THE EFFECT OF PHONOLOGICAL NEIGHBORS AND HOMOPHONES ON SPOKEN WORD RECOGNITION IN MANDARIN CHINESE

BHAMINI SHARMA

PhD

The Hong Kong Polytechnic University

2020

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

The Effect of Phonological Neighbors and Homophones on Spoken Word Recognition in Mandarin Chinese

Bhamini Sharma

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

February 2020

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

____

Bhamini Sharma

Abstract

This dissertation examined the effects of phonological neighborhood and homophone mates on spoken word recognition in Mandarin. Phonological neighbors are words that sound similar, and homophones are words with exactly the same pronunciation. Neighborhood density (number of phonological neighbors) and neighbor frequency (average frequency of the neighbors) are the main measures of a phonological neighborhood. Homophone density (number of homophones) and homophone frequency (average frequency of the homophone mates) are the main measure of a homophone family.

A lot of research has been devoted to phonological neighborhood effects in English and other European languages, and found overall inhibitory effects of phonological neighborhoods. Words with many neighbors and high frequency neighbors are responded with more errors and longer response time as compared to words with fewer neighbors and lower-frequency neighbors, respectively. But there has not been much research on tone languages. On the other hand, much less is known about the effects of homophone mates on spoken word recognition, probably because most languages like English do not have a large number of homophones in the lexicon.

In this thesis, the language under investigation is Mandarin, a tonal language with a high density of homophone mates. The main goal of the current research is to compare the roles of phonological neighbors and homophone mates in the process of Mandarin spoken word recognition. I conducted two experiments: an auditory lexical decision experiment and an auditory naming experiment. Mixed-effects regression analyses of auditory lexical decision results showed facilitatory effects of phonological neighbor frequency, homophone density, and homophone frequency for real monosyllables and inhibitory effects of neighborhood density for pseudo-syllables. In addition, the models also showed a significant interaction of homophone density and homophone frequency for real monosyllables. The results from the auditory naming experiment showed no significant effects of either phonological neighborhoods or homophone mates.

Taken together, the current research showed both similarity and differences between phonological neighbors and homophone mates in the processing of spoken Chinese words. Implications for models of Mandarin lexicon are discussed.

Publications arising from the thesis

Yao, Y., & Sharma, B. (2017). What is in the neighborhood of a tonal syllable? Evidence from auditory lexical decision in Mandarin Chinese. *Proceedings of the Annual Meeting of the Linguistic Society of America*, Vol 2, Austin, TX, Jan 5-8, 2017.

Sharma, B. (2018). Effects of homophone density on spoken word recognition in Mandarin Chinese. *Proceedings of the Interspeech*, Hyderabad, Sept 2-6, 2018.

Acknowledgements

First and foremost, I would like to sincerely thank my supervisor Dr. Yao for her support and guidance throughout these four years of my graduate studies. Dr. Yao shaped my research insight towards this novel research area.

I would like to extend my heartfelt thanks to my co-supervisors Dr. Leung and Dr. Lau for their feedback and suggestions from time to time.

I would like to thank the other members of my thesis committee, Dr. Angel Chan (chair), Dr. Peggy Mok (external examiner), and Dr. Bin Li (external examiner) for their valuable suggestions and feedback on my thesis.

Thank you is a small word when it comes to my parents, for the amount of support, love, motivation and encouragement you both gave me. From the time you taught me how to hold a pencil and up till now, you were always been my support system. Love you both. You both mean the world to me.

I would like to thank my husband, Akshay, for sticking by my side from the day one of my PhD studies, in all ups and downs. Your support and motivation means a lot.

I feel extremely lucky to have a wonderful circle of friends who made these four years smooth and easy for me. Thank you Wenchun, Hong Lan, Sonya, Liu Chang, Luo Jin, Zhou Mi, and Katrina.

A big thank you to Joe and Mr. Tong for their excellent technical support. I would also like to thank Loretta and Canny from the research office for their admin-related help.

Last but not the least, I would also like to thank all the subjects who participated in the experiments conducted as a part of my dissertation. Without their contribution, this study was not possible.

Table of contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Introduction to the study

Spoken word recognition is a process of recognizing words via auditory modality. This process is very short but complicated. When we hear a word, acoustics of the word activates the sound. The activated sounds further activate word candidates in the mental lexicon. The word candidate that surpasses the threshold activation level gets selected. If the selected word candidate is the target word, successful recognition process completes. On the contrary, if the selected word is not the target word, then the recognition process fails. The recognition performance can be evaluated in terms of accuracy (i.e. whether or not the recognition process is successful) and speed (i.e. how long it takes to complete the recognition process). The speed and accuracy of spoken word recognition may be influenced by a variety of factors. For example, when the listener is in a noisy environment, his or her ability to recognize spoken words will in general deteriorate. On the other hand, the ease or difficulty of recognizing a spoken word may also be influenced by properties of the word. The most well-known example is the effect of word frequency. High frequency words are recognized faster and more accurately than low frequency words (Luce & Pisoni, 1998).

Another type of lexical effects has to do with the presence of words that are somehow similar or related to the target word in the same lexicon. As mentioned above, in the process of spoken word recognition, along with the target word, a group of other non-target words are also activated to different extents. In the case of a failed recognition process, it is one of the non-target co-activated candidates that somehow gets the highest activation (higher than the target word) and therefore is (wrongly) selected as the word spoken. These co-activated words typically share some properties with the target word, in terms of meaning, spelling, or pronunciation. In other words, the lexicon can be viewed as an interconnected network, where words that share properties are connected to each other. These co-activated words influence the processing of the spoken word by either helping or by impeding the target word via the connections.

One of the possible ways to connect related words is through phonological similarity (i.e. similarity in pronunciation). Words with high phonological similarity, i.e. similar-sounding words, are known as phonological neighbors. In practice, phonological neighbors are most commonly defined by the one-phoneme difference rule (Luce & Pisoni, 1998), that is, any two words that differ by one and only one phoneme, by addition, deletion or substitution. For example, *cat* and *mat* are phonological neighbors as they share the phonemes in medial and final positions, and only differ in the initial position. The phonological neighborhood of a word is the set of phonological neighbors the word has. A phonological neighborhood can be measured by neighborhood density and neighborhood frequency (Luce & Pisoni 1998). Neighborhood density is the number of neighbors; neighborhood frequency is the average frequency of the neighbors (Luce & Pisoni, 1998). Thus, a word with many neighbors is in a dense neighborhood, while a word with few neighbors is in a sparse neighborhood.

Previous research has shown that phonological neighbors can influence the process of word recognition. Usually, Luce and Pisoni (1998), and Vitevitch and Luce (1999) found that words from dense neighborhoods take more time and are recognized less accurately as compared to words from sparse neighborhoods. This shows that there is more competition from neighbors in a dense neighborhood than from neighbors in a sparse neighborhood; in other words, neighborhood density has an inhibitory effect on spoken word recognition. Similarly, these studies also show that words with high neighborhood frequency are recognized slower and less accurately than words with low neighborhood frequency, suggesting that high frequency neighbors introduce more competition compared to low frequency neighbors. In other words, neighborhood frequency also has an inhibitory effect on spoken word recognition.

However, these effects of phonological neighborhood on spoken word recognition are not consistent across languages. The seminal works on phonological neighborhoods as reviewed above (Luce & Pisoni, 1998; Vitevitch & Luce, 1999) are all focused on English. Much less is known about how neighborhood effects work in other languages. In fact, previous studies have shown that phonological neighborhood may have different effects in non-English languages, such as Japanese (Amano & Kondo, 2000; Yoneyama, 2002), French (Dufour & Frauenfelder, 2010; Ziegler, Muneaux, & Grainger, 2003), and Spanish (Vitevitch & Rodríguez, 2005). For example, Vitevitch and Rodriguez (2005) studied the effects of phonological neighborhood in Spanish. They found facilitatory effects of neighborhood density and neighborhood frequency. These findings are in contradiction to the results found in English. These conflicting effects could be due to language specific properties of the neighborhood structure, morphology, and the structure of word in the language.

Furthermore, previous studies on neighborhood effects have, in general, only examined non-tonal languages. What will happen in a tonal language? The crucial feature of a tonal language is the use of pitch patterns to distinguish word meanings (i.e. lexical tones). The presence of lexical tones raises an important question, that is, how to define a phonological neighborhood in a tonal language? Should tone be considered in the evaluation of phonological similarity? If so, how should the consideration of tones be integrated with the consideration of segments? Take Chinese as an example. Lexical tones are used in all the varieties of Chinese. The most widely-studied Chinese varieties are Mandarin, which has four lexical tones (Duanmu, 2007), and Cantonese, which has six lexical tones (Yip, 2002). Previous studies have shown that tones and segments are equally important for Chinese spoken word recognition (Malins & Joanisse, 2012; McBride-Chang et al., 2008). A small number of studies have examined the effects of phonological neighborhoods in Mandarin (Neergaard, 2018; Tsai, 2007) and Cantonese (Kirby & Yu, 2007). But these studies used different definitions of phonological neighborhoods, and reported mixed results regarding the effects of phonological neighborhoods on Chinese word recognition.

In this dissertation, I further the investigation of phonological neighborhood effects in Mandarin, following Neergaard's (2018) general methods, I test multiple possible definitions of phonological neighborhoods with two spoken word recognition tasks: auditory lexical decision and auditory word naming. However, both experimental design and statistical analysis are significantly improved in my study. Instead of using a small set of word stimuli, I use the complete set of possible syllables that are available in the language, so that the experimental results paint a more comprehensive picture of the effects under investigation. The statistical analysis used in this dissertation is carefully planned to control for all the factors that may affect the spoken word recognition process, so that the critical effects can be interpreted unambiguously.

In addition to phonological neighbors, this thesis will also include a discussion of the effects of homophones. In the existing literature of phonological neighborhoods (Luce & Pisoni, 1998; Vitevitch & Luce, 1999), the focus has always been on words that sound similar but not exactly the same. What about words that sound exactly the same as the target word? Do homophones also play a role in the process of spoken word recognition? These questions are difficult to discuss with data from English (or many other languages), because there are, in general, not many homophones in the language. Mandarin provides a rare opportunity to address these questions. With a small syllable inventory of only around 1300 monosyllables with tone gives rise to a lot of homophones in the language. Mandarin has a wide range of homophones ranging till 40 homophones. For comparison, English words have around 2-4 homophones. There have been few studies in Mandarin looking at the effects of homophones that found inhibitory effects on spoken word recognition (Wang, Li, Ning, & Zhang, 2012; Zhou, 2015). Further, there have been few studies that have examined the effects of homophone density and its interactions within the frequency, in Mandarin (Li, Fang, & Lou, 2011; Zhou, 2015). These studies suggest interactions between homophone density, and frequency. However, these interactions are not very well understood. Therefore, to contribute to this set of literature and further the investigation on the effects of homophones using an exhaustive dataset of 1259 Mandarin monosyllables, the current study was conducted.

In sum, this dissertation addresses: First, how to define phonological neighborhoods for a tonal language like Mandarin and what are the effects of phonological neighbors on Mandarin spoken word recognition? Second, what are the effects of homophones on Mandarin spoken word recognition? Do homophone mates have the same effects as phonological neighbors? Third, how to explain the effects of phonological neighbors and homophone mates in Mandarin in theoretical models of spoken word recognition?

1.2 Dissertation outline

This dissertation is organized as follows: Chapter 2: Literature Review, reviews the relevant literature on models of spoken word recognition and the effects of phonological neighborhood and homophones on spoken word recognition. Chapter 3: Research Methods, introduces the general methods (e.g. stimuli preparation, procedures of model building and analysis) that are shared by both experimental studies. Chapter 4: Auditory lexical decision, presents the detailed experimental procedure and results of the auditory lexical decision experiment. Chapter 5: Auditory naming, presents the detailed experimental procedure and results of the auditory naming experiment. Chapter 6: Discussion and conclusion, contains a general discussion of the experimental results and the theoretical implications. Limitations of the current study and future directions are also discussed towards the end of Chapter 6: Discussion and conclusion.

1.3 Significance of the dissertation

The significance of this dissertation lies in three aspects. First, this dissertation will provide a comprehensive view of the effects of phonological neighborhoods on spoken word recognition in Mandarin. In the current dissertation, I report results from well-controlled experiments and well-grounded data analysis. Findings from this dissertation will illuminate the nature of phonological neighborhood effects in Mandarin; they will also extend the existing literature on phonological neighborhoods by contributing valuable insight from lexical processing in a tonal language.

Second, this dissertation will provide important insights into the effects of homophones on spoken word recognition. This issue is rarely addressed in the literature, because most of the commonly researched languages like English do not have many homophones. Mandarin, which has an abundance of homophones, gives a unique opportunity to investigate this issue.

Last but not the least, this dissertation explores the continuum of phonological similarity and how different degrees of similarity may affect lexical processing. This dissertation investigates both similar-sounding neighbors and same-sounding homophones. To the best of my knowledge, this is the first study that compared the effects of these two types of phonological similarity in spoken word recognition. Based on the nature of effects of phonological neighbors and homophones, this dissertation tries to associate the experiment results with the existing theoretical models of spoken word recognition to provide theoretical explanation to the effects found in the current dissertation.

Chapter 2: Literature Review

The process of recognizing words is known as spoken word recognition. In the recognition process, the incoming auditory information is mapped onto the word representations stored in the listener's mental lexicon. This mapping of auditory input onto mental lexicon would have been done very easily if we lived in an ideal world with no noise, no variation in speech across people. However, that's not the case. Although, this process is done effortlessly by the listener, but the underlying mechanism is quite complex. there have been several theoretical models of the underlying mechanisms of spoken word recognition. In this chapter, I review the main theoretical models of spoken word recognition in Section 2.1 (Models of spoken word recognition). There are many factors that can influence the speed and accuracy of spoken word recognition. The most relevant factors for this dissertation are the effects of phonological neighborhoods and the effects of homophones. In Sections 2.2 (Phonological neighborhood) and 2.3 (Homophones), I review the relevant literature of these effects.

2.1 Models of spoken word recognition

There have been several models and/or theories that have attempted to explain the process of spoken word recognition. These models lay foundation on how acoustic signals are perceived as words in the mental lexicon. Some of the seminal models of spoken word recognition are Logogen model (Morton, 1969), Cohort theory (Marslen-Wilson, 1987), Shortlist model (Norris, 1994), TRACE model (McClelland & Elman, 1986), neighborhood activation model (Luce & Pisoni, 1998) etc.

Morton (1969) proposed *Logogen Model* of word recognition. This model was based on the assumption that each word is associated with a logogen. Logogens are units that contain information about a word in terms of phonetics, semantics, and syntax. When a word is presented, logogens that matches information with the input gets activated. With more match between the input speech and the logogen, the activation level of the logogen rises. This activation was also enhanced by the context. The first logogen that reaches the threshold of activation is recognized. The threshold of recognition varies depending on various factors. For example, word frequency: high frequency words have lower threshold compared to low frequency words. This results in faster recognition of high frequency words because the raise in activation required to reach the threshold is less compared to low frequency words. The model does consider the effects of context and word frequency on the word recognition process. However, the model strictly depends on the activation levels. It does not matter if there are other words possible or not. Word is only recognized when the activation reaches the threshold. Also, the model only talks about the uni-directional rise in the activation level. This indicates that a non-word will only be realized by the end of the input signal, which is usually not the case.

Later, Marslen-Wilson (1987) introduced the Cohort model of spoken word recognition. This model is based on three stages of processing, namely access, selection, and integration. During access stage, acoustic-phonetic features of the input speech signal activate a group of word candidates in the mental lexicon referred to as *cohort*. In the selection stage, words that differ from the input speech signal by more than a one feature gets eliminated from the cohort. Syntactic and semantic properties are incorporated by the integration function. For example, words that mismatch in context get eliminated from the cohort. The words in each cohort compete with one another until only one word is left in the mental lexicon. This model emphasizes on the early recognition of words. Recognition of a word occurs at the point at which a word is uniquely different from rest of the words in the cohort. Therefore, the word can be recognized before the word ends based on their uniqueness points. This model also supports the word frequency aspect that implies a faster recognition for high frequency words. However, the model suggests that context does not play a role in the formation of cohort. Only when a cohort of word is selected, contextual information helps in cutting down the candidates. Also, the model is based strictly on the sequence of input signal from initial to final position. The model fails to account for misperceptions, i.e. when the correct word candidate gets eliminated or fails to make an entry in the selected cohort. Also, the selection process does not make use of contextual information in the selection of cohort of words, which also leads to misperceptions.

TRACE model (McClellland & Elman, 1986) was the first computational model of word recognition. This model was based on the interactive-activation framework. This means that the higher levels of contextual information can interact with the lower processing levels at all times. The model comprises of three levels, (1) feature level, (2) phoneme level, and (3) word level. Feature level consists of information about various speech characteristics such as voicing, burst, etc. Phoneme level and word level represents various phonemes and words respectively. Among these levels there were two types of connections: excitatory connections, and inhibitory connections. The excitatory connections run across the three levels while the inhibitory connections are active within the level. The excitatory connections transmit information from one level to another. When a word is presented, the incoming speech signal activates corresponding speech features in the feature level. The activated features activate phonemes in the phoneme level, which in turn activates the word in the word level through the excitatory connections. Simultaneous to this process, inhibitory connections within the levels inhibit or restrict the activation of certain phonemes and word at their respective levels via inhibitory connections. Since the model is not strictly sequential like the cohort model, this allows the model to use context to recognize correct word candidates. It should be noted that the model assumes all the features at the feature level to be equally important for identification and classifies all the features on a same rating scale. Also, the model indulges in unreasonable duplication of network at each level of processing, which can be handled only by a small lexicon.

Shortlist model by Norris (1994) was developed with an aim to instill the merits of TRACE. It consists of two stages. In the first stage, based on the input information word candidates were selected or shortlisted. In the second stage, selected candidate competes with each other. The process of shortlisting the word candidates is entirely based on the information provided by the input speech signal. The word candidates with lowest activation get eliminated to make space for the higher activated candidates in case there are many candidates. So, the process of updating activation scores is a continuous process to maintain the size. This makes it possible to consider a realistic size of lexicon. The shortlisted word candidates are wired with the overlapping word candidates through inhibitory links. These shortlisted candidates then compete with each other. The word with highest level of activation gets identified. The model does account for word frequency effect.

Neighborhood Activation Model (NAM; Luce & Pisoni, 1998) of spoken word recognition is based on acoustic processing as well as processing due to formation of groups of identical words. According to NAM, when a word is heard, a set of acoustic-phonetic patterns that represents a word gets activated. These set of words are similar sounding that differs maximally by one-phoneme from the input word. The activation level depends on the match with the input signal. These activation levels then activate word decision units. The activation of word decision unit depends on the acoustic-phonetic patterns and higher-level lexical information like word frequency. A word is recognized when the decision unit crosses certain threshold level. The model explains the structural organization of acoustic stimulus and the role of neighborhood structures and word frequency in word recognition in the mental lexicon. Most of the previous models on spoken word recognition considered word frequency effects intrinsic to activation process and thus could not fully explain these effects. However, NAM suggests frequency bias on the decision unit to explain the effects of frequency on spoken word recognition. Also, none of the previous models on spoken word recognition discuss about the structural organization of the mental lexicon and the influence it has on the processing of words.

