and NF_Comp	ND_Comp	-0.00053	-1.0	[-0.001, 0.0005]
	NF_Comp	-0.012	-1.9	[-0.02, 0.0002]
	HomoD	0.0033	0.6	[-0.006, 0.01]

Table 5. Summary of neighborhood effects in the RT models for real-syllable trials, with alternative neighborhood metrics.

3.2. ACCURACY IN REAL-SYLLABLE TRIALS. Overall, models on the accuracy of real-syllable trials (14495 trials, 554 items) revealed no significant effects of either neighborhood density or neighbor frequency, regardless of which neighborhood metrics were used (all  $p > .01$ ). Nevertheless, a significant positive effect of homophone density was observed in three of the four models ( $p < .005$ ), on top of a significant effect of syllable frequency, which was observed in all four models (all  $p < .001$ ). That is to say, the more homophones a syllable can be associated with—and the more frequent these homophones are—the more likely the syllable will be correctly classified as “real”. As an example, the summary of fixed-effects terms in the model with ND\_SegT and NF\_SegT is shown in Table 6. Critical fixed effects in alternative models are summarized in Table 7.

<b>Fixed effects</b>	$\beta$	$t$	$p$
Intercept	2.74	25.7	<.001
ND_SegT	-0.00028	-0.02	1
NF_SegT	-0.047	-0.7	0.5
HomoD	0.25	3.0	0.003
SyllFrq	0.25	9.23	<.001
SyllDur	-0.28	-0.57	0.6
Tone=T1	-0.05	-0.52	0.6
Tone=T2	-0.04	-0.34	0.7
Tone=T3	0.05	0.4	0.7

Table 6. Summary of fixed effects in the accuracy model for real-syllable trials, with ND\_SegT and NF\_SegT as neighborhood metrics.

<b>Models with alternative neighborhood metrics</b>	<b>Critical fixed effects</b>	$\beta$	$z$	$p$
Using ND_Seg and NF_Seg	ND_Seg	0.0095	0.72	0.4
	NF_Seg	-0.025	-0.3	0.7
	HomoD	0.24	2.85	<0.01
Using ND_CompT and NF_CompT	ND_CompT	0.015	1.21	0.2
	NF_CompT	0.0043	0.06	1.0
	HomoD	0.24	2.85	<0.01
Using ND_Comp and NF_Comp	ND_Comp	0.02	2.444	0.01
	NF_Comp	-0.026	-0.26	0.8
	HomoD	0.21	2.41	0.01

Table 7. Summary of neighborhood effects in the accuracy models for real-syllable trials, with alternative neighborhood metrics.

Taken together, for the processing of real syllables in the lexical decision task, neighborhood measures mainly affect the speed of classification, while homophone density affects the accuracy of classification. Among the four sets of neighborhood metrics, ND\_SegT and NF\_SegT are the only set that show an inhibitory effect of neighborhood density; none of the other three sets showed any sensitivity between the neighborhood measure and response time.

3.3. RT AND ACCURACY IN PSEUDO-SYLLABLE TRIALS. Three out of the four models on RT in pseudo-syllable trials (11269 trials, 314 items) revealed a strong positive effect of neighborhood



density on response time (all  $|t| > 3$ ). In other words, pseudo-syllables with many real-syllable neighbors took more time to respond to, as more effort was needed to suppress the co-activated real syllables that interfered with the correct, “non-word” response. Neighbor frequency was not a significant predictor in any model (all  $|t| < 1.5$ ). In addition, there is also a strong effect of item type (tonal gaps vs. segmental gaps), in that tonal gaps elicited significant longer response times (all  $|t| > 4$ ) than segmental gaps, which corroborated the interference from real-syllable neighbors that only differ in tone. Table 8 shows the summary of a sample model with ND\_SegT and NF\_SegT; Table 9 shows the summary of critical effects in models with alternative neighborhood measures.