All the theoretical models reviewed above recognize that multiple word candidates are activated in the process of recognition. The terms used for the candidates may be different across models (e.g. *cohort* in the cohort model, and *neighbor* in the NAM), but all the models agree that the candidates usually share something in common with the target word, in one or more lexical aspects (pronunciation, spelling, meaning, etc.). In this dissertation, I focus on the aspect of pronunciation and the effects of co-activated phonologically similar word candidates in the process of spoken word recognition. For this purpose, I adopt a model that is most similar to *NAM*. In the next two Sections, I will review relevant empirical evidence for the effects of phonologically similar or even identical word candidates.

2.2 Phonological neighborhood

2.2.1 Quantifying phonological neighborhood

As shown in the discussion above, many models of spoken word recognition assume the co-activation of a group of words along with the stimulus word. These co-activated words are often phonologically similar to the stimulus word. These phonologically similar words are also known as phonological neighbors. But, how to define phonological neighbors and how to quantify the amount of co-activation produced by these co-activated phonological neighbors is quite difficult. There are several ways or rules proposed to define similarity of sounds. Some of the commonly used rules are neighborhood word probability rule (Luce & Pisoni, 1998), phi-square rule (Iverson, Bernstein, & Auer Jr, 1998) and one-phoneme difference rule (Luce & Pisoni, 1998). Luce and Pisoni (1998) proposed the neighborhood probability rule, which describes the amount of influence from the neighbors based on the confusability of sounds in the presence of noise. The more similar two words sounds, the more confusing they are. Neighborhood probability rule provides a measure to predict the probability of identifying the stimulus word, given all other words in the lexicon. The calculation of neighborhood probability involves two formulas. Formula (1) calculates the confusability between the stimulus word and any other given word in the lexicon by multiplying the conditional probabilities of recognizing every phoneme in the stimulus word as the phoneme in the corresponding position in the non-stimulus word. In Luce and Pisoni's work, the word-based confusability measure is referred to as neighborhood word probability (NWP). The more confusable the stimulus word is with the neighbor word, the higher the NWP. In the most extreme case, the NWP can be as high as 1, when the stimulus word is always mis-recognized as the neighbor word.

(1)

$$\boxed{NWP = \prod_{i=1}^{n} p(PN_i|PS_i)}$$

where PNi is the *i*th phoneme of the neighbor, PSi is the *i*th phoneme of the stimulus word and *n* is the number of phonemes.

For example, the probability of recognizing the stimulus word "cat" as "bad" can be calculated by multiplying the probability of identifying /k/ as /b/, the probability of identifying /æ/ as /æ/, and the probability of identifying /t/ as /d/. See formula (2).

(2) *p*(bæd|kæt) = *p*(b|k) * *p*(æ|æ) * *p*(d|t)

The calculation of neighborhood probability involves another formula (see formula (3)), which computes the amount of overall competition posed by the lexicon onto the stimulus word by summing the NWPs with all other words in the lexicon (i.e. the complete set of possible neighbor words). Here, for simplicity, we refer to this sum of NWPs as NWP density. It should be noted that neighborhood probability rule does not make categorical distinction between neighbors and non-neighbors. Instead all the words in the lexicon are considered as potential neighbors, but the amount of influence of each word to the stimulus word may vary depending upon the similarity a word shares with the stimulus word. For example, in a lexicon of N words, the overall completion on stimulus word can be expressed as given in formula (3).

(3) NWP density = $p(\text{word}_1|\text{stimulusword}) + p(\text{word}_2|\text{stimulusword}) + \ldots + p(\text{word}_N|\text{stimulusword})$

As shown above, NWP density gives the overall probability of a stimulus word being misrecognized when only phonemic content is being considered. This measure does not consider the possible influence of word frequency or the relative; neither does it compare the probability of correctly identifying the stimulus word (stimulus word probability (SWP); see formula (4)). Luce and Pisoni proposed a more comprehensive measure, the frequency-weighted neighborhood probability rule (FWNPR), which calculates the overall probability of a word being correctly identified (i.e. p(ID)) given all other words in the lexicon. In order to capture the potentially higher influence from frequently-used neighbors, one can calculate frequency-weighted neighborhood probability rule (FWNPR) by weighting the individual SWP and NWP with the frequency of the stimulus word and the frequency of the neighbor word respectively (see formula (5)).

(4)

$$\boxed{SWP = \prod_{i=1}^{n} \quad p(PS_i|PS_i)}$$

(5)

$$\boxed{FWNPR = p(ID) = \frac{\prod_{i=1}^{n} \quad p(PS_i) * Freq_s}{\{PS_i)] * Freq_s\} + \sum_{j=1}^{nn} \quad \{PS_i)] * Freq_{N_j}\}}}$$

where PS_i is the probability of the *i*th phoneme of the stimulus word, PN_{ij} is the probability of the *i*th phoneme of the *j*th neighbor, *n* is the number of phonemes in the stimulus words and the neighbor word, $Freq_s$ is the frequency of the stimulus words, $Freq_{Nj}$ is the frequency of the *j*th neighbor, and *nn* is the number of neighbors.

Both NWP density and FWNPR have been used in the literature. Luce and Pisoni (1998) tested the effectiveness of FWNPR in their investigation of phonological neighborhood effects in spoken word recognition in English. They conducted a perceptual identification experiment with 918 three-phoneme monosyllables. The auditory stimuli were embedded in white noise at +15 dB, +5 dB, and -5 dB signal to noise ratios, in order to increase the level of difficulty of the identification task and the chance of observing perceptual errors. Ninety participants participated in this experiment, where the participants were asked to provide their best guess for each word they heard in a typed response. The participant's performance was evaluated in terms of accuracy. FWNPR was calculated for each stimulus word, following the formula in (5); usage frequency of words was collected from Kucera and Francis (1967). Results from this experiment revealed that FWNPR significantly correlated with identification scores on all three SNR conditions. Identification scores were high (mean = 64.03%) when frequency-weighted neighborhood word probability was low and frequency-weighted stimulus word probability was high. While identification scores were low (mean = 37.76%) when frequency-weighted neighborhood word probability was high and frequency-weighted stimulus word probability was low. Results from this study show that FWNPR does predict effects of phonological neighborhood on word recognition.

NWP density has also been shown to be effective for predicting phonological neighborhood effects (e.g. Strand & Sommers (2011)). Strand and Sommers (2011) tested neighborhood probability measures along with other neighborhood measures in auditory spoken word recognition. The stimuli were 180 CVC words randomly selected from English lexicon project (Balota et al., 2007). Two tasks were conducted: phoneme identification and word naming. In phoneme identification, participants were presented with series of audio clips of a phoneme and they were instructed to identify the phoneme in a forced-choice task. In word naming task, participants were presented audio clips of word and they were asked to repeat the word. The authors used NWP density as a measure to quantify neighborhood. Seventy-two participants took part in this study. There was a significant negative correlation between recognition accuracy and density measures ($r = -0.16$, $p < 0.05$). Words with less lexical competition were identified more accurately compared to words with higher lexical competition. This study also tested other neighborhood measures, which will be discussed in later parts of this Section.

Although both NWP density and FWNPR are effective measures to predict the effects of word recognition, they pose a few limitations in its use of phoneme-based conditional probability as a means to estimate perceptual similarity. Most importantly, while these conditional probabilities do capture the similarity between two sounds, the exact value of the conditional probability also depends on how many other perceptually similar phonemes there are. For example, if /f/ and /v/ are a pair of highly similar phonemes that are hard to distinguish, then the identification accuracy will be around 50% if listeners are guessing at the chance level (see Table 1). However, if /tʃ/, /dʒ/, /ʃ/ and /ʒ/ are highly similar to one another, then the identification accuracy will be around 25% if the listeners are guessing at the chance level (see Table 1) due to four perceptually similar options available. In other words, it may seem that, /f/ and /v/ are twice as confusing as /tʃ/ and /dʒ/, or /ʃ/ and /ʒ/, or /tʃ/ and /ʃ/, or /dʒ/ and /ʒ/. But in fact, all these pairs are at the highest level of similarity that causes total confusion.

Another limitation in conditional probability is the influence of response bias. As mentioned before that conditional probabilities capture similarity between two sounds, however, if participant response is biased towards a certain phoneme, it would appear that this phoneme is highly confusable with many other phonemes in the language. If that happens, then many phonemes in the language appear to be highly confusable with this phoneme. As a resultant, confusability among truly confusable phonemes will not be observed. In other words, the confusability among these other phonemes will be masked. For example, in a language people tend to select phoneme /h/ more often for unknown reason. When that happens, it becomes difficult to observe similarity between truly perceptually similar phonemes like /f/ and /v/. In that case, the conditional probability of /f/ and /v/ gets masked due to the response bias of participants to choose /h/ over truly perceptual phonemes like /f/ and /v/.

On the other hand, phi-square statistics that compares the response distributions of each response (entire row of Table 1) remains unaffected by frequency of each phoneme (individual cell of Table 1) that overcomes the problem of response bias.

Table 1: A toy confusion matrix.

	m	l	f	v	tʃ	ʃ	dʒ	ʒ	h	Total
m	95	2	0	3	0	0	0	0	0	100
l	2	85	4	8	0	0	0	0	1	100
f	0	0	50	50	0	0	0	0	0	100
v	0	0	50	50	0	0	0	0	0	100
tʃ	0	0	0	0	25	25	25	25	0	100
ʃ	0	0	0	0	25	25	25	25	0	100
dʒ	0	0	0	0	25	25	25	25	0	100
ʒ	0	0	0	0	25	25	25	25	0	100
h	0	0	5	0	0	2	2	1	90	100

In order to overcome the limitation of conditional probability used in neighborhood word probability rule, Iverson et al. (1998) proposed a phi-square rule (based on phi-square statistic). As the phi-square statistic compares response distribution across all response options (entire row of Table 1) instead of individual responses (each cell in Table 1), the resultant is said to be unaffected by response percentages (individual cell of Table 1) and response bias (individual cell of Table 1).

The calculation of phi-square statistic involves three formulas. The first formula estimates the similarity for each pair of phonemes. This similarity estimate is done by calculating the phi-square statistic on the distribution of responses given to these two phonemes (see formula (6)).

(6)

$$\varphi^2 = \sqrt{\frac{\sum \frac{(x_i - E(x_i))^2}{E(x_i)} + \sum \frac{(y_i - E(y_i))^2}{E(y_i)}}{N}}$$

Where, x_i and y_i are the frequencies of identification of phonemes x and y as response category i, $E(x_i)$ and $E(y_i)$ are the expected frequencies of response for x_i and y_i if phonemes x and y are identical as category i, and N is the total of all responses to phoneme x and y.

The expected values, $E(x_i)$ and $E(y_i)$, are determined by summing the frequency with which phoneme x was identified as category i and the frequency with which phoneme y was identified as category i, divided by 2. Therefore, $E(x_i)$ and $E(y_i)$ are always identical. The idea behind this being that phoneme x and y should be identified as the members of a given category with equal frequency if they are perceptually identical. The phi-square value equals 0, when the distribution of response to two phonemes is identical i.e. participants select each response alternative equally for both phonemes *x* and *y*. For example, phonemes /f/ and /v/ in Table 1. The phi-square statistic equals 1, when there is no overlap in the response distribution of two phonemes i.e. participants do not make use of the same response categories for phonemes *x* and *y*. For example, phonemes /f/ and /tʃ/ in Table 1. Unlike phoneme-based conditional probabilities that estimates the confusion between tow phonemes, the phi-square value estimates the similarity in pattern of responses to the two phonemes.

Just like NWP (see formula (2)), the second formula calculates the similarity between the stimulus word and another word in the lexicon by multiplying the phi-square values of recognizing a phoneme in the stimulus word as the phoneme in the same position in the non-stimulus word (see the formula in (7)). Formula (7) gives the expression of computing phi-square value of the stimulus word to another word in the lexicon.

For example, the phi-square similarity of "cat" and "bad" can be calculated by multiplying the phi-square value of identifying /k/ as /b/ with phi-square value of identifying /æ/ as /æ/ with phi-square value of identifying /t/ as /d/. See formula (7)

(7) $\varphi\ (\text{bæd}|\text{kæt}) = \varphi\ (\text{b}|\text{k}) * \varphi\ (\text{æ}|\text{æ}) * \varphi\ (\text{d}|\text{t})$

Just like NWP density, the last formula (see formula (8)) calculate the overall competition posed by the lexicon onto the stimulus word i.e. phi-square density. All words in the lexicon are considered as potential neighbors. For a given stimulus word, its overall competition in the lexicon can be quantified by summing the phi-square values of stimulus word to all the other words in the lexicon. For example, in a lexicon of N words, the overall completion on stimulus word can be expressed as given in formula (8).

(8) phi-square density = φ ($word_1$|stimulusword) + φ ($word_2$|stimulusword) +…+ φ ($word_N$|stimulusword)

Strand (2014) provides an online Phi-square lexical competition database (Phi-Lex) that gives access to auditory and visual lexical competition in English. Phi-Lex database contains three and four-phoneme English words extracted from English Lexicon Project (Balota et al., 2007). The phi- square values in the database were based on the confusion matrix data from Luce (1986).

There are similarities between Luce and Pisoni's neighborhood probability rule and Iverson's phi-square rule. Both the rules are based on perceptual confusability of sounds. Both the rules do not make a binary distinction between neighbors and non-neighbors. The main difference between neighborhood probability rule and phi-square rule is that neighborhood probability rule describes the confusion between two phonemes on how likely these two phonemes are confused between each other whereas phi-square rule compares the responses to the two phonemes and describes how similar the response patterns are.

Strand and Sommers (2011) tested phi-square measures along with other neighborhood measures in auditory spoken word recognition. Similar to the results with NWP density reviewed above, phi-square density also proved to be significant predictor of word identification accuracy. Specifically, there was a significant negative correlation between recognition accuracy and density measures (r - -0.32, p <0.01). In other words, words with less lexical competition were identified more accurately compared to words with higher lexical competition.

In addition to neighborhood probability rule, Luce and Pisoni (1998) also introduced one-phoneme difference rule. One-phoneme difference rule is the most commonly and widely used working definition of phonological neighbors. According to one-phoneme difference rule, any two words that differ by one phoneme, by addition, deletion, or substitution, are considered phonological neighbors. For example, /bæt/, /kæp/, /æt/, /kæst/ and /kit/ are phonological neighbors of /kæt/, as they differ only by one phoneme. /bæt/, /kit/ and /kæp/ differs from /kæt/ by one phoneme via substitution at initial, medial and final position respectively; /æt/ and /kæt/ differs by deletion of a phoneme; /kæst/ and /kæt/ differ by addition of a phoneme. Webster's Pocket dictionary (Webster's Seventh Collegiate Dictionary, 1967) was used to estimate the neighborhood measures. Luce and Pisoni (1998) used only familiar words to estimate phonological neighborhood using one-phoneme difference rule in order to exclude words that are very infrequent. The frequency count was based on Kucera and Francis corpus (Francis & Kucera, 1967). Familiarity rating were obtained from Nusbaum et al.'s study (Nusbaum, Pisoni, & Davis, 1984) where subjects rated familiarity on a seven-point scale, ranging from "don't know the word (1) to "know the word and know its meaning" (7). Luce and Pisoni (1998) chose words with familiarity rating of 5.5 or above to estimate neighborhood. The neighborhood metrics used in many English studies is from Hoosier mental lexicon

(Nusbaum et al., 1984). Neighborhood matrix in Hoosier mental lexicon is based on 20,000 words from Webster's pocket dictionary (Webster's Seventh Collegiate Dictionary, 1967).

Apart from English, there are lexical database in other languages too. The lexical database for monosyllabic French words is LEXOP (Peereman, 1999). LEXOP contains 2449 monosyllabic word forms extracted from BRULEX (Content, Mousty, & Radeau, 1990), a psycholinguistic database of 35,746 words for French. LEXOP database constitutes details on phonological and orthographical characteristics. The phonological and orthographical neighborhood in LEXOP was defined by one-phoneme/letter substitution. This database does not mention anything about familiarity ratings. The lexical database in Spanish is LEXESP: Léxico informatizado del español (Sebastián-Gallés, 2000). The lexical database in Japanese is the NTT database (Amano & Kondo, 2003) that does consider the familiarity ratings.

In NWP density and phi-square density, the amount of influence by neighbors is quantified by the conditional probabilities and phi-square values respectively, as given in formula (3) and (6). Under the one-phoneme difference rule, which makes categorical distinction between neighbor and non-neighbors, the two most often used ways to quantify the amount of influence from the neighbors are neighborhood density (i.e. the number of neighbors in the neighborhood) and neighborhood frequency (i.e. the average frequency of neighbors). A word with many neighbors is said to be in a dense neighborhood, while a word with few neighbors is said to be in a sparse neighborhood.

Luce and Pisoni (1998) examined the effects of one-phoneme difference rule in auditory lexical decision and auditory naming (the same study also examined neighborhood probability rule in perceptual identification in noise, which has been reviewed earlier under neighborhood probability rule). In their auditory lexical decision task, participants were presented with auditory stimulus and were asked to decide whether the stimulus was a real word or a non-word as quick and accurately as possible. In this task both accuracy as well as speed of response were measured. 918 monosyllabic real words (same set as the stimuli word in perceptual identification task) along with 304 three-phoneme non-words were used as stimuli. The real words were divided into high (mean = 254.12) and low frequency (mean = 5.22), high (mean = 21.92) and low neighborhood density (mean = 11.07), and high (mean = 370.32) and low neighborhood frequency (mean = 46.29). Non-words were divided into high (mean = 17.78) and low neighborhood density (mean = 8.10), and high (mean = 156.96) and low neighborhood frequency (mean = 11.84). Three stimulus files were constructed with 306 real words and 304 non-words. Each set of stimulus file was present to 10 participants. In total, thirty participants took part in this experiment. Results from real words suggest faster and more accurate responses for high frequency real words (mean reaction time = 390 ms; mean accuracy = 93.43%) than low frequency real words (mean reaction time = 445 ms; mean accuracy = 86.03%); words with high neighborhood frequency were responded slower and less accurately (mean reaction time = 426.25 ms; mean accuracy = 89.04%) than words with low neighborhood frequency (mean reaction time = 408.75 ms; mean accuracy = 90.42%); and words with high neighborhood density were responded more accurately and slowly (mean reaction time = 424.25 ms; mean accuracy = 91.4%) compared to words with low neighborhood density(mean reaction time = 410.75 ms; mean accuracy = 88.04%). Results from non-words suggest significant phonological neighborhood effects. Non-words with high neighborhood density were responded slowly and less accurately (mean reaction time = 451 ms; mean accuracy = 86.55%) compared to non-words with low neighborhood density (mean reaction time = 411.5 ms; mean accuracy = 90.03%). Similar effects were seen for neighborhood frequency. Non-words with high neighborhood frequency were responded slowly and less accurately (mean reaction time = 437 ms; mean accuracy = 86.85%) compared to non-words with low neighborhood frequency (mean reaction time = 425.5 ms; mean accuracy = 89.74%).

In their auditory naming task, participants were presented with auditory stimulus and were asked to repeat the word as quick and accurately as possible. 400 CVC monosyllabic words were chosen from their real-word stimuli used in auditory lexical decision task. They selected words to construct 8 cells with 50 words in each cell. These eight cells were orthogonally constructed combining high and low levels of word frequency, neighborhood density and neighborhood frequency. They selected stimuli words on the basis of an algorithm that first rank-ordered each of the 918 real words on each of their three independent variables. Further, in order to ensure that cells that were matched on a given variable (e.g. both high neighborhood density) were maximally alike and that cells intend to differ on a given variable (e.g. one high and one low neighborhood density) were maximally different, they employed a method that used minimized and maximized squared deviations of successively ranked words. The words were divided into high (mean = 145.95) and low frequency (mean = 4.33), high (mean = 22.12) and low neighborhood density (mean = 11.44), and high (mean = 245.17) and low neighborhood frequency (mean = 60.50). Eighteen participants took part in this experiment. Results showed that words with high neighborhood density were named slowly and less accurately (mean reaction time = 822 ms; mean accuracy = 97.8%) than words with low neighborhood density (mean reaction time = 720 ms; mean accuracy = 98.08%). No effects of neighborhood frequency and word frequency were found.

Overall, it was found that words from dense neighborhood had lower accuracy rates and took longer to respond compared to words from sparse neighborhood. This indicated that phonological neighbors act like competitors and pose competition to the stimulus word during word recognition process. Therefore, words with high neighborhood density face more competition due to more competitors resulting in lower accuracy and longer latency. Similar effect was found for neighborhood frequency i.e. words with higher neighborhood frequency were responded slower and less accurately compared to words with lower neighborhood frequency. Luce and Pisoni (1998) found that results from frequency weighted neighborhood probability rule and one-phoneme difference rule showed similar effects. The results were comparable and they did not differ significantly between the two measures of defining neighborhood.

Strand and Sommer (2011) also examined the effects of neighborhood measures under the one-phoneme difference rule. They found that one-phoneme difference rule significantly predicts the effects of phonological neighborhood ($r = -0.20$, $p < 0.05$). In general, they found that neighborhood measures from all the three rules, neighborhood probability rule ($r = -0.16$, $p < 0.01$), phi-square rule ($r = -0.32$, $p <$

0.05) and one-phoneme difference rule ($r = -0.20$, $p < 0.05$), are significantly effective in predicting the effects of phonological neighborhood. Phi-square density accounted for an additional variance compared to NWP density and neighborhood density. In general, all measures of neighborhood density significantly predicted the effects of phonological neighborhood.

So far one-phoneme difference rule is the most often used neighborhood definition in the literature, due to its simplicity and effectiveness. In this dissertation, I use a modified version of one-phoneme difference rule for investigating the effects of similar-sounding words on Chinese word recognition, taking into consideration the syllable structure of Chinese lexical items. Details on the modified version of one-phoneme difference rule used in this dissertation are presented in Chapter 3: Research Methods. In the next Section (Section 2.2.2 Phonological neighborhood effects), I am going to review additional experimental evidence on phonological neighborhood assuming the one-phoneme difference rule.

2.2.2 Phonological neighborhood effects

There has been a voluminous body of literature on the effect of phonological neighborhood on spoken word recognition (Amano & Kondo, 2000; Dufour & Frauenfelder, 2010; Luce & Pisoni, 1998; Vitevitch, 2002; Vitevitch & Luce, 1998, 1999; Vitevitch & Rodríguez, 2005; Vitevitch, Stamer, & Sereno, 2008; Yoneyama, 2002; Ziegler et al., 2003). Most of these studies have used one-phoneme difference rule to define phonological neighborhood. In this Section, I am reviewing studies using one-phoneme difference rule to define phonological neighborhood.

Apart from the seminal work by Luce and Pisoni (1998) that suggested that words from sparse neighborhood are recognized faster and with less errors compared to words from dense neighborhood, their group has published a few more studies along these lines using different experiment paradigms. Studies from their research group reinforced their previous findings via auditory naming task (Vitevitch & Luce, 1998), speeded-match, auditory lexical decision, and semantic categorization task (Vitevitch & Luce, 1999). In all these studies, words with high neighborhood density were found to have longer reaction times and lower accuracy rates than words with low neighborhood density.

Nevertheless, the investigation of neighborhood density effects is often complicated by the correlation between neighborhood density and other lexical measures. The most highly correlated variable with neighborhood density is phonotactic probability. Phonotactic probability is the frequency with which a segment or a sequence of segments occurs at a particular position in a word. Positional segment frequency and positional biphone frequency are the most commonly used measures to determine phonotactic probability: Positional segment frequency is the frequency of occurrence of a particular segment in a given position in a word, while positional biphone frequency is the segment-to-segment co-occurrence probability. Phonotactic probability has been shown to positively correlate with neighborhood density (for example, Vitevitch, Luce, Pisoni and Auer (1999) found $r = 0.61$, $p < 0.0001$ for the English lexicon), because words that have a lot of neighbors have segments or biphones that occur frequently in the lexicon.

Vitevitch and Luce (1998) investigated the effects of phonological neighborhood and phonotactic probability. Two-hundred forty nonwords and one-hundred fifty real words (3-phoneme) divided into high neighborhood density, high probability real-words (mean frequency weighted neighborhood density = 56.43) and nonwords (mean frequency weighted neighborhood density = 44.61), and low neighborhood density, low probability real-words (mean frequency weighted neighborhood density = 40.00) and nonwords (mean frequency weighted neighborhood density = 13.46) were used as stimuli. Thirty native English speakers participated in a standard auditory naming task. The authors found that overall real words were responded faster ($F_1(1,28) = 17.76$, $p < 0.001$; $F_2(1,386) = 447.04$, $p < 0.001$) than nonwords. Furthermore, the authors also looked at words and nonwords separately and found opposite effects. Among real words, low-probability and low-density words were responded faster ($F_1(1,14) = 16.40$, $p < 0.001$; $F_2(1,148) = 3.89$, $p < 0.05$) than high-probability and high-density words; but for nonwords, high-probability and high-density words were responded faster ($F_1(1,14) = 4.56$, $p < 0.05$; $F_2(1,238) = 3.80$, $p < 0.05$) than low-probability and low-density words. To explain these results, the authors proposed that real word processing is mainly affected by neighborhood density, in an inhibitory manner as shown in previous research (Luce & Pisoni, 1998); on the other hand, nonword processing is mainly affected by phonotactic probability, which acted in a facilitative manner as shown in (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). As a result, words with high-density (also high-probability) have greater processing difficulty than words with low-density (also low-probability); while nonwords with high-probability (also high-density) are easier to process than nonwords with low-probability (also low-density).

However, the account proposed by Vitevitch and Luce (1998) may be limited to the task of auditory naming. Evidence from other tasks suggested that inhibitory neighborhood effects are evident in both words and nonwords. As discussed above in Section 2.2.1, Luce and Pisoni (1998) found significant inhibitory effects of neighborhood density for both words and nonwords in lexical decision. It is possible that tasks like auditory naming used only word stimuli; while for other task like an auditory lexical decision task, both nonwords and words were used as stimuli and in order to successfully identify a non-word, one has to reject all the neighbor words that gets activated when a non-word is heard.

A major caveat of Vitevitch and Luce's study is that neighborhood density and phonotactic probability covary in the experiment design, therefore it's not easy to differentiate the effects of neighborhood density and phonotactic probability. There has been some effort to tease apart the effects of neighborhood density and phonotactic probability, but the evidence comes from the study of speech production. Storkel, Armbrüster and Hogan (2006) conducted a picture naming experiment with 16 CVC novel words with English-speaking participants. The stimuli follow a 2-by-2 design of phonotactic probability (high, low) × neighborhood density (high, low). Overall, as suggested by Vitevitch and Luce (1998), the results showed that these two measures do have opposite effects on word processing, in spite of the high correlation between neighborhood density and phonotactic probability. Specifically, phonotactic probability has a inhibitory effect (i.e. stimuli with high phonotactic probability were less accurately named compared to the stimuli with low phonotactic probability), whereas neighborhood density showed facilitatory effects (i.e. stimuli with high neighborhood density were more accurately named than those with low neighborhood density). One might notice that these effects are in the opposite directions of the corresponding effects in word recognition. The notable contrast in effect direction has been attributed to differences between spoken word recognition and production (Peramunage, Blumstein, Myers, Goldrick, & Baese-Berk, 2011; Vitevitch, 2002; Vitevitch & Sommers, 2003). A production task is driven by semantics because of which sematic neighbors are strongly activated. While phonological neighbors will be co-activated but are weakly activated (Chen & Mirman, 2012) as they do not match the idea or semantics. According to Chen and Mirman (2012), weak neighbors lead to a facilitative effect of phonological neighbors on production task. Since the focus of this dissertation is on spoken word recognition, details on the aspects of speech production is beyond the scope of the current dissertation.

Apart from phonotactic probabilities, neighborhood density is also correlated with word length, at least for the English lexicon. Charles-Luce and Luce (1990) reported that neighborhood density decreases with increase in word length. Three-phoneme words have neighborhood density that goes till forty neighbors while the maximum neighborhood density for four- and five- phoneme words decreases to eighteen and twelve, respectively. Furthermore, 97% of three-phoneme words have more than five neighbors. Therefore, as the word length increases, neighborhood density decreases. As a result, majority of studies use three-phoneme monosyllables. Along with this, another reason to choose monosyllables as stimuli was probably to avoid the possible complication arising from lexical stress. (I will be talking more about suprasegmentals later in this Section.)

Despite the trend of focusing on shorter, monosyllabic words, there are a few studies that have investigated the effects of phonological neighborhood on longer words. Among others, Vitevitch, Stamer and Sereno (2008) investigated the effects of neighborhood density in bisyllablic English words. In this study, 56 bisyllablic words with strong-weak stress patterns were divided into dense (mean = 11.72; SD = 1.58) and sparse neighborhood (mean = 4.43; SD = 1.99). The stimuli were controlled for word frequency and neighborhood frequency. They conducted two experiments: (1) perceptual identification in noise, and (2) auditory lexical decision. In perceptual identification task, stimuli were presented at +12 dB signal-to-noise ratio. Thirty-seven participants participated in this task. Results from perceptual identification experiment showed that words from dense neighborhood were identified significantly less accurately (mean accuracy = 77.1%; SD= 7.5) compared to words from sparse neighborhood (mean accuracy = 80.3%; SD = 9.2). In auditory lexical decision experiment, same set of bisyllabic words were used along with 56 bisyllabic nonwords. Forty right-handed participants participated in this experiment. Results from lexical decision experiment showed that words with sparse neighborhood were responded faster and more accurately (mean reaction time = 833 ms; mean accuracy = 94.3%) than the words from dense neighborhood (mean reaction time = 846 ms; mean accuracy = 92.0%). Taken together, bisyllabic English words were identified more accurately and quickly when they belonged to sparse neighborhood than their counterparts i.e. words from dense neighborhood. These findings were in agreement with the previous findings (Luce & Pisoni, 1998). Although Vitevitch and Luce (1999) confirmed the effects of neighborhood density in bisyllablic English words, however in this study, stress was controlled by using words with strong-weak stress pattern only. Thus, the study avoided potential effects of stress pattern on the calculation of neighborhood matrix. Note that the definition of neighborhood does not say anything about stress, therefore it is unclear how to define the neighborhood for words with stress on one of the syllables (e.g. are SUBject and subJECT neighbors because they differ in stress pattern?) Overall, there is little discussion about the effects of suprasegmentals on the definition of neighborhood in the literature from English.

Besides the studies in English, studies have also been conducted in other languages including French (Dufour & Frauenfelder, 2010; Ziegler et al., 2003), Spanish (Vitevitch & Rodríguez, 2005), and Japanese (Amano & Kondo, 2000; Yoneyama, 2002). Ziegler et al. (2003) examined the effects of phonological neighborhood in French word recognition while controlling for word frequency, phonotactic probability, orthographic neighborhood density (i.e. the number of words that differ the target word by one letter). The stimuli words were 2-5 phonemes long (mean word length = 3.3 phonemes) monosyllables, and both auditory lexical decision, and auditory naming task were used. In order to control for orthographic neighborhood, the authors used phonological neighbors that differ by one-phoneme through substitution alone. Eighty monosyllabic words and eighty monosyllabic nonwords, balanced across phonological neighborhood density (mean high = 28.8; mean low = 13.5) were used as stimuli. LEXOP (Peereman, 1999) lexical database was used to obtain neighborhood measures. Overall, they found an inhibitory effect of neighborhood density on response time and accuracy ($F_1(1,31) = 206.9$, $p < 0.0001$; $F_2(1,74) = 8.01$, $p < 0.01$). In the naming task, forty-six participants were recruited. Same set of eighty real words monosyllables as used in auditory lexical decision task were used in this experiment. The results echoed the findings from their auditory lexical decision experiment. Phonological neighborhood showed an inhibitory effect on response time ($F_1(1,45) = 35.1$, $p < 0.0001$; $F_2(1,74) = 2.89$, $p < 0.10$). It can be observed that the effects for auditory naming was weaker as evident from the marginally significant main effects in the auditory naming task ($F_1(1,45) = 35.1$, $p < 0.0001$; $F_2(1,74) = 2.89$, $p < 0.10$) compared to the effects seen in auditory lexical decision task ($F_1(1,31) = 206.9$, $p < 0.0001$; $F_2(1,74) = 8.01$, $p < 0.01$). In order to balance phonological neighborhood and orthographical neighborhood, Ziegler et al.'s choice of stimuli was based on substitution neighbors alone in the calculation of phonological neighborhood. Therefore, words that differ by deletion or addition were not considered as phonological neighbors.

Later, Dufour and Frauenfelder (2010) examined the effects of phonological neighborhood using one-phoneme difference rule that includes phonological neighbors by substitution, addition and deletion. They found similar effects of neighborhood density and neighbor frequency in French spoken word recognition using a lexical decision task. Thirty-six (phoneme length ranging from three-to-four phonemes) monosyllabic French words, with high-density words (mean = 13.08) and low-density words (mean = 4.17); and thirty-six nonwords were selected as stimuli. Brulex database was used to calculate neighborhood measures. As observed in English, neighborhood density showed a significant inhibitory effect ($F_1(1,22) = 26.12$, $p < 0.0001$; $F_2(1,32) = 4.33$, $p < 0.05$) on reaction time in French monosyllables. Words from the sparse neighborhood were responded faster and more accurately (mean RT = 759 ms; error rate = 2.90%) compared to the words from dense neighborhood (mean RT = 826 ms; error rate = 5.98%). When words from low density neighborhoods were varied in neighborhood frequency, significant inhibitory effects of neighborhood frequency were observed on reaction time ($F_1(1,25) = 113.30$, $p < 0.0001$; $F_2(1,17) = 6.18$, $p < 0.05$). Words with no high frequency neighbor were responded more quickly and more accurately (mean RT = 817 ms; error rate = 2.31%) compared to words with high frequency neighbors (mean RT = 892 ms; error rate = 5.00%). It can be noted here that the effects of neighborhood frequency were more pronounced for sparse neighborhood. Overall, the effects of phonological neighborhood were similar to that seen in English, when focused on monosyllables.

Neighborhood effects were also investigated in Spanish, but the results were contradictory. Vitevitch and Rodriguez (2005) examined the effects of neighborhood density and neighborhood frequency in Spanish using an auditory lexical decision task. Eighty bisyllabic real words and eighty bisyllabic nonwords were used as stimuli in the experiment. The stimuli were categorized into high (log-transformed mean = 2.3) and low (log-transformed mean =1.3) word frequency (log-transformed), dense (mean = 15) and sparse (mean = 6.9) neighborhood, and high (log-transformed mean = 2.2) and low (log-transformed mean = 1.6) neighborhood frequency. Thirty-eight native Spanish speakers participated in this experiment. Results indicated that high frequency words were responded more quickly and accurately (mean reaction time = 932 ms; mean accuracy = 94%) than low frequency words (mean reaction time = 979 ms; mean accuracy = 87%); words from dense neighborhood were responded faster and more accurately (mean reaction time = 945 ms; mean accuracy = 93%) compared to words from sparse neighborhood (mean reaction time = 966 ms; mean accuracy = 88%); and words with high neighborhood frequency were responded more quickly and accurately (mean reaction time = 942 ms; mean accuracy = 92%) than words with low neighborhood frequency (mean reaction time = 968 ms; mean accuracy = 89%). These findings were in opposite direction compared to the earlier findings in English, in that both neighborhood density and neighborhood frequency showed facilitatory effects in Spanish word recognition (e.g. words with high neighborhood density were responded faster and more accurately than words with low neighborhood density). Similarly, words with high neighborhood frequency were responded faster and more accurately than words with low neighborhood frequency. The authors suggested that the opposite effects of phonological neighborhood as seen in Spanish could have been due to language-specific properties of the lexicon. For example, Spanish words tend to be longer than English words. Weiner and Miller (1946) reported that longer words were identified more accurately compared to shorter words. This could be one of the reasons why Spanish language showed opposite effects. In addition, Spanish words are more heavily inflected than English words, with the use of affixes indicating gender, number and tense. As a result, there are more words in Spanish that are both phonologically and semantically similar than in English. For example, the Spanish words *nino* and *nina* mean "male child" and "female child", respectively. The confluence of phonological similarity and semantic similarity in Spanish words may have caused the overall facilitatory effects observed in Spanish.

Although, Vitevitch and Rodriguez's (2005) results suggested cross-linguistic differences in the effects of phonological neighborhood. Previous studies, as discussed above, have mostly focused on European languages like English (Luce & Pisoni, 1998; Vitevitch & Luce, 1999), French (Dufour & Frauenfelder, 2010; Ziegler, Muneaux, & Grainger, 2003), and Spanish (Vitevitch & Rodríguez, 2005).