<b>Fixed effects</b>	<b><math>\beta</math></b>	<b><math>t</math></b>	<b>95% CI</b>
Intercept	6.86	503.2	[6.8, 6.9]
ND_SegT	0.0093	6.8	[0.007, 0.01]
NF_SegT	-0.0022	-0.6	[-0.009, 0.005]
SyllDur	0.33	9.3	[0.3, 0.06]
Tone=T1	-0.0088	-1.2	[-0.02, 0.006]
Tone=T2	-0.016	-2.3	[-0.03, -0.002]
Tone=T3	-0.0072	-0.7	[-0.03, 0.01]
ItemType=Tonal gaps	0.04	4.1	[0.02, 0.06]

Table 8. Summary of fixed effects in the RT model for real-syllable trials, with ND\_SegT and NF\_SegT as neighborhood metrics.

<b>Models with alternative neighborhood metrics</b>	<b>Critical fixed effects</b>	<b><math>\beta</math></b>	<b><math>t</math></b>	<b>95% CI</b>
Using ND_Seg and NF_Seg	ND_SegT	0.0048	4.7	[0.003, 0.007]
	NF_SegT	0.0064	1.3	[-0.003, 0.02]
Using ND_CompT and NF_CompT	ND_CompT	0.0058	5.5	[0.004, 0.008]
	NF_CompT	-0.0043	-0.9	[-0.01, 0.005]
Using ND_Comp and NF_Comp	ND_Comp	0.0014	1.8	[-0.0001, 0.003]
	NF_Comp	-0.0052	-0.7	[-0.02, 0.008]

Table 9. Summary of neighborhood effects in the accuracy models for pseudo-syllable trials, with alternative neighborhood metrics.

The inhibitory effect of neighborhood density for pseudo-syllable trials was also found in the accuracy models (12840 trials, 314 syllable types). In all four models, neighborhood density had a significant negative effect on response accuracy (all  $p < .001$ ). No neighbor frequency effect was found in any model (all  $p > .01$ ). Consistent with the findings from the RT models, tonal gaps elicited more errors (i.e. lower accuracy) than segmental gaps. Table 10 shows the summary of the model with ND\_SegT and NF\_SegT; Table 11 lists the critical effects in all the alternative models.

In sum, our results showed that neighborhood density played an overall inhibitory role in auditory lexical decision. Real monosyllables from dense neighborhoods had longer response times than those from sparse neighborhoods; pseudo monosyllables with more real-syllable neighbors had both longer response times and lower accuracy than those with fewer neighbors. Neighbor frequency, on the other hand, had a curious facilitative effect in terms of response

times for real monosyllables, but was not significant in any other models. Among the four sets of neighborhood metrics tested in the current study, ND\_SegT and NF\_SegT—defined based on the one-phoneme/tone difference rule—turned out to be the most sensitive to lexical decision performance. Homophone density was tested only in the models for real syllables. Unlike previous studies (e.g., Wang et al., 2012), we found a facilitative effect of homophone density in terms of response accuracy. Real monosyllables associated with more homophonic morphemes (characters) are recognized more accurately than those associated with fewer morphemes (characters).

<b>Fixed effects</b>	$\beta$	$t$	$p$
Intercept	2.97	18.0	<.001
ND_SegT	-0.13	-4.4	<.001
NF_SegT	0.087	1.1	0.27
SyllDur	2.35	3.2	0.001
Tone=T1	-0.10	-0.7	0.5
Tone=T2	0.37	2.6	0.01
Tone=T3	-0.34	-1.7	0.09
ItemType=Tonal gaps	-0.33	-1.6	0.1

Table 10. Summary of fixed effects in the accuracy model for pseudo-syllable trials, with ND\_SegT and NF\_SegT as neighborhood metrics.

<b>Models with alternative neighborhood metrics</b>	<b>Critical fixed effects</b>	$\beta$	$z$	$p$
Using ND_Seg and NF_Seg	ND_Seg	0.089	-4.6	<.001
	NF_Seg	0.22	2.3	0.02
Using ND_CompT and NF_CompT	ND_CompT	-0.10	-5.0	<.001
	NF_CompT	0.20	2.2	0.02
Using ND_Comp and NF_Comp	ND_Comp	-0.051	-3.7	<.001
	NF_Comp	0.29	2.3	0.02

Table 11. Summary of neighborhood effects in the accuracy models for pseudo-syllable trials, with alternative neighborhood metrics.