Therefore, there are relatively few studies that have examined the effects of phonological neighborhood in non-European languages except a few in Japanese (Amano & Kondo, 2000; Yoneyama, 2002). Yoneyama (2002) investigated the effects of phonological neighborhood via a series of word recognition experiments that included auditory naming, word identification in noise, and semantic categorization in Japanese. Seven-hundred trisyllabic CVCVCV Japanese words extracted from the NTT database (Amano & Kondo, 2003) were used as stimuli. Also, they evaluated the three different measures of computing phonological neighborhood: First, based on one-phoneme difference rule; Second, based on modified version of one-phoneme difference rule, where pitch accent was taken into consideration. Any two words that differ either by one phoneme or by pitch accent by addition, substitution or deletion were neighbors (e.g. *ana* and *a'na* are neighbors that differ in initial pitch accent); Third, based on auditory similarity between two words by mapping each word of the lexicon onto psychological mental space. This is done by mapping word onto time-frequency representation for each word in the lexicon. All three measures of computing neighborhood showed significant results. Interestingly, the results are mixed. On one hand, similar to English and French, results from word identification experiment showed inhibitory effects of neighborhood density on accuracy scores. On the other hand, similar to Spanish, results from auditory naming, and word identification experiment showed facilitatory effect of neighborhood density on reaction time. Interestingly, for semantic categorization task, the two types of neighborhood density effects coexisted. Neighborhood density based on auditory similarity rule showed inhibitory effects while neighborhood density based on one phoneme difference rule and one-phoneme/pitch-accent rule showed facilitative effects on reaction time. In a semantic categorization task, the authors reported that the two types of effects (inhibitory for auditory similarity rule and facilitative for one-phoneme difference and one-phoneme/pitch accent difference rule) on neighborhood density coexist. The authors speculated that the two types of effects (inhibitory and facilitative) on semantic categorization can coexist based on previous literature on phonotactic probability (Vitevitch et al., 1997). The facilitative effects can be interpreted as effects of phonotactic probability while the inhibitory effects are indicative of neighborhood density effects

As earlier mentioned, studies in European languages did not take suprasegmentals into account. Since Japanese is a pitch-accent language, contribution of suprasegmentals is inevitable. Therefore, as discussed above, Yoneyama (2002) proposed a rule that incorporates pitch-accent to define neighborhood measures. Moreover 70% of the world's languages are tone languages (Yip, 2002), therefore, use of suprasegmentals is of key importance to the majority of human languages. Thus, examining the role of suprasegmentals in phonological neighborhood becomes utmost necessary.

Majority of research has focused on spoken word recognition in English or other European languages. Thus, the exact patterns of phonological neighborhood effects in tone languages is still unclear. A tone language makes use of pitch patterns (i.e. "lexical tones") to differentiate word meanings. Therefore, the existence of lexical tones in tone languages gives rise to an important question on how to define phonological neighborhood in tone languages? Whether or not lexical tones should be included in the calculation of phonological neighbors? Previous studies have shown that tones and segments are equally important for Chinese spoken word recognition (Malins & Joanisse, 2012; McBride-Chang et al., 2008). Malins and Joanisse (2010) who studied the spoken word recognition in Mandarin using eye tracking found that tones and segments are processed simultaneously. Their findings suggest that tone plays an equally important role as segments and cannot be ignored. McBride-Chang et al. (2008) compared word recognition in English and Chinese for Cantonese-speaking children learning English. They found that in recognition of Chinese characters, lexical tones played a significant role, while in English, only phonemes were found to be essential. Cutler and Chen (1995) studied phonological similarity effects in Cantonese for spoken word recognition lexical decision task. They found that lexical tones and segments were processed similarly. Lee and Nusbaum (1993) studied interactions in processing of segmental and suprasegmental information in English and Mandarin. English does not have meaning attached to the suprasegmentals as compared to Mandarin and thus the processing takes place in a different manner for the two languages. The authors suggested that a comprehensive theory is needed to account for segmental and suprasegmental information processing of the acoustic stimulus in comprehending spoken language. Zhao, Guo, Zhou and Shu (2011) studied the time course of processing of spoken words in Chinese with monosyllabic words using N400. They found that Chinese was more sensitive to complete syllable in comparison to segments. Also, rime and tone both were found to be equally important. They also reported time differences in processing English and Chinese that could be due to structural differences in the two languages. Based on this evidence, it can be observed that lexical tone plays a quintessential role in spoken word recognition and thus, cannot be ignored.

Furthermore, there is a dearth of literature on phonological neighborhood in tone languages. Among the few studies on phonological neighborhood in tone languages that have used one-phoneme difference rule, either tone has been suggested to be processed similar to a phoneme (Kirby & Yu, 2007) or has completely been ignored with the focus mostly on segments (Tsai, 2007). Kirby and Yu (2007) used one-phoneme difference rule to define phonological neighborhood in Cantonese. They conducted a word likeliness judgement task where participants judge how likely the stimuli are a word on a 7-point scale where "1" indicated highly unlikely to be a Cantonese word and "7" indicates highly likely to be a Cantonese word. They found a positive correlation between neighborhood density and well-formedness (R^2 = 0.3277, F (1, 430) = 166, $p < 0.01$) for both words and non-words. In other words, words and non-words with many neighbors sounded more like real words to Cantonese speakers. It should be noted that in this study the authors completely ignored tone in defining phonological neighborhood in Cantonese.

As lexical tone modulates the meaning of words in tone languages, it is a very important aspect that needs further investigation to determine whether or not lexical tones should be considered, or it can be ignored while defining neighbors in spoken word recognition. Tsai (2007) studied the effect of neighborhood density in Mandarin monosyllables using an auditory naming task. Phonological neighborhood measures were obtained from the *Guoyu Cidian Jianbianben Bianjitzliau Tztspin Baugau* (Word Frequency Statistic Report of the Database for National Language Concise Lexicon) and Frequency Statistics of the Academia Sinica Balanced Corpus of Modern Chinese (Academia Sinica, 1997). In this study, phonological neighbors were defined as syllables that differ from target syllable in one of the three phoneme position (initial, nucleus and coda). This study calculated neighborhood density using one-phoneme difference rule that does not consider the contribution of lexical tone towards neighborhood. Ninety-one monosyllables were used as stimuli. The stimuli were divided into three-word sets based on type of neighbors and neighborhood density: Nasal-final neighbors, vowel-final neighbors, and large density difference neighbors. Nasal-final neighbors were words with nasal-final consonant neighbors e.g., /li/ and /lin/. Words were categorized into high (mean = 27.94) and low (mean = 25.94) neighborhood density. Vowel-final neighbors were words that did not have any nasal-final neighbors. Words were divided into high (mean = 28) and low (mean = 26.19) neighborhood density. In the large difference condition, words had greater difference between high (mean = 24.63) and low (mean = 16.75) neighborhood density. Twenty-six participants were recruited in this study. This study revealed an inhibitory effect of neighborhood density on reaction time (t (25) = 3, p<0.01). Monosyllables from dense neighborhood were responded slower compared to monosyllables from sparse neighborhood. This significant inhibitory effect of neighborhood density was seen only in condition with vowel-final neighbor. There are two major concerns regarding this study. First, the difference in high and low neighborhood density in vowel-final neighbor condition, that showed significant results, was relatively low (average high-low difference = 1.81). Surprisingly, no significant effect of neighborhood density was found for large density difference condition (average high-low difference = 7.88). If the effects of neighborhood density exist, it should be evident on large density difference condition. However, this was not the case. Second, the authors did not consider lexical tones in defining neighborhood.

Quite recently, Neergaard and colleagues (Neergaard, Xu, & Huang, 2016) have also worked in this line of research. In particular, Neergaard, Xu and Huang (2016) generated a database of Mandarin neighborhood statistics, using lexical information from SUBTLEX-CH corpus (Cai & Brysbaert, 2010). The database provides lexical statistics for words and nonwords. The database contains information on phonological neighborhood density, neighborhood frequency, homophone density, word frequency, and homophone density of words and neighborhood density and neighborhood frequency of nonwords in Mandarin. Along with lexical statistics, the database also provides description of words that includes pinyin, phoneme length, lexical tone, syllable structure, etc. This database used multiple measures to calculate neighborhood density based on different ways of segmenting a word. These segmentation schemes were based on modified one-phoneme difference rule where instead of phoneme difference, a unit difference is considered as a neighbor. This unit can be a phoneme or a group of phonemes. The segmentation schemes used were C_V_C, C_VVX, C_V_V_X, C_V_VX, CV_V_X, CV_VX, and CVVX where C stands for consonant, V for vowel, X for second vowel of diphthong or glide and the underscore defines the unit of segmentation. For example, in C_V_C scheme, word that differs in consonant or vowel by addition substitution or deletion are neighbors while in C_VVX scheme, words that differ either by consonant or rhyme by addition substitution or deletion are neighbors. The segmentation schemes are broadly classified as one that considered lexical tone in the calculation of phonological neighborhood (7 schemes) and others that did not consider lexical tone in in the calculation of phonological neighborhood (7 schemes). For example, C_V_C scheme considers neighbors that differ in consonant or vowel by addition substitution or deletion while C_V_C_T scheme consider neighbors that differ in consonant or vowel or tone by addition, substitution or deletion.

Later, Neergaard (2018) examined the effects of phonological neighborhood in Mandarin by 16 different definitions of phonological neighborhood (using database reported in Neergaard, Xu and Huang (2016) in an auditory naming and an auditory lexical decision task. However, there are multiple issues about the study that undermine the reliability of the results. The first major issue is with experimental design. The auditory naming task used one-hundred fifty-four mono- and bisyllablic words, and auditory lexical decision task used seventy-five monosyllables as stimuli, but it is unclear how these stimuli were selected and what factors were considered or controlled. Possible differences in the stimuli (e.g. word frequency, stimulus duration, etc.) that might affect naming performance were not addressed in either stimuli selection or data analysis (see below).

The second issue has to do with data analysis. To test the effects of different neighborhood measures, Neergaard constructed 144 models (9 predictors X 16 definitions of neighborhood), presumably with only one fixed-effect predictor (e.g. neighborhood density, or word frequency) in each model. It is unclear why it was necessary to build separate models for the same non-neighborhood-related predictor (e.g. word frequency), as the predictor does not change with the neighborhood definition. A more serious problem is the practice of modeling different predictors separately, which fails to control for potential interaction and interference. Furthermore, some critical control variables (such as stimulus duration, length of word in syllables) that are known to affect perceptual recognition were completely missing from their models. In addition, in auditory lexical decision task, 11.82% of the data were excluded from analysis without a credible explanation. Furthermore, he results were analyzed on reaction time alone and no analysis was conducted for accuracy scores while it has been shown in multiple previous studies that accuracy scores are sensitive to neighborhood effects (Luce & Pisoni, 1998). Neergaard reported that the best-fitted model was the one that considered both tone and segments in defining phonological neighbors in Mandarin, and that the neighborhood effect was facilitatory. However, given the serious concerns about the methodology, the reliability of these results is called into question.

Overall, it is quite evident that there is relatively small number of studies that have examined the effects of phonological neighborhood in Mandarin. Moreover, it can be elucidated from the review of these studies that there are mixed findings on the effects of phonological neighborhood in Mandarin. This dissertation would focus on portraying a complete picture of the effects of phonological neighborhood on spoken word recognition using a complete set of Mandarin monosyllables in Mandarin. An exhaustive set of statistical analyses controlling for factors that affects spoken word recognition like word frequency, stimulus duration etc. will be conducted in order to avoid any errors while interpreting critical effects.

2.2.3 Summary

In the above Section, I have reviewed the literature on effects of phonological neighborhood across languages. As described above, most of the studies conducted in this area converge to a common finding that dense phonological neighborhood and higher neighborhood frequency lead to inhibitory effects in spoken word recognition.

However, it should be noted that there is a great amount of variability associated with the effects of phonological neighborhood. First of all, there is a variability in the effect of phonological neighborhood across languages. Most of the studies examining the effects of phonological neighborhood have been conducted in English or other European languages (Dufour & Frauenfelder, 2010; Luce & Pisoni, 1998; Vitevitch & Rodríguez, 2005; Ziegler et al., 2003). Vitevitch et al. (2004) suggests that the effects of phonological neighborhood vary across languages. These differences in findings from studies conducted in different languages could be attributed to language-specific properties of the lexicon. Therefore, the manner in which words are processed in different languages may depend on specific features of that language. Majority of previous research was focused on English or other non-tone languages like French and Spanish. However, majority of the languages in the world are tone languages. Earlier studies in European languages did not take suprasegmentals into account. The most widely accepted definition of neighborhood based on one-phoneme difference rule (Luce & Pisoni, 1998) do not consider suprasegmentals in the calculation of neighborhood. However, suprasegmental features like pitch, stress and intonation are of high importance that cannot be avoided or ignored as 70% of the world languages are dominated by suprasegmental features.

Not just across languages, the effects of phonological neighborhood tend to vary across tasks. Phonological neighborhood influence word processing either by posing competition to the target word or by helping the target word depending on the type of the task involved. In speech perception, the effects of phonological neighborhood on spoken word recognition are predominantly inhibitory (Luce & Pisoni, 1998;

Vitevitch & Luce, 1999). Words from dense neighborhood are recognized less accurately and take longer due to more competition compared to words from sparse neighborhood. In comparison, during word production, the presence of neighbor aids in processing by passing on their activation to the target word, in other words, facilitating the process of word production (Peramunage et al., 2011; Vitevitch, 2002; Vitevitch, Armbrüster, & Chu, 2004; Vitevitch & Sommers, 2003). Words with high neighborhood density show facilitatory effect compared to words with low neighborhood density. In sum, words in a dense neighborhood are difficult to perceive but easier to produce.

Apart from cross-linguistic and across-task differences, another important issue is that some languages are richer than others in the number of homophones (i.e. similar sounding words with different meaning). However, most of the investigated languages are European languages that do not have enough homophones. In comparison, Mandarin, a tone language is known to have abundant homophones. Therefore, it is important to not to ignore the effects of homophones on spoken word recognition. In the next Section, I will be reviewing literature on the effects of homophones on spoken word recognition.

2.3 Homophones

As discussed above in the neighborhood model, phonological neighbors are similar sounding words, and they affect each other in lexical processing. An additional issue is whether words that sound the same, i.e. homophones, are phonological neighbors as well. Homophones are words with same phonological form but with different meaning and spelling, e.g. *new* and *knew* are homophones in English. This issue is interesting because on one hand, homophones are similar to phonological neighbors in terms of phonological similarity, given that homophones are phonologically same words. On the other hand, homophones are different from phonological neighbors as homophonous words have the exact same phonological form while phonological neighbors are similar sounding words with different phonological form. This issue has not been fully explored mainly because of paucity of homophones in the languages that were studied. In this Section, I will first discuss the representation of homophones in the mental lexicon, and then review previous literature on the effects of homophones.

Based on previous studies, there are two types of model of homophones. The first type of model of homophones is the shared lexical representations of homophones (Dell, 1990; Jescheniak & Levelt, 1994) where homophonous words share a common phonological form. Henceforth, I will be calling this type of homophone representation in the lexicon as *shared representation*. All the models reviewed in Section 2.1 Models of spoken word recognition (Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994) represent shared representation. Under shared representation, homophone mates have a common phonological form (at lexeme level) but separate representations at the lemma level (see Figure 1). This is in contrast to phonological neighbors, which have separate representations at both the lemma level and the lexeme level. For example, as shown in Figure 1, *bear* and *bare* are homophones, and they have separate representation at lemma level but common representation at lexeme level. On the other hand, *beer* and *hear* are phonological neighbors, which have separate representation at both lexeme level and lemma level.

In this model, homophone relationship and phonological neighborhood are represented differently, therefore the model allows for the possibility that homophones may have different effects than phonological neighbors on the target word.

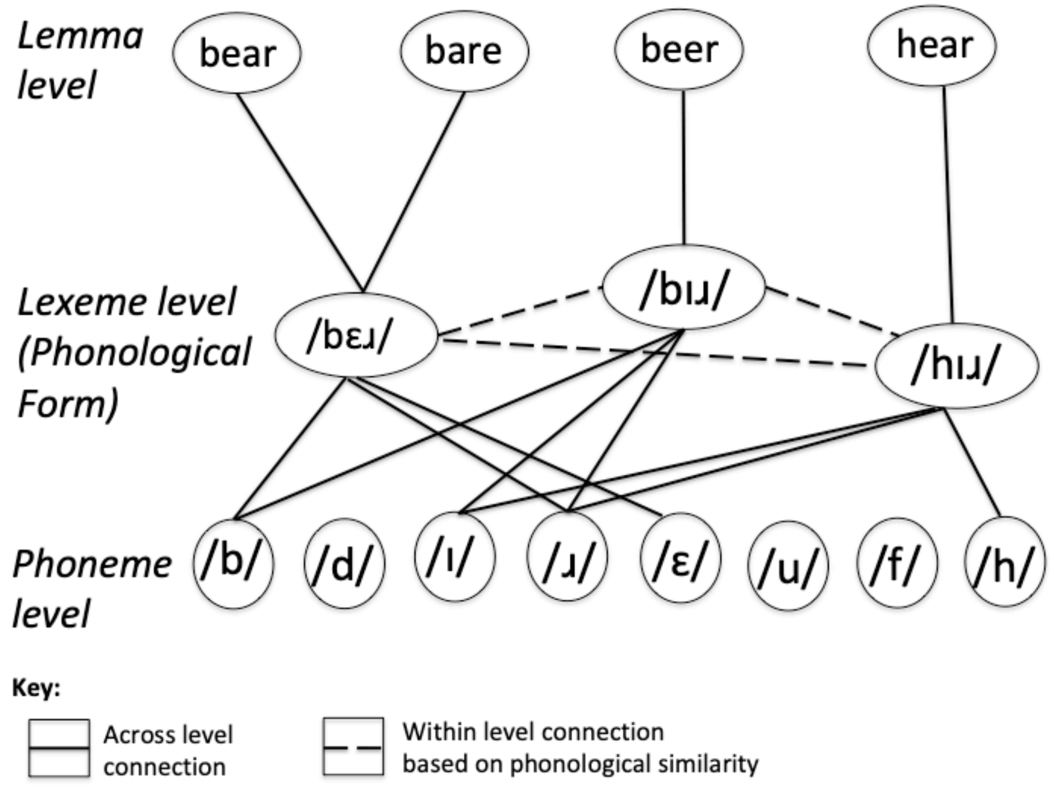

Figure 1: The shared lexical representation of homophonous words where solid line represents the connection between levels, and dotted line represents connections within lexeme level among phonologically similar words. The general structure of the model follows NAM (Luce & Pisoni, 1998).

The second type of representation of homophonic relationship assumes independent lexical representations of homophones (Caramazza, Costa, Miozzo, & Bi, 2001). In this model, each member of the homophone family is associated with a unique phonological form (see Figure 2). Henceforth, I will be calling this type of homophone representation as independent representation. Under independent representation, homophone mates have separate representations at both the lexeme and the lemma level. This makes the representation of homophone mates similar to that of phonological neighbors, which also have separate representations at both the lemma level and the lexeme level. As shown in Figure 2, *bear* and *bare* are homophone mates, and they have separate representation (different from Figure 1) at both lexeme and lemma levels, just like phonological neighbors *beer* and *hear*. The words *bear* and *bare* have the same phonemic sequence, but their phonetic realizations may not be exactly the same, due to differences in duration, vowel targets, or other aspects of phonetic detail (e.g. Gahl (2008); see the discussion below). The fact that they are phonetically different makes them a type of phonological neighbors, in the broad sense as words that sound similar to each other. Furthermore, compared to the canonical phonological neighbors, which have different phonemic sequences, homophone mates enjoy higher similarity in pronunciation because their differences are sub-phonemic. In other words, in the independent representation model, homophone mates should produce similar effects on lexical processing as canonical phonological neighbors, and to a greater extent than canonical phonological neighbors.

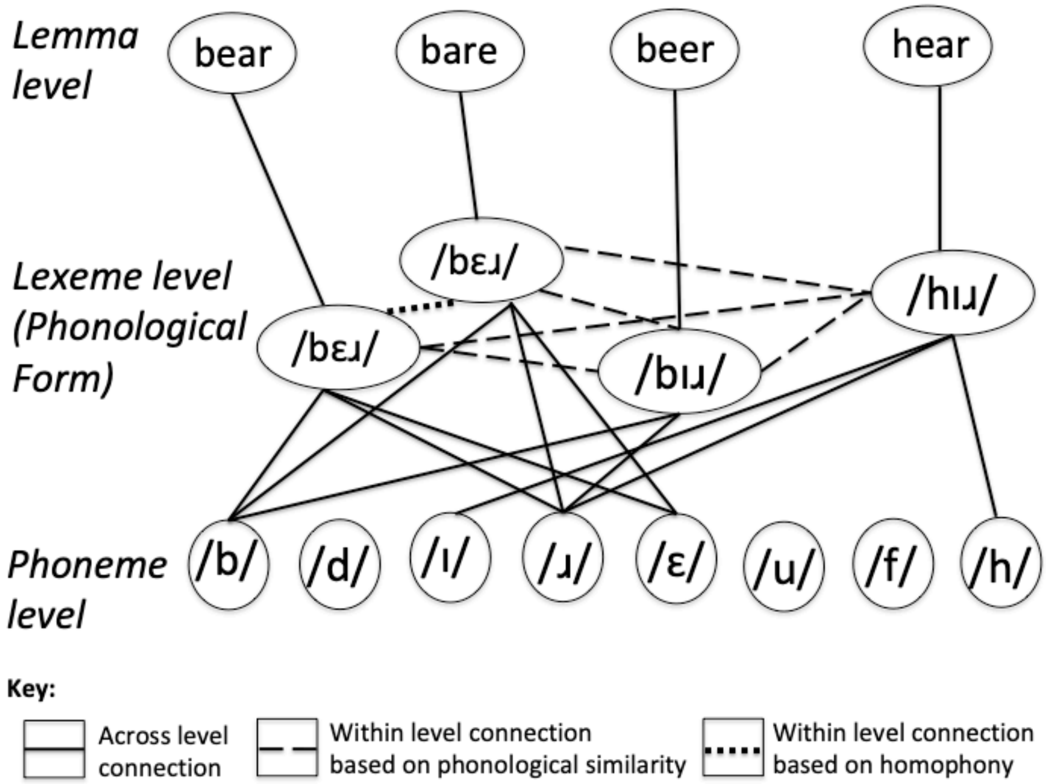

Figure 2: The independent lexical representation of homophonous words where solid line represents the connection between levels and dotted line represents connections within lexeme level. The general structure of the model follows NAM (Luce & Pisoni, 1998).

As discussed above, the shared representation model assumes that homophone mates, unlike phonological neighbors, share lexical representation at lexeme level, and hence allows the possibility for homophones to have different effects than phonological neighbors; in contrast, the independent representation model assumes that homophone mates have separate lexical representation at lexeme and lemma level, just like phonological neighbors, and thus predicts that homophones have similar effects as phonological neighbors. There is evidence for both types of homophone representations in the lexicon. The evidence for shared representation of homophones comes from frequency inheritance, which is the phenomena of low frequency homophone mates behaving similar to high frequency homophone mates.

Jescheniak and Levelt (1994) found that low frequency homophone mates have similar naming latency as high frequency homophone mates in a production task. They examined the naming latency of homophonous words in a translation task. The participants were presented with English words visually, and were asked to produce their Dutch translations. In this study, low frequency words with high frequency homophones were used as experimental condition, and low frequency words without homophones and high frequency words without homophones as control condition. The stimuli constituted of 11 items in experimental condition items and 33 items as fillers for control condition. Dutch being very transparent language has homophones with same spelling but different meaning, e.g. *bos* in Dutch means "*forest"* and "*bunch".* The authors reported that low frequency words with high frequency homophones (mean RT = 796 ms) were produced faster than low frequency control words (mean RT = 888 ms) but comparable to high frequency control words (mean RT = 765 ms). In other words, these findings suggest that the naming latencies cannot be explained by lemma frequency. Therefore, they can only be explained by lexeme frequency. These findings would be possible only if low frequency homophone mates can inherit the frequency of their high frequency homophone mates. In the two models of homophone representation discussed above, the shared model makes it possible for these low frequency homophone mates to share the frequency of their high frequency mates. In fact, in the shared model, all the homophone mates share representation at the lexeme level, their lexeme would have the cumulative frequency of all the homophone mates. As a result, the naming latencies of high frequency homophone and low frequency homophone of the word with the same phonological form would not differ because the cumulative frequency of the common phonological form remain the same. Therefore, these results provide support for a shared representation at the lexeme level of homophone mates.

As for the evidence for the independent representation model of homophones, Gahl (2008) showed that homophone mates are indeed phonetically different. According to Gahl (2008), there are word duration differences among homophone mates. The author analyzed the word durations of around 90,000 English homophones from the Switchboard corpus. The author found that high frequency homophonous words were significantly shorter in duration compared to its low frequency counterparts, after controlling for speaking rate, syntactic category, bigram probability, proximity to pauses, and orthography. This difference in word durations of high frequency homophonous words and low frequency homophonous words is suggestive of separate representation of homophones in the mental lexicon, providing support for the independent representation model of homophones.

More importantly, the main evidence for the independent representation model of homophones comes from the absence of frequency inheritance. Caramazza et al. (2001) replicated the experiment of Jescheniak and Levelt (1994) and found opposite results. Caramazza and colleagues conducted a series of experiments with picture naming and translation tasks for English, Chinese and Spanish. In the English picture naming task, low frequency words with high frequency homophones (N=26) were used as critical words, compared to control words that were either low frequency words without homophones (N=26) or high frequency words without homophones (N=26). The authors reported that the naming latency of critical words (mean RT = 764 ms) was much longer than that of high frequency control words (mean RT = 714 ms, $p < 0.005$) in both by-subject and by-item analyses. This finding suggests a lack of frequency inheritance. If there were frequency inheritance, then we expect to see critical words to inherit the cumulative frequency of the homophone family which is comparable to the high frequency control words. Therefore, we would expect these critical words to have similarly short naming latency as high frequency control words. In fact, the naming latency of critical words is closer to the low frequency control words than the high frequency control words. These results suggest a lack of frequency inheritance. In the two models of homophone representation discussed earlier, the shared representation model would predict frequency inheritance while the independent representation model does not. Thus, these results provide evidence for the independent representation model of homophones but not for the shared representation model. The support also comes from similar experiments with Mandarin and Spanish. The picture naming experiment with Mandarin also found that that critical words ($N = 32$) had significantly longer naming latencies (mean RT = 783 ms) than high frequency control words ($N= 32$, mean RT = 717 ms, $p < 0.002$), in both by-item and by-subject analyses. This again provides evidence of lack of frequency inheritance. The third experiment was a translation task from Spanish-to-English, similar to the English-to-Dutch translation task used in Jescheniak and Levelt (1994). The results showed that the translation latency of low frequency words with high frequency homophones was significantly longer than that of (mean RT = 1058 ms) high frequency control words (mean RT = 852 ms, $p < 0.001$).

However, what Caramazza and colleagues did not discuss in these studies is the possible difference between critical words (low frequency words with high frequency homophone mates) and low frequency control words (low frequency words with no homophone mates). In the English picture naming task, the average naming latency of critical words (mean RT = 764 ms) was longer than that of the low frequency control words (mean RT = 752 ms), but only significant in the by-subject analysis ($p < 0.004$) and not in the by-item analysis (Fs < 1). Similarly, in the Mandarin picture naming task, the critical words (mean RT = 783 ms) had longer naming latencies than low frequency control words (N=32, mean RT = 749 ms), but the difference was only significant in the by-subject analysis ($p < 0.001$), but not in the by-item analysis ($p = 0.10$). In the Spanish task, however, there was no difference in naming latency between the critical words and low frequency control words (mean RT for critical words = 1058 ms; mean RT for low frequency control words = 1060 ms; Fs < 1). Overall, it is fair to say that there also seems to be a slight trend for the critical words to have longer naming latency than low frequency control words, although the trend is not always significant. How to interpret such results? They seem to suggest that there may be some competition from high frequency homophone mates in the processing of the critical words, which is not present in the processing of low frequency control words. Neither model of homophone representation directly predicts the inhibition among homophone mates in production tasks, however, it is also true that neither model rules out inhibition among homophone mates. In the shared representation model, interaction among homophone mates may arise through shared phonological forms (see Figure 1). Similarly, in the independent representation model, interaction may arise through shared phonemes (see Figure 2). Neither model specifically predicts that the interaction should be inhibitory (i.e. competition). In fact, at least in the independent representation model, one may argue that the interaction among homophone mates may be facilitative in speech production tasks such as picture naming, because homophone mates have the same representations as phonological neighbors, which according to Vitevitch and colleagues (Vitevitch & Luce, 1999) facilitate each other in speech production (however, see Sadat, Martin, Costa and Alario (2014) for different opinions on this). This raises a question of whether there are interactions among homophone mates during lexical processing? If so, are the interactions facilitative or inhibitory? On a related note, are the interactions among homophone mates similar to the interactions among phonological neighbors? These are all questions that I will address in the current thesis. In particular, I will be using speech perception tasks, for which the effects of phonological neighbors are better understood.

Initially the difference between the results from Caramazza et al. (2001) and Jescheniak and Levelt (1994) were attributed to cross linguistic differences between Dutch, and English and Chinese. Dutch has homophones with identical spellings (but different meanings) while the homophones in English and Chinese have different orthographic forms. For example, *bos "bunch" and bos "forest"* are homophone mates in Dutch, whereas *nun* and *none* are homophones in English. As a result, homophones with the same spelling may facilitate frequency inheritance, leading to the observation of frequency inheritance for Dutch but not for other languages. To test this hypothesis, Caramazza et al. (2001) compared same-spelling homophones with different-spelling homophones in English. If orthography played a role then same spelling homophones were expected to be named faster compared to different spelling homophones. However, the authors found no significant difference in naming latency between the two groups of words, failing to support the hypothesis of orthographic influence. Caramazza and colleagues also suggested that the discrepancy between the studies could be due to the size of the stimuli set. Jescheniak and Levelt (1994) study had only 11 items per group, whereas Caramazza et al.'s studies had 22-32 items per group. However, it is unclear whether a size difference of this scale could result in opposite effects of frequency inheritance since this has not been tested with larger stimuli sets.

As mentioned above, there are two hypotheses regarding homophone representations: (1) assuming the shared representation model of homophones (Jescheniak & Levelt, 1994), and (2) assuming independent representation model of homophones (Caramazza et al., 2001). They both have some support from empirical research from production studies: evidence for shared representation comes from frequency inheritance in Dutch (Jescheniak & Levelt, 1994); evidence for independent representation comes from frequency non-inheritance in English, Mandarin and Spanish (Caramazza et al., 2001), and pronunciation variation among homophone mates in English (Gahl, 2008). Overall, there is more evidence for independent representation of homophones based on multiple evidence provided by Caramazza et al. (2001) in three different languages. However, neither of the models on homophone representation specifically addresses the nature of interaction among homophone mates, even though interaction among homophone mates is suggested in some of these studies (Caramazza et al., 2001). In addition, it should also be noted that the stimuli set used in previous behavioral experiments were in general quite small (Caramazza et al., 2001; Jescheniak & Levelt, 1994). This is partly due to the fact that most of the examined languages, e.g. English and other European languages, do not have many homophones (even when Caramazza et al. (2001) used Mandarin, which has a high density of homophones (see Section 2.4 Background on

Mandarin), they only used 32 homophone pairs in the stimuli set).

In this thesis, I am going to focus on Mandarin and include a much larger set of homophones in the experimental stimuli. Mandarin provides an ideal testing ground for homophone effects, because of the high density of homophones in the lexicon. It is estimated that Mandarin has an average homophone density of 5 characters per syllable (Duanmu, 2007) i.e. each syllable in Mandarin has 5 homophones on an average. A monosyllable with same phonological form and same lexical tone can be represented by up to 40 characters in Mandarin. For example, the syllable /i/ with a Tone 4 (falling) can have up to 39 characters. This makes Mandarin a suitable tool to test the effect of homophones (more details on Mandarin phonology in Section 2.4). In the following, I will review a few studies that investigate the homophone effects in Mandarin (Fang, Li, & Luo, 2014; Li, Fang, & Lou, 2011; Li, Wang, & Li, 2011; Wang, Li, Ning, & Zhang, 2012; Zhou, 2015). Also, these studies are mostly focused on word recognition, which is less investigated than word production in the homophone research. As reviewed in Section 2.2.2 Phonological neighborhood effects, canonical phonological neighbors inhibit spoken word recognition. Therefore, based on the assumptions of the two models, the shared representation model would predict that the effects of homophone mates are different from phonological neighbors on spoken word recognition, and the independent representation model would predict inhibitory effects of homophone mates, probably even stronger than phonological neighbors on spoken word recognition. In general, the studies on Chinese homophone effects in word recognition found evidence for independent representation model and more importantly, interactions between homophone effects and word frequency.

Wang et al. (2012) investigated the effect of homophones using an auditory lexical decision task in Mandarin. Most commonly used measure of homophones is homophone density that refers to the number of words that share same phonological form with the target word. In this study, the authors used thirty monosyllabic words with high homophone density (more than 9 homophones, mean = 13.3), thirty with low homophone density (less than 9 homophones, mean = 4.9), and sixty non-words as stimuli. All lexical measures of stimulus words were obtained from Dictionary of frequently used words (Yuan, Ci, & Dian, 1990). In their study, they controlled for syllable frequency, frequency of the highest homophone in the homophone family, salience (ratio of highest homophone frequency to the second highest homophone frequency), and strokes frequency. All stimuli used in this study were high frequency monosyllables. The authors reported significant inhibitory effects of homophone density. Reaction times were longer and accuracy scores were poorer for words with high homophone density (mean reaction time = 836 ms; mean accuracy = 92.4%) than the words with low homophone density (mean reaction time = 794 ms; mean accuracy = 94.9%). Based on these findings, the authors suggested that words with more homophone mates pose stronger competition than words with lesser homophone mates in lexical processing. This evidence suggests that the nature of interaction among homophone mates is competition. It should be noted that the observed effects of homophone density in Wang et al.'s research strongly reminds us of the well-established inhibitory effects of phonological neighborhood density. As discussed in Section 2.2.2 (Phonological neighborhood effects), phonological neighbors have an inhibitory effect on spoken word recognition. This suggest that there is seemingly parallel similarity between homophone mates and phonological neighbors in spoken word recognition.

Later, Zhou (2015) conducted an eye-tracking experiment to investigate the effect of homophone density on word recognition while controlling for syllable frequency (cumulative frequency of the homophone family), and tonal probability. In this study, 32 monosyllables were used as critical items and 64 monosyllables as fillers. In each trail, participants were presented a slide with four characters. Participants were asked to identify the heard target word by clicking the character on the slide. Fixation data was collected continuously throughout the experiment. Two mixed effects models were used to analyze the data, one on reaction time in identification responses and the other on reaction time in fixation responses. Model results reveal no significant main effects of homophone density, syllable frequency, and tonal probability in either model. Although words with high homophone density (mean RT = 1231.88 ms) were longer than words with low homophone density (mean RT = 1089.33 ms). However, the model on reaction time in identification responses (but not the model on reaction time in fixation responses) shows a significant 3-way interaction among homophone density, syllable frequency, and tonal probability (β = 810.39, $t = 2.361$, $p < 0.05$). The direction of the interaction indicates that when all three variables are high, the reaction time becomes longer, although the authors did not go into detail of the patterns underlying the interaction effect. By contrast, Wang et al.'s study only used high frequency stimuli.

To summarize, some similarity between homophone mates (homophone density) and phonological neighbors (phonological neighborhood density) has been observed, mostly in spoken word recognition and less clear in speech production. To my best knowledge, the observed interactions among homophone mates are all inhibitory in nature. Furthermore, the research on Mandarin spoken word recognition suggested an interaction between frequency and homophone density, in that when frequency is higher, the inhibitory effect of homophone density is higher. But this research has overall focused more on the high frequency range of the lexicon. What is lacking is a comprehensive investigation of the full scale of the phenomenon. In this dissertation, I will fill this gap by examining the complete continua of frequency and homophone density in the Mandarin lexicon.

Before I go into the details of my study, I will first provide a brief description of the phonological system of Mandarin.

2.4 Background on Mandarin

Mandarin is a tone language spoken mainly in China. Monosyllables are considered as the building blocks in the Mandarin lexicon. In Mandarin, there are 24 onset including null onset. The structure of rime is a bit complex. Rime consists of a vowel or a diphthong with or without coda. There are 33 rimes in Mandarin. Onset and rime are also referred to as initial and final respectively, but in this dissertation, I am using onset and rime. Table 2 lists all the onset and rimes in Mandarin. Further, there are four lexical tones in Mandarin: *high-level* (Tone 1), *high-rising* (Tone 2), *dipping* (Tone 3), and *high-falling* (Tone 4). Tone is a lexical entity in Mandarin in the sense that a syllable in combination with different tones mean different. For example, syllable /ma/ when spoken with *high-level* tone (Tone 1) /ma1/ means ‘mother’ vs. *high-rising* tone (Tone 2) /ma2/ means ‘hemp’ vs. *dipping* tone (Tone 3) /ma3/

means ‘horse’ vs. *high-falling* tone (Tone 4) /ma4/ means ‘to scold’. Therefore, a legal monosyllable in Mandarin is a combination of onset, rime, and lexical tone. A Mandarin monosyllable can be represented as $CGVV^T$ or CGV^TC, where C is a consonant, G is a glide, VV is either a vowel or a diphthong, and T is a lexical tone (Duanmu, 2007).

Table 2: List of all the onset and rime (in IPA) in Mandarin.

Onset	Rime
p, pʰ, m, f, t, tʰ, n, l, ts, tsʰ, s, tʂ, tʂʰ, ʂ, ʐ, tɕ, tɕʰ, ɕ, k, kʰ, x, w, y, 0 (null)	a, ɔ, ɣ, i, ɚ, ai, ei, ɑʊ, oʊ, an, ən, ɑŋ, əŋ, ʊŋ, ja, jɑʊ, jɛ, joʊ, jɛn, in, jɑŋ, iŋ, jʊŋ, u, wa, wɔ, waI, weI, ɥɛn, yn, wɑŋ, v, ɥœ

Due to phonotactic constraints, not all the combinations of onset and rime are possible. In addition, there are also accidental gaps in the combination of onset, rime, and tone. For the purpose of this study, we divide gaps in the syllable inventory into two types: segmental gaps and tonal gaps. Segmental gaps are impossible combinations of onset and rime. For example, the combination of onset sound /ʂ/ and rime /joʊ/ is not possible in Mandarin for any of the lexical tones. Therefore, monosyllables /ʂjoʊ1/, /ʂjoʊ2/, /ʂjoʊ3/, and /ʂjoʊ4/ do not exist in Mandarin. Tonal gaps, on the other hand, are impossible combinations of onset, rime, and tone, where the combination of onset and rime may occur with other tones. For example, monosyllables /an1/ /an3/, and /an4/ exist in Mandarin, but /an2/ does not, which makes /an2/ a tonal gap. Given the existence of these gaps, Mandarin has an inventory of around 1300 (tonal) monosyllables (Duanmu, 2007).