**4. Discussion.** The current study aimed to investigate the effects of phonological neighborhood measures (neighborhood density and neighbor frequency) and homophone density on spoken word recognition in Mandarin Chinese. Our results showed that all three variables were, to some extent, predictive of Mandarin-speaking subjects’ performance in auditory lexical decision. Specifically, we found both inhibitory and facilitative effects from neighborhood measures: while high neighborhood density lengthens response times (for both real and pseudo syllables), high neighbor frequency predicts for faster responses (for real syllables only). The co-existence of inhibitory and facilitative neighborhood effects on response time can be explained by the nature of a lexical decision task. During each trial, the subject only needs to make a binary decision about whether the syllable is “real” or not, without having to identify which syllable they have heard. Thus, when the target response is “real”, the subject may arrive at a correct, “real” response via two routes: (a) when the target syllable is highly activated and correctly identified, or (b) when a non-target albeit “real” syllable is highly activated. Phonological

neighbors may impede route (a) by introducing more competition during the recognition process and thus slow down the response, meanwhile, they may also facilitate route (b) and thus speed up the response by quickly identifying a non-target syllable. This also explains why the facilitative effect of neighborhoods on response time was only found for real syllables but not for pseudo-syllables: when the target response is “not real”, route (b) will result in a wrong response, which will in turn be excluded from the response time analysis.

When compared with previous reports of Mandarin auditory lexical decision such as Wang et al. (2012), the current results presented two seeming discrepancies. First, the current subjects’ performance is much slower (mean RT = 1054 ms) and less accurate (mean accuracy = 84.8%) than those reported in Wang et al.’s study (mean RT around 800 ms; mean accuracy over 94%). Second, Wang et al. found an inhibitory effect of homophone density on both response time and accuracy, whereas the current study found a null effect on RT and a facilitative effect on accuracy. The first discrepancy can be reconciled by the fact that Wang et al.’s spoken stimuli are significantly shorter (mean duration = 430 ms) than those in the current study (mean duration = 626 ms), which would result in an earlier recognition point and thus shorter response times. Furthermore, Wang et al. only used high-frequency syllables, which we know would elicit faster and more accurate responses. As for the second discrepancy, a few potential sources can be speculated although we cannot offer a definitive answer for now. For example, it is possible that the inclusion of low-frequency syllables makes the current task much more difficult than that in Wang et al.’s study, which encouraged current subjects to adopt a task-specific strategy that trades response speed off for accuracy. As a result of accuracy being prioritized over speed, syllables associated with more homophonic morphemes (characters) are more likely to be correctly recognized than those with fewer homophones. After all, it is somewhat counter-intuitive to conceive an inhibitory effect of homophone density on recognition accuracy (i.e. more homophones lead to higher rates of incorrect “non-word” responses), unless we assume that the subject has set up a fast pace through the experiment—as they probably have in Wang et al.’s task—and could not wait for the initial competition among homophone mates to be resolved.

Among the four sets of neighborhood metrics we tested, ND\_SegT and NF\_SegT turned out to be most predictive. The inhibitory effect of neighborhood density (only observed with ND\_SegT) on response time for real syllables is consistent with the more widely observed inhibition from neighborhood density for pseudo syllables. Thus, the current results provide evidence that the one-phoneme/tone difference rule is a more accurate definition of Chinese phonological neighborhoods than the other three alternatives.

The general findings of neighborhood effects from this research can be interpreted in Zhou and Marslen-Wilson’s (1994, 2009) model of the Chinese lexicon. During a real-syllable trial in the auditory lexical decision task, the acoustic input activates simultaneously a number of monosyllables at the phonological form level, including both the target syllable and its phonological neighbors, which are one phoneme/tone away from the target syllable. Each activated syllable further passes the activation to the morphemes (and characters) it is associated with. When the activation of one morpheme (character) surpasses a certain threshold, a decision of “real syllable” will be reached. During the recognition process, the co-activated phonological neighbors act as competitors of the target syllable, suppressing the activation of the target syllable (and the associated morphemes) and delaying a response. Meanwhile, if there exists some very strong competitors (e.g. high-frequency neighbors), they may quickly accumulate enough activation to cause a coincidentally correct “real-syllable” response.