In Mandarin morphology, most of the words consist of one or two monosyllables, where each syllable representing a morpheme. Due to the small size of syllable inventory, each monosyllable may correspond to multiple morphemes, where each morpheme has a unique meaning and a unique orthographic form (i.e. character). In other words, each monosyllable can be written with a number of different characters representing different meanings. As a result, there is a high density of homophones in Mandarin, to the extent that a monosyllable can correspond to up to 40 morphemes. On average, each monosyllable corresponds to a homophone family with 5 members (Duanmu, 2007), and according to my calculation, 77.2% of the monosyllables have at least 2 homophone mates in the family. For example, the monosyllable /kaɪ4/ correspond to 5 characters 丐 "beggar", 概 "general", 溉 "irrigation", 盖 "cover", and 钙 "calcium".

It is quite evident that the structure of Mandarin is quite different from English and other European languages. All the spoken word recognition models reviewed earlier in Section 2.1 (Models of spoken word recognition) hold good for English and other European languages. To understand the underlying mechanism of spoken word recognition in Mandarin, Zhou and Marslen-Wilson (1994, 2009) proposed a multi-level model for word recognition in Mandarin. The model postulates an existence of highly interconnected three levels of processing towards word recognition: syllable level, morpheme level, and word level. Syllable level represents the phonological forms while morpheme and word levels represent morphemes and word forms, respectively. There exist connections both between and within these levels of processing. Between the levels, the processing takes place in a hierarchical manner from syllable to morpheme to word while within each level, there exist weaker links between forms that share some features. Morphemes and words that has same first syllables are connected within level. Although there is no direct connection between words that share same morpheme at word level, but they are indirectly connected through connection within morpheme level. During the process of word recognition, there is a spread of activation along the connections at different levels and within levels that influence the word recognition process in Mandarin. Figure 3 shows a pictorial representation of the lexicon adapted from Zhou and Marslen-Wilson (1994) model. In Figure 3, connections within syllable level indicate phonological similarity.

As discussed earlier in Section 2.3 (Homophones), there are two types of lexical representation of homophones: (1) Shared representation, and (2) Independent representation. Given Zhou and Marslen-Wilson (1994, 2009) multi-level model for Mandarin, Figure 4 and Figure 5 depict the revised models for the two types of representations specifically for Mandarin. Homophones are words that share the phoneme sequence and tone but are not exactly the same, based on the assumption of independent model. As shown in Figures 4 and 5, the key difference between the two models is whether homophones have shared representation (as seen in Figure 4) or independent representation (as seen in Figure 5) at syllable level. Since the stimuli used in this dissertation are monosyllabic morphemes so I will only be focusing on the interactions among the units at or below morpheme level.

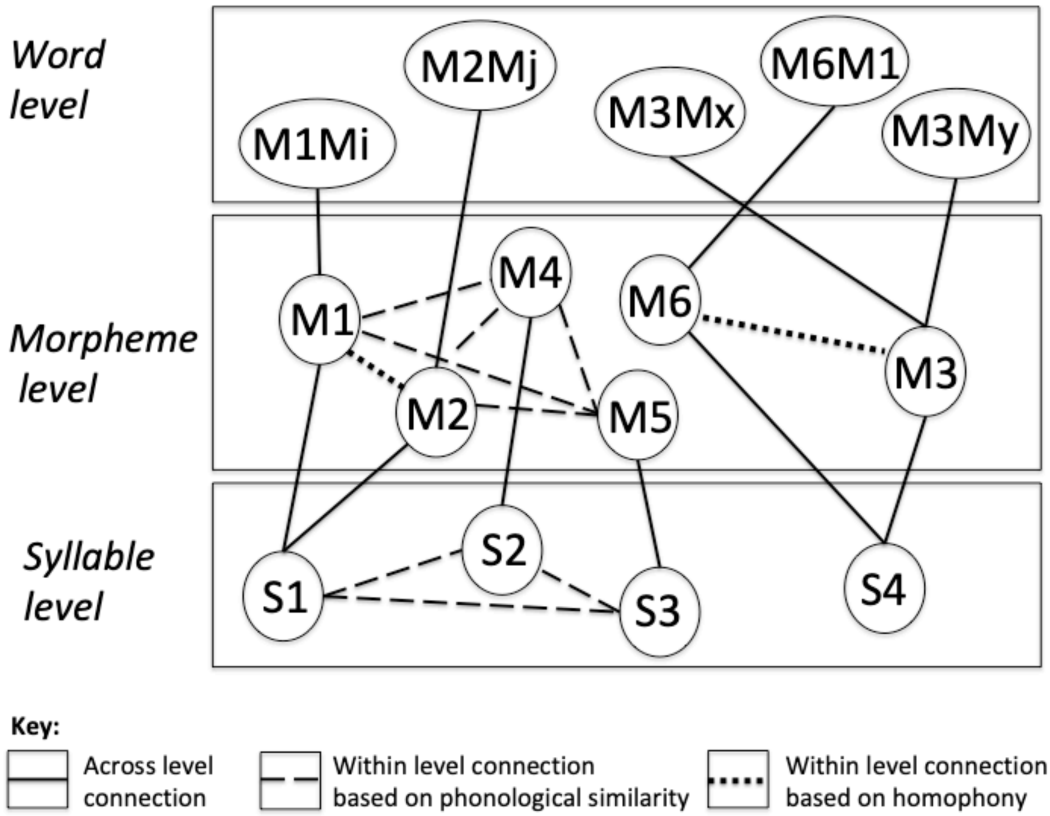

Figure 3: Lexical representation model adapted from Zhou and Marslen-Wilson (1994) multi-level model for word recognition in Mandarin. Here solid line indicates connections between levels, dashed line indicates connections within level, S stands for syllable, and M stands for morpheme. Homophones are connected through dotted lines, while phonological neighbors are connected through dashed lines. Please note that in the original model there are also connections at the word level among words that share the first syllable. However, this is not the focus of the current study, thus, these connections are not included.

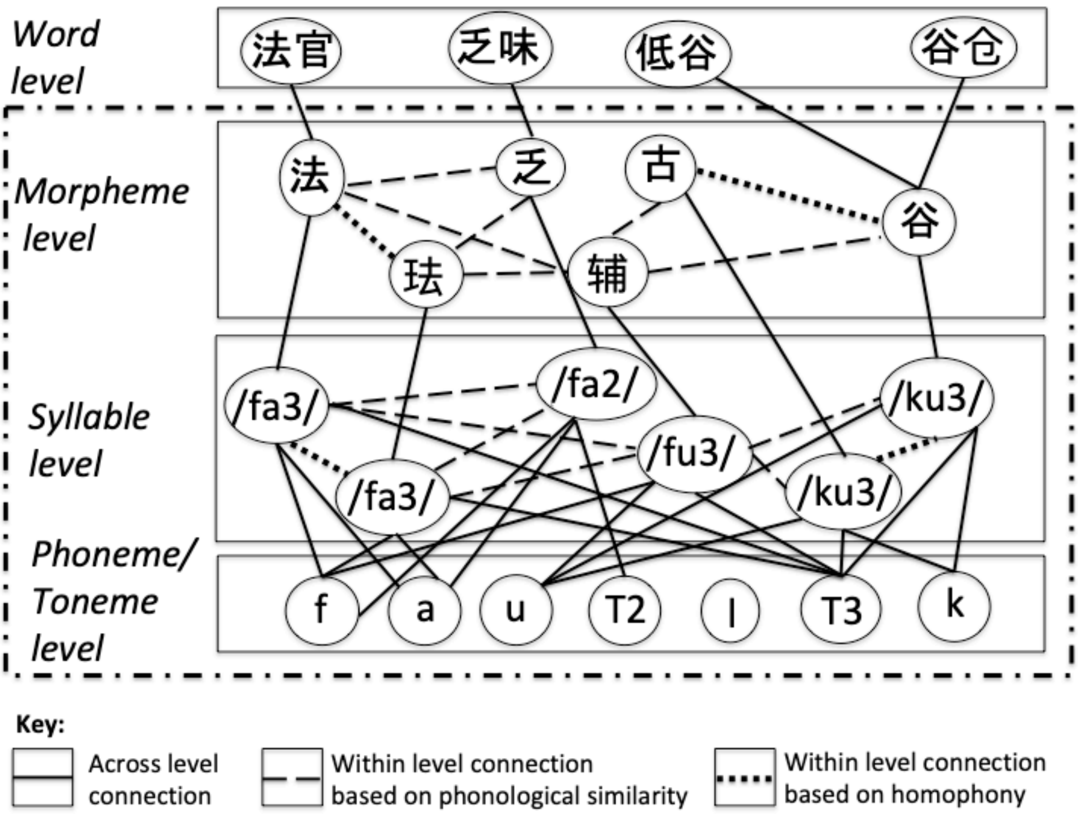

Figure 4: Adapted shared representation model for Mandarin. Here solid line indicates connections between levels. Homophones are connected through dotted lines, while phonological neighbors are connected through dashed lines. Please note segments and tones are both represented (separately) in the phoneme/toneme level. This is only for the purpose of indicating that both segments and tones are important components of a Mandarin syllable; whether or not segments and tones are both included in the measures of phonological similarity will be explored in the main study. The dashed-dotted rectangle refers to level focused in this dissertation.

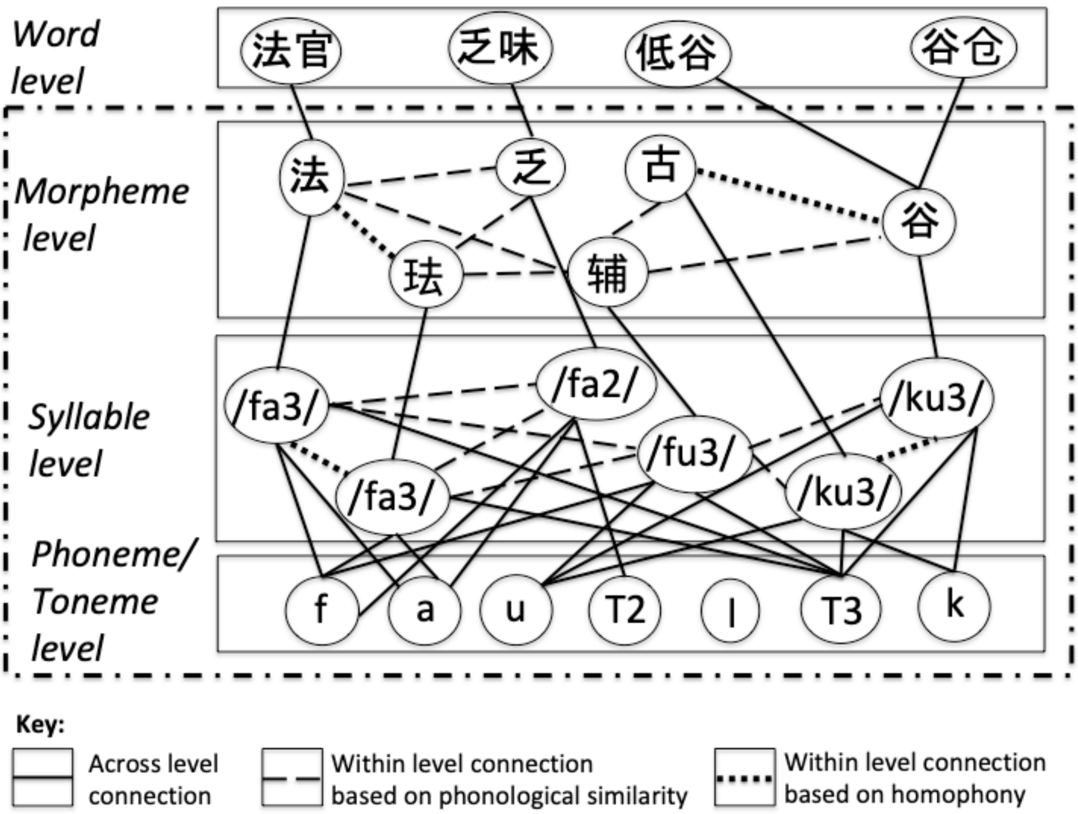

Figure 5: Adapted independent representation model for Mandarin. Here solid line indicates connections between levels. Homophones are connected through dotted lines, while phonological neighbors are connected through dashed lines. Please note segments and tones are both represented (separately) in the phoneme/toneme level. This is only for the purpose of indicating that both segments and tones are important components of a Mandarin syllable; whether or not segments and tones are both included in the measures of phonological similarity will be explored in the main study. The dashed-dotted rectangle refers to level focused in this dissertation.

2.5 Current study

To summarize, the goals of the present dissertation are two-fold: First, to provide a comprehensive view of the effects of phonological neighborhood and homophones on spoken word recognition in Mandarin; Second, this dissertation focuses on exploring whether or not phonological neighbors and homophones have similar effects on spoken word recognition. As discussed in Section 2.3 (Homophones), there are two types of homophone representations i.e. shared and independent. Based on shared homophone representation (Jescheniak & Levelt, 1994), the answer to this question would be that phonological neighbors and homophones have different effects. However, based on independent representation (Caramazza et al., 2001), the answer would be that phonological neighbors and homophones have similar or stronger effects in the same direction. In addition, neither of the models on homophone representation predicts possible interactions among homophone mates nor do they negate the possibility of interaction among homophone mates. Given Wang et al.'s and Zhou's results it seems that homophone density does affect speech perception and possibly interacts with frequency. Therefore, there can be two sets of hypotheses:

(1) Hypothesis 1A:

Given the independent representation model, homophone mates and phonological neighbors should have similar effects on the recognition of spoken real monosyllables in Mandarin. In other words, either both are inhibitory, or both are facilitative, or both have null effects on spoken word recognition. Given previous findings for both English and Mandarin (Luce & Pisoni, 1998; Wang et al., 2012), the most likely pattern is for both phonological neighbors and homophone mates to be inhibitory. Furthermore, the two should also have the same interaction patterns with word frequency. Given previous studies on Chinese homophone effects (Wang et al., 2012; Zhou, 2015), I predict that homophone mates have stronger influence for high frequency words than for low frequency words. Thus, phonological neighbors are predicted to show the same interaction with frequency.

Hypothesis 1B:

Given the shared representation model, homophone mates and phonological neighbors may not have similar effects on the recognition of spoken real monosyllables in Mandarin. In other words, either they will have opposite effects i.e, inhibitory for one and facilitatory for other, or one is inhibitory or facilitaotory while the other has null effects on spoken word recognition. Given previous findings in Mandarin (Wang et al., 2012; Zhou, 2015), the most likely pattern for homophone mates is inhibitory. So, phonological neighbors might have facilitative or null effects in Mandarin. Furthermore, the two should not have the same interaction patterns with syllable frequency. Given previous studies on Chinese homophone effects (Wang et al., 2012; Zhou, 2015), I predict that homophone mates have stronger influence for high frequency syllables than for low frequency syllables. Thus, phonological neighbors are predicted to have no interaction with syllable frequency.

(2) Hypothesis 2:

Following from Luce and Pisoni (1998), I predict inhibitory effects of phonological neighbors on the identification of pseudo-syllables. In other words, pseudo-syllables with many similar-sounding real monosyllables and high neighborhood frequency will be recognized as pseudo-syllables slower and less accurately than pseudo-syllables with fewer similar-sounding real monosyllables and low neighborhood frequency.

In order to achieve these aims, I am conducting two experiments: auditory lexical decision experiment, and auditory naming experiment. These two experiments share stimuli and statistical analysis. Therefore, in the next chapter (Chapter 3: Research

Methods), I am discussing the common features of both the experiments before going into the details of each experiment in Chapter 4: Auditory lexical decision, and Chapter 5: Auditory naming.

Chapter 3: Research Methods

As described in Chapter 2, the goal of this dissertation is to investigate the effects of phonological neighborhoods and homophones on spoken word recognition in Mandarin. To achieve this goal, two experiments were conducted: an auditory lexical decision experiment and an auditory naming experiment. This chapter describes the common aspects of the methodology of these two experiments, in terms of stimuli, lexical measures, and methods of statistical analysis. Methodological differences between the two experiments (for example, the auditory lexical decision task uses both real monosyllables and pseudo-syllables while the auditory naming experiment only uses real monosyllables) are mentioned in this chapter and further explained, together with other detail of experimental methods and results in Chapter 4 and Chapter 5, respectively.

3.1 Stimuli and recording

As described in Chapter 2 Section 2.4, monosyllabic morphemes (with tone), which are mapped to orthographic forms (i.e. characters), are the building blocks of the Mandarin lexicon. The segmental structure of a tonal monosyllable can be expressed as $(C)(G)V(V)^T$ or $(C)(G)V(C)^T$ where C is a consonant, G is a glide, V is a monophthong, VV is a diphthong, and T is a lexical tone. In the onset-rime (or initial-final) structure, the initial C is the onset, and the rest (G)V(V) or (G)V(C) is the rime. As mentioned in Chapter 2 Section 2.4, there are total 24 onset, 33 rimes, and 4 tones in Mandarin, and the complete inventory of monosyllables consists of around 1200 unique monosyllables (not all possible combinations of onset, rime, and tone are attested).

In this dissertation, I started with the exhaustive set of tonal monosyllables in Mandarin when constructing the stimuli set of real monosyllables. A total of 1,259 monosyllables attested in Xinhua dictionary (5^{th} edition) were included in the set. Complete list of real monosyllables used in the study can be found in Appendix A.

The set of pseudo-syllables consisted of 768 unique syllables, all of which are *accidental* gaps given the Mandarin syllable inventory. *Accidental* gaps are lexical gaps that are phonotactically legal but do not occur in Mandarin due to unknown reasons, in contrast with *systematic* gaps, which are phonotactically illegal (for example, the CV sequence in /ʃi/ is not allowed by Mandarin phonotactic rules; Kirby & Yu, 2007). I further divide accidental gaps into tonal gaps and segmental gaps. Tonal gaps are illegal syllables due to a gap in tone; in other words, the same segmental composition is attested in Mandarin lexicon with a different tone. For example, /tsweɪ2/ is not attested in Mandarin but /tsweɪ4/ is; thus /tsweɪ2/ is a tonal gap. Segmental gaps, on the other hand, are illegal syllables that are not attested in any tone although the segmental composition conforms to the phonotactic rules, e.g. /mʊŋ1/, /mʊŋ2/, /mʊŋ3/ and /mʊŋ4/ are all segmental gaps. In the set of pseudo-syllables used in this study, 356 are tonal gaps, and 412 are segmental gaps. Systematic gaps are intentionally avoided when designing the stimuli set because phonotactically illegal segmental sequences are too difficult to pronounce for native speakers, which makes it hard to produce natural-sounding auditory stimuli for the experiments. Complete list of pseudo-syllables used in this study can be found in Appendix B.

While both auditory lexical decision and auditory naming experiments used real monosyllable stimuli, only the auditory lexical decision experiment used pseudo-syllables in the stimuli. All the auditory stimuli (both real monosyllables and pseudo-syllables) were recorded by a female native Mandarin speaker in her late 20s. The speaker has extensive training in Mandarin phonetics and has previously worked in the profession of teaching Chinese as a foreign language. The recording took place in a sound-attenuated room using a uni-directional microphone that was routed to Digi design recording system. Each item was produced three times, each time preceded by a carrier phrase ("item number"). For each item, the token with the best sound quality—typically the second production—was chosen as a stimulus of the experiments. All the stimuli were intensity-normalized at 70 dB in Praat (Boersma & Weenink, 2010).

The duration of the stimulus ranged from 0.35s to 0.99s for real monosyllables and 0.35s to 0.86s for pseudo-syllables. The mean stimulus duration for real monosyllables were 0.62s (SD = 0.1) and for pseudo-syllables were 0.58s (SD = 0.09). Figure 6 and Figure 7 shows the distribution of raw and log-transformed stimulus duration for real monosyllables and pseudo-syllables.

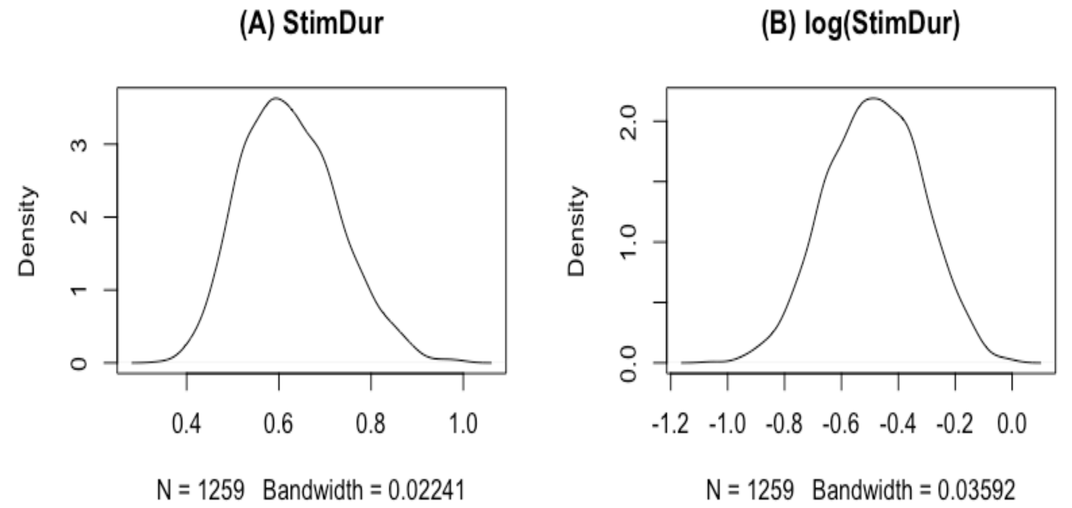

Figure 6: Distribution of raw and log-transformed stimulus duration (StimDur) for real monosyllables. X-axis indicates syllable duration in s.

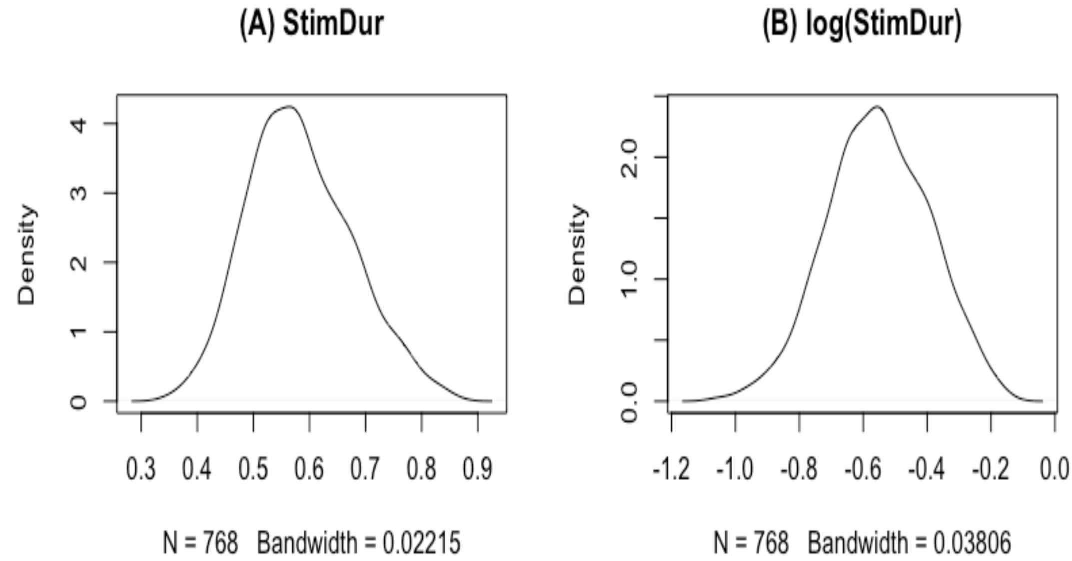

Figure 7: Distribution of raw and log-transformed stimulus duration for pseudo-syllables. X-axis indicates syllable duration in s.

3.2 Lexical measures

A number of lexical measures were compiled for each stimulus item. Given the goals of this study, the critical lexical measures are related to phonological neighborhood (neighborhood density and neighbor frequency) and homophone families (homophone density and frequency of homophone mates). While neighborhood measures are available for both real monosyllables and pseudo-syllables (since pseudo-syllables may sound similar to real syllables), homophone-related measures are only available for real monosyllables. In addition, a small number of real monosyllables do not have entries in the lexical database I used (see detail below), probably due to extremely low frequency of occurrence in everyday use of the language. As a result, the total number of items with lexical measures varies slightly from 1,259.

3.2.1 Neighborhood measures

The two key measures of phonological neighborhoods are neighborhood density and neighborhood frequency. Neighborhood density refers to the number of neighbors of the target monosyllable, and neighborhood frequency refers to the frequency of the neighbors, which is usually calculated as either the sum or average of neighbor frequency. In this study, I use average neighbor frequency, because it is not as correlated with neighborhood density as neighborhood density is.

What is less clear, however, is the definition of neighborhood. As discussed in Chapter 2, multiple versions of neighborhood definition have been proposed, ranging from neighborhood probability rules, the phi-square rule, to the one-phoneme difference rule. In the current literature on phonological neighborhood effects, the most widely used definition of phonological neighborhood is the one-phoneme difference rule, according to which any words that differ by one phoneme due to addition, deletion or substitution of a phoneme are considered as phonological neighbors.

While the one-phoneme difference rule has achieved much success in the study of English and other European languages, it remains to be seen whether it can be applied to the study of a tonal language. As discussed in Chapter 2 Section 2.2.2, previous studies of Mandarin have generally applied the one-phoneme difference rule, but differed in whether tone was ignored or treated with the same status as a phoneme. Furthermore, if the one-phoneme difference rule is construed as a one-unit difference rule in general, there should be at least two ways of applying the rule to Mandarin monosyllables, in accordance with the two ways of segmenting a Mandarin monosyllable (Duanmu, 2007), i.e. at the phoneme level and the component level (onset, rime), respectively.

In this dissertation, I consider a total of four possibilities of defining a phonological neighborhood for Mandarin monosyllables (i.e. four neighborhood schemes): (1) one-segment/tone difference rule, (2) one-segment difference rule, (3) one-component/tone difference rule, and (4) one-component difference rule. Overall, all four rules follow the general spirit of the one-unit difference rule, but differ in terms of (1) whether tone is considered as a unit or not, and (2) whether the unit for segmental sequence is segment or component. Specifically, under the one-segment/tone difference rule, monosyllables that differ in one and only one segment (phoneme) or tone via addition, deletion or substitution are considered neighbors. For example, /ɕjaʊ3/, /laʊ3/, /ljaŋ 3/ and /ljaʊ2/ are all neighbors of

/ljaʊ3/, but /ɕjaʊ1/ is not. Under the one-segment difference rule, monosyllables that differ in one and only one segment through addition, deletion or substitution are considered neighbors. For example, /ɕjaʊ2/, /laʊ3/, /ljaŋ3/, and /ɕjaʊ1/ are all neighbors of /ljaʊ3/. Under the one-component/tone difference rule, monosyllables that differ by one and only one component (onset or rime) or tone via addition, deletion or substitution are considered neighbors. For example, /laŋ3/, /mjaʊ3/ and ljaʊ2/ are all neighbors of /ljaʊ3/. Lastly, under the one-component difference rule, monosyllables that differ by one and only one component (onset or rime) via addition, deletion or substitution are considered neighbors. For example, /laʊ/, /lu/, /pjaʊ/ are neighbors of /ljaʊ/.

The rationale for considering four different neighborhood schemes is twofold. First, for the current study, it is important to clarify whether any observed neighborhood effects on spoken word recognition are only specific to a certain definition of neighborhood or whether the effects are robust enough to be observed across different ways of defining the neighborhood. Secondly, comparing neighborhood measures from four different neighborhood schemes also gives us insight on the relative fitness of the neighborhood definitions.

To be sure, there is a couple of existing databases that provide phonological neighborhood measures, such as Neergaard, Xu and Huang (2016) and Sun, Hendrix, Ma and Baayen (2018) databases. Neergaard et al.'s database included neighborhood measures under various definitions of phonological neighborhood, including some similar to the four listed in Table 3. However, the entries in Neergaard et al.'s database are words—which may be monosyllabic, disyllabic or multisyllabic—instead of monosyllabic morphemes. In other words, Neergaard et al.'s database is word-based, and monosyllabic morphemes that are not used as standalone words (i.e. bound morphemes) will not be listed in the database; neither will monosyllabic bound morphemes be included for the calculation of neighborhood measures. By contrast, the current study is focused on lexical effects at the morpheme level, thus the phonological neighborhoods concerned in this study are constructed by all monosyllabic morphemes.

Sun et al.'s Chinese lexical database provides neighborhood measures for both words and morphemes. However, Sun et al.'s database only provides neighborhood measures for one neighborhood scheme, which is based on the one-segment/tone difference rule. As mentioned above, including four different neighborhood schemes in the current study will shed light on the robustness of neighborhood effects (if any) and the relative fitness of the neighborhood definition for modeling neighborhood effects in Mandarin.

In view of the limitations of existing neighborhood measure databases, I calculated the neighborhood measures used in this study separately, based on character-based usage frequency counts from the SUBTLEX-CH corpus (Cai & Brysbaert, 2010). The SUBTLEX-CH corpus is a corpus of film and television subtitles of around 47 million characters. The authors of the SUBTLEX-CH corpus also published, based on the corpus data, a list of character frequency counts that includes 5936 unique characters. I used the character frequency list from SUBTLEX-CH to calculate four sets of neighborhood measures, one for each neighborhood scheme. Probably due to extremely low usage frequency, there are 123 monosyllables that are in the Xinhua dictionary that did not appear in the SUBTLEX-CH corpus, and are therefore excluded from the calculation of both neighborhood density and neighbor frequency in all four neighborhood schemes. The same methods of calculating neighborhood measures for real monosyllables are applied to the calculation of neighborhood measures for pseudo-syllables.

Table 3: Neighborhood measures under four neighborhood schemes.

Neighborhood scheme	**Neighborhood density (ND) measure**	**Neighbor frequency (NF) measure**
One-segment/tone difference rule	ND_SegT	NF_SegT
One-segment difference rule	ND_Seg	NF_Seg
One-component/tone difference rule	ND_CompT	NF_CompT
One-component difference rule	ND_Comp	NF_Comp

Figure 8 and Figure 9 shows the distribution of ND measures across neighborhood schemes for real monosyllables and pseudo-syllables. Figure 10 and Figure 11 shows the distribution of NF measures across neighborhood schemes for real monosyllables and pseudo-syllables. Because frequency measures are highly skewed, they are log-transformed before being entered into regression models. Table 4 and Table 5 lists the summary statistics of NF measures across neighborhood schemes.

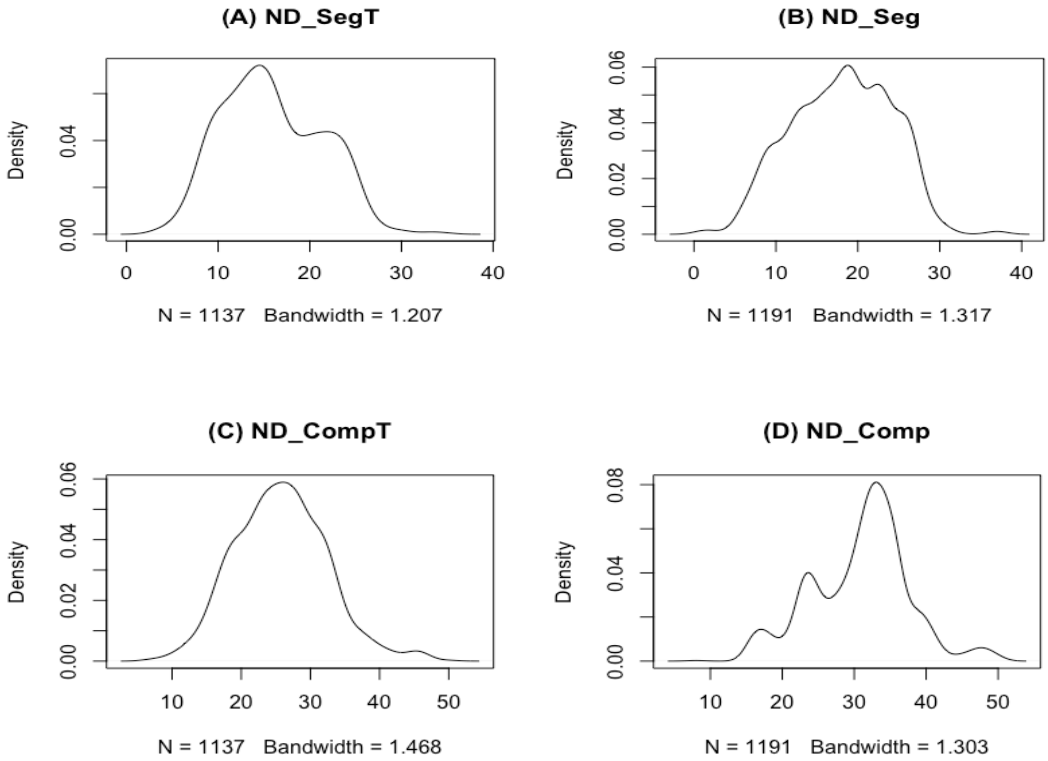

Figure 8: Distribution of ND across different neighborhood schemes for real monosyllables.

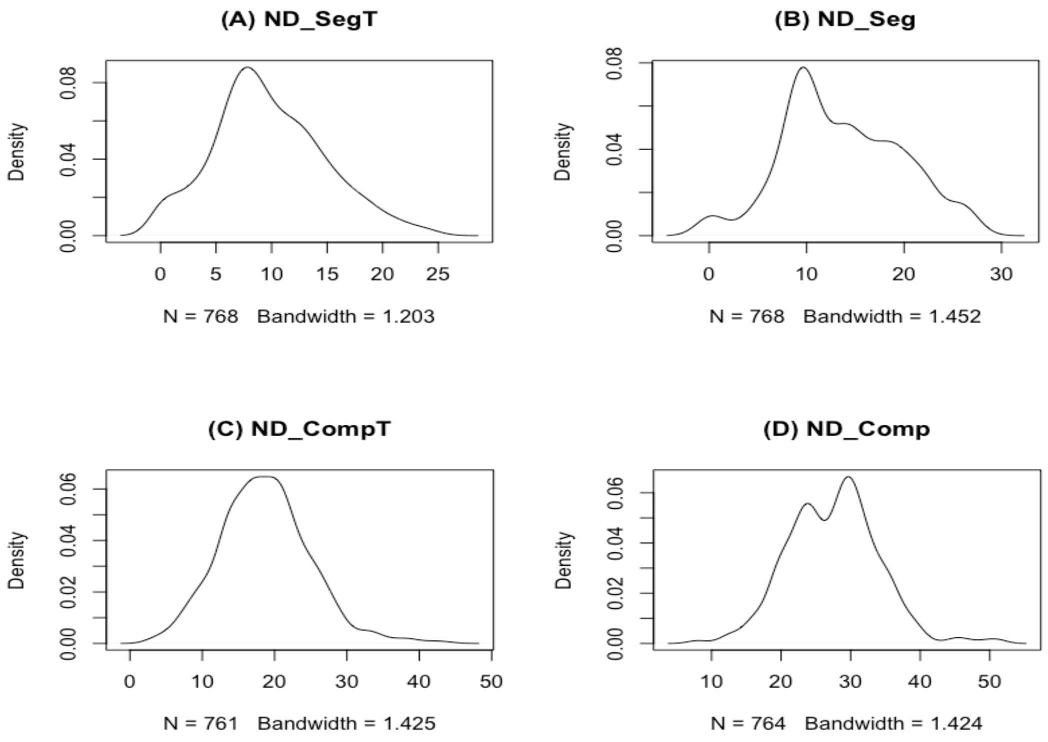

Figure 9: Distribution of ND across different neighborhood schemes for pseudo-syllables.