To conclude, the current study found complex effects of phonological neighborhoods (both inhibitory and facilitative) and a facilitative effect of homophone density in spoken word recognition in Mandarin Chinese. It should be noted that the observed effects may be specific to the current task (auditory lexical decision) and experimental design. Our future research will further explore the effects of phonological neighbors and homophones in spoken word recognition in Mandarin Chinese, using different perceptual tasks and focusing more on the possible differences between strong and weak neighbors (Chen & Mirman, 2012). In doing so, we hope to achieve a more comprehensive understanding of the roles of similar-sounding words in lexical processing.

## References

- Amano, Shigeaki & Tadahisa Kondo. 2000. Neighborhood and cohort in lexical processings of Japanese spoken words. *ISCA Tutorial and Research Workshop (ITRW) on Spoken Word Access Processes*.
- Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. *R package version 1(7)*.
- Boersma, Paul & David Weenink. 2010. *Praat: doing phonetics by computer [Computer program], Version 5.1. 44*.
- Cai, Qing & Marc Brysbaert. 2010. SUBTLEX-CH : Chinese word and character frequencies based on film subtitles. *PLoS ONE* 5(6). <https://doi.org/10.1371/journal.pone.0010729>.
- Chen, Hsuan-Chih & Anne Cutler. 1997. Auditory priming in spoken and printed word recognition. In Chen, H.-C. (ed) *The Cognitive Processing of Chinese and Related Asian Languages*, 77–81. Hong Kong: The Chinese University Press.
- Chen, Qi, & Daniel Mirman. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, 119(2), 417–430. <https://doi.org/10.1037/a0030049>
- Cluff, Michael S. & Paul A. Luce. 1990. Similarity neighborhoods of spoken two syllable words: Retroactive effects on multiple activation. *The Journal of the Acoustical Society of America* 87(S1), S125–S126.
- Dufour, Sophie & Ulrich Hans Frauenfelder. 2010. Phonological neighbourhood effects in French spoken-word recognition. *Quarterly Journal of Experimental Psychology (2006)* 63(2). 226–38. doi:10.1080/17470210903308336.
- Gahl, Susanne. 2008. *Time and thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language* 84(3). 474–496. <https://doi.org/10.1007/s13398-014-0173-7.2>.
- Gahl, Susanne, Yao Yao & Keith Johnson. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* 66(4), 789–806.
- Goldinger, Stephen D., Paul A. Luce & David B. Pisoni. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language* 28(5), 501–518.
- Luce, P. A. & D. B. Pisoni. 1998. Recognizing spoken words: the neighborhood activation model. *Ear and Hearing* 19(1). 1–36.
- Luce, Paul A, Stephen D. Goldinger, Edward T. JR. Auer & Michael S. Vitevitch. 2000. Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics* 62(3), 615–625.

- Malins, Jeffrey G. & Marc F. Joanisse. 2010. The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language* 62(4). 407–420. <https://doi.org/10.1016/j.jml.2010.02.004>.
- Malins, Jeffrey G. & Marc F. Joanisse. 2012. Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia* 50(8). Elsevier. 2032–2043. doi:10.1016/j.neuropsychologia.2012.05.002.
- Neergaard, Karl & Chu-Ren Huang. 2016. Graph theoretic approach to Mandarin syllable segmentation. Proceedings of the *15th International Symposium on Chinese Languages and Linguistics (IsCLL-15)*.
- Neergaard, K. D., Hongzhi Xu & Chu-Ren Huang. 2016. Database of Mandarin Neighborhood Statistics. Proceedings of the *Language Resources and Evaluation Conference (LREC 2016)*.
- O'Seaghdha, Padraig G., & Joseph W. Marin. (2000). Phonological competition and cooperation in form-related priming: sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 57–73. <https://doi.org/10.1037//0096-1523>
- Peramunage, Dasun, Sheila E Blumstein, Emily B Myers, Matthew Goldrick & Melissa Baese-Berk. 2011. Phonological Neighborhood Effects in Spoken Word Production : An fMRI Study. *Journal of Cognitive Neuroscience* 23(3). 593–603. doi:10.1162/jocn.2010.21489.
- Sadat, Jasmin, Clara D. Martin, Albert Costa & F. Xavier Alario. 2014. Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology* 68, 33–58. <https://doi.org/10.1016/j.cogpsych.2013.10.001>.
- Schneider, W., A. Eschman & A. Zuccolotto. 2007. E-Prime 2.0 Professional. *Pittsburgh: Psychology Software Tools*.
- R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2013.
- Su, Xinchun & Jinzhan Lin. 2006. A statistic analysis of the counts of Putonghua syllables and the distribution of characters among syllables: A case study of the phonetic notation materials in The Contemporary Chinese Dictionary. *Zhongguo Yuwen* 312, 274–288.
- Tsai, Pei-Tzu. 2007. *The effects of phonological neighborhoods on spoken word recognition in Mandarin Chinese*. Master's thesis at University of Maryland, College Park.
- Vitevitch, Michael S. 2002. The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28(4). 735–747. <https://doi.org/10.1037/0278-7393.28.4.735>.
- Vitevitch, Michael S, Jonna Ambrüster & Shinying Chu. 2004. Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(2). 514–529.
- Vitevitch, Michael S & Paul A Luce. 1999. Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language* 40. 374–408.
- Vitevitch, Michael S & Eva Rodríguez. 2005. Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders* 3(1). 64–73. <https://doi.org/10.1080/14769670400027332>.

- Vitevitch, Michael S & Mitchell S Sommers. 2003. The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition* 31(4). 491–504.
- Vitevitch, Michael S & Melissa K Stamer. 2006. The curious case of competition in Spanish speech production. *Language and Cognitive Processes* 21(6). 760–770. doi:10.1080/01690960500287196.
- Vitevitch, Michael S & Melissa K. Stamer. 2009. The influence of neighborhood density (and neighborhood frequency) in Spanish speech production: A follow-up report. *University of Kansas, Spoken Language Laboratory Technical Report No. 1*, 1–6.
- Wang, Wenna, Xiaojian Li, Ning Ning & John X. Zhang. 2012. The nature of the homophone density effect : An ERP study with Chinese spoken monosyllable homophones. *Neuroscience Letters* 516, 67–71. <https://doi.org/10.1016/j.neulet.2012.03.059>.
- Yao, Yao & Christine Meunier. 2014. Effects of phonological neighborhood density on phonetic variation: The curious case of French. *Oral presentation at the 14th Laboratory Phonology Conference*. Tokyo, Japan.
- Ye, Yun & Cynthia M Connine. 1999. Processing Spoken Chinese : The Role of Tone Information. *Language and Cognitive Processes* 14(5-6). 609–630.
- Yip, Moira. 2002. *Tone*. Cambridge University Press.
- Yoneyama, Kiyoko. 2002. Phonological neighborhoods and phonetic similarity in Japanese word recognition. PhD thesis at the Ohio State University.
- Zeng, Biao & Sven L. Mattys. 2011. Separability of tones and rhymes in Chinese speech perception: Evidence from perceptual migration. *Proc. ICPhS*, vol. 17, 2264–2267.
- Zhou, Xiaolin & William Marslen-Wilson. 1994. Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes* 9(3). 393–422. doi:10.1080/01690969408402125.
- Zhou, Xiaolin & William D. Marslen-Wilson. 2009. Pseudohomophone effects in processing Chinese compound words. *Language and Cognitive Processes* 24(7/8). 1009–1038.
- Ziegler, Johannes C., Mathilde Muneaux & Jonathan Grainger. 2003. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language* 48, 779–793. [https://doi.org/10.1016/S0749-596X\(03\)00006-8](https://doi.org/10.1016/S0749-596X(03)00006-8).