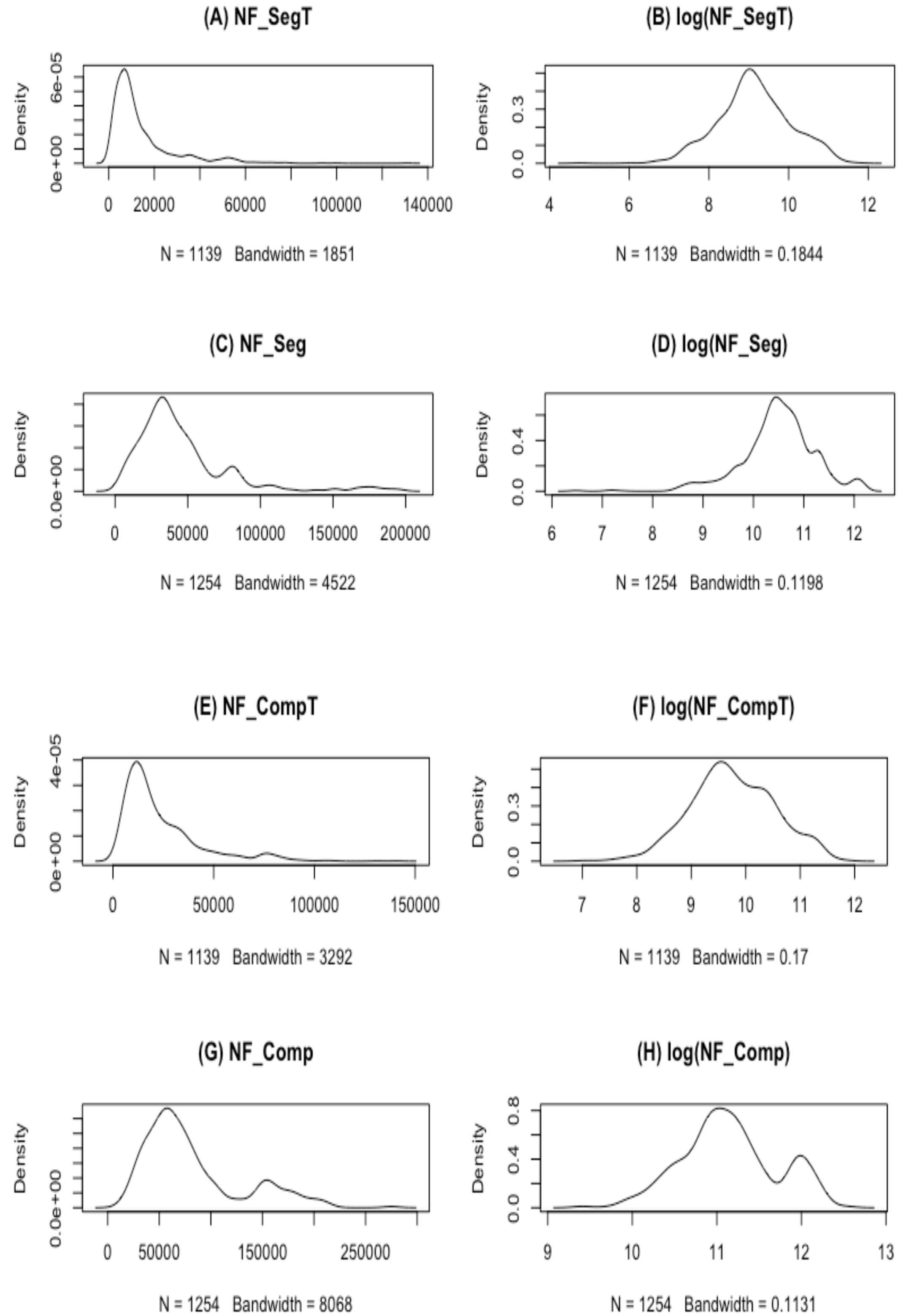

Figure 10: Distribution of raw and log-transformed neighborhood frequency across different neighborhood schemes for real monosyllables. X-axis indicates number of occurrences per million.

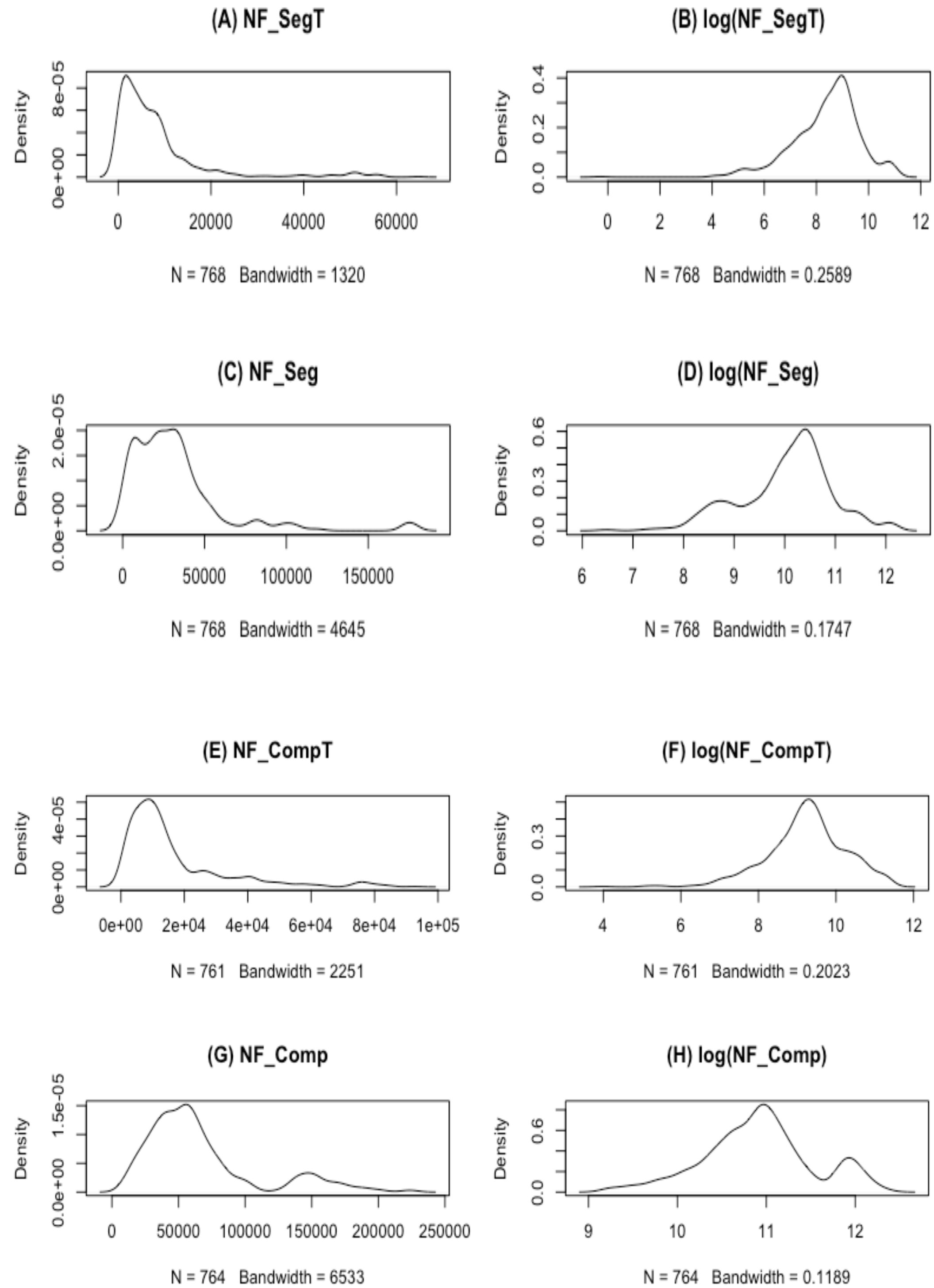

Figure 11: Distribution of raw and log-transformed neighborhood frequency across different neighborhood schemes for pseudo-syllables. X-axis indicates number of occurrences per million.

3.2.2 Homophone measures

Critical homophone measures include homophone density and homophone frequency. Homophone density refers to the number of homophone mates in a homophone family. Homophone frequency describes the usage frequency of the homophone mates in a homophone family. Unlike the definition of neighborhood, the definition of homophone family is much clearer: only items that are identical in pronunciation are considered as homophones.

There are multiple ways of estimating homophone frequency: sum homophone frequency (i.e. total frequency of the monosyllable), maximum homophone frequency (i.e. the frequency of the highest-frequency homophone mate), and average homophone frequency (i.e. sum homophone frequency divided by homophone density). In my dataset, all three frequency measures are quite highly correlated (correlation coefficient $r > 0.9$ among log-transformed homophone frequency measures). In addition, both sum homophone frequency and maximum homophone frequency are also highly correlated with homophone density ($r > 0.9$) while average homophone frequency is much less correlated with homophone density ($r = 0.1$). To avoid having highly correlated variables in the regression models, I use average homophone frequency as a measure of the frequency of homophone mates in the family.

Both homophone density (HD) and homophone frequency (HF) are calculated from the character frequency list from the SUBTLEX-CH corpus (Cai & Brysbaert, 2010). A total of 287 monosyllables out of 1259 real monosyllables have only one homophone mate in the homophone family (for example, / kən2, / liɛ1/, / paɪ2/ etc.), and the remaining real monosyllables have at least two homophone mates in the family. Figure 12 shows the distribution of HD. Figure 13 shows the distribution of raw and log-transformed HF. Summary statistics of HD and HF are included in Table 4.

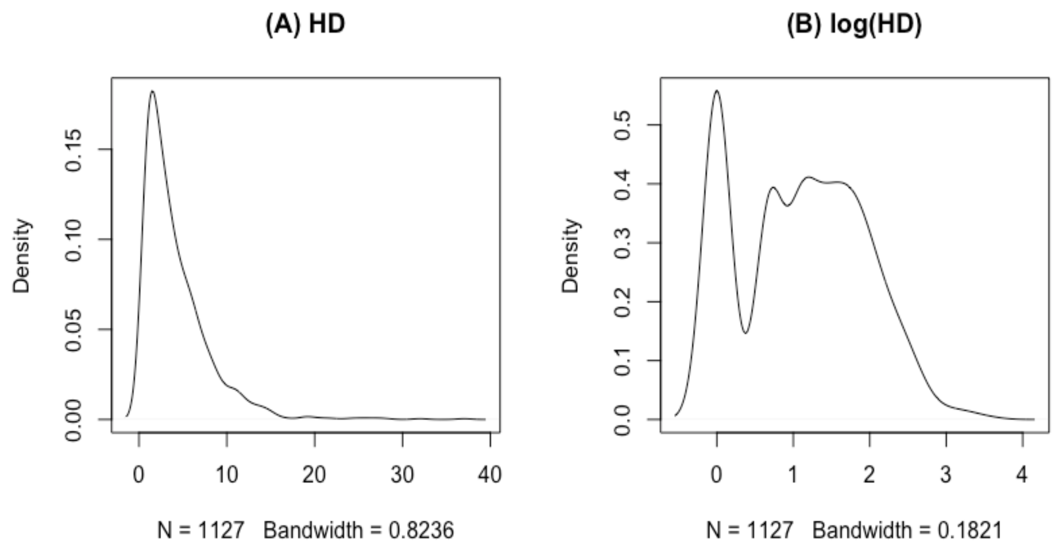

Figure 12: Distribution of raw and log-transformed homophone density.

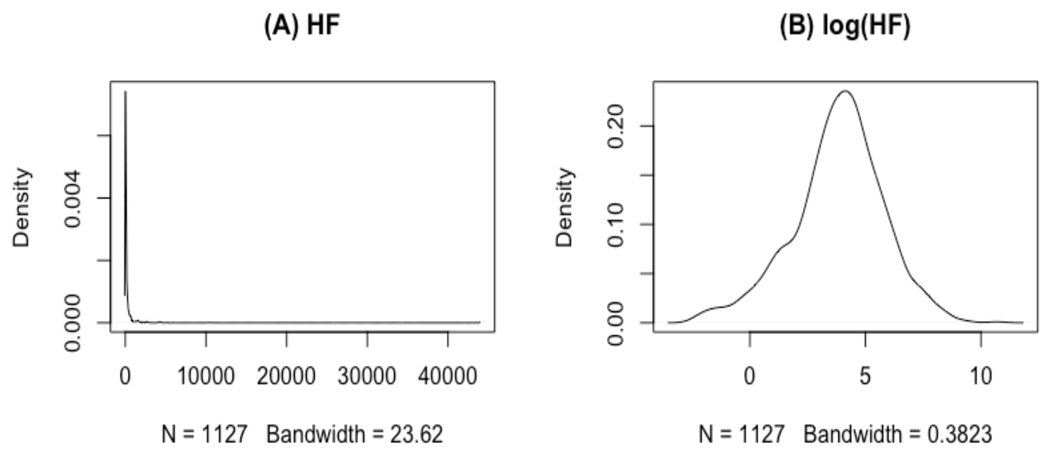

Figure 13: Distribution of raw and log-transformed HF.

Table 4: Summary statistics of lexical measures for real monosyllables. The unit of NF and HF measures is number of occurrences per million.

	Minimum	Maximum	Mean	Median	SD
ND_SegT	3.00	35.00	15.89	15.00	5.48
NF_SegT	19.52	7708.19	885.39	609.81	881.82
ND_Seg	1.00	37.00	18.02	18.00	6.03
NF_Seg	328.60	13088.21	2631.52	2090.81	2119.27
ND_CompT	7.00	50.00	25.76	26.00	6.66
NF_CompT	91.14	4853.71	887.80	688.02	668.13
ND_Comp	8.00	50.00	30.89	32.00	6.70
NF_Comp	450.11	9017.82	2602.18	2261.01	1377.63
HD	1.00	37.00	4.19	3.00	3.85
HF	0.09	43956.70	269.51	51.85	1476.39

Table 5: Summary statistics of lexical measures for pseudo-syllables. The unit of NF measures is number of occurrences per million.

	Minimum	Maximum	Mean	Median	SD
ND_SegT	0.00	25.00	9.79	9.00	5.05
NF_SegT	0.00	11429.53	781.62	518.01	975.62
ND_Seg	0.00	28.00	13.69	13.00	6.09
NF_Seg	0.00	13589.81	2313.20	1846.38	2331.91
ND_CompT	3.00	44.00	18.67	18.00	6.26
NF_CompT	8.93	6650.76	860.79	585.63	878.90
ND_Comp	8.00	51.00	27.49	28.00	6.47
NF_Comp	520.43	9777.48	2369.82	2037.16	1394.38

3.2.3. Other lexical measures

In addition to neighborhood and homophone measures, I also included *Tone* (T1, T2, T3, T4) as a control variable in the regression analysis.

3.3 Statistical analysis

The complete set of real monosyllables and pseudo-syllables ranged in syllable length (i.e. number of phonemes) from 1 to 4 phonemes. As previous research has pointed out, neighborhood density decreases significantly when syllable length increases. In other words, shorter monosyllables naturally have higher neighborhood density compared to longer syllables, and it will be unfair to compare neighborhood density across syllable lengths. The same correlation pattern is observed in the current stimuli set, regardless of which neighborhood scheme is used (see Figure 14). Thus, to control for syllable length, the statistical analysis in this dissertation only examines 3-phoneme syllables (N = 677 for real monosyllables, N = 387 for pseudo-syllables), which comprises the largest group in the stimuli set, for both real monosyllables and pseudo-syllables.

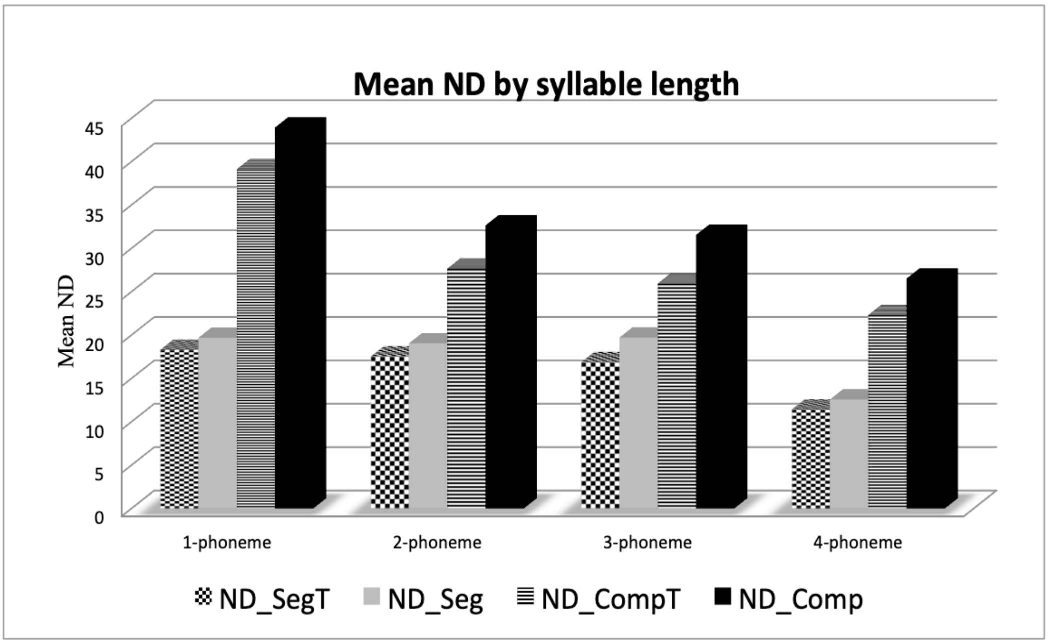

Figure 14: Mean neighborhood density by syllable length (i.e. number of phonemes in the syllable) by neighborhood scheme, based on the stimuli set of real monosyllables.

Data from both auditory lexical decision experiment and auditory word naming experiment were analyzed by mixed-effect regression models. In this Section, I describe the common methods of constructing, trimming and selecting the mixed-effect models that are shared by the two experimental studies.

3.3.1 Model construction

Both studies made use of two types of mixed effect models: linear mixed-effects models and generalized linear mixed-effects models and. Both models contain fixed-effects predictors and random-effects predictors, but in a linear mixed-effects model is used to describe variation in a continuous variable while a generalized linear mixed-effects model is used to describe variation in a binary outcome variable.

In this dissertation, linear mixed-effects models were used to model reaction time (RT), while generalized linear mixed-effects models were used to model response accuracy (Accuracy; “0” = incorrect, “1” = correct). All models were constructed in using *lmer* function in the lme4 package (Bates, Maechler, Bolker, Walker, & others, 2014) in R (Team, 2014). Reaction time was log-transformed before being entered into the model.

In the auditory lexical decision experiment, data from real monosyllables and data from pseudo-syllables are modelled separately (among other things, pseudo-syllables do not have lexical measures such as homophone density or homophone frequency). The auditory naming experiment only had real monosyllables, but the experiment consisted of two different tasks: instantaneous naming and delayed naming. Data from instantaneous naming and data from delayed naming are modelled separately.

According to the hypotheses laid out in Chapter 2 Section 2.5 the critical effects under investigation for real monosyllables are the main effects of neighborhood density, neighbor frequency, homophone density, and homophone frequency as well as the interaction of neighborhood density and syllable frequency and the interaction of homophone density and syllable frequency. To avoid the high correlation between syllable frequency and homophone density, I used homophone frequency (i.e. syllable frequency divided by homophone density) as a proxy for syllable frequency. In addition, all the models on real monosyllables also included log-transformed stimulus duration (StimDur) and Tone as control predictors, as well as intercepts for Subject and Item as random effects to control for individual differences across subjects and items. Furthermore, the models for delayed naming task in the auditory naming experiment also included Delay condition (600ms, 1200ms) as a control predictor.

(9) and (10) below give the general initial formula for models on real monosyllables in the auditory lexical decision experiment and the instantaneous naming task in the auditory naming experiment. ND and NF refer to measure of the neighborhood density and neighbor frequency.

(9) log(RT) ~ ND +log(NF) + log(HD) + log(HF) + ND:log(HF) + log(HD):log(HF) + log(StimDur) + Tone + (1|Subject) + (1|Item)

(10) Accuracy ~ ND + log(NF) + log(HD) + log(HF) + ND:log(HF) + log(HD):log(HF) + log(StimDur) + Tone + (1|Subject) + (1|Item)

(11) and (12) below give the initial model formula for models on the delayed naming task in the auditory naming experiment.

(11) Accuracy ~ ND +log(NF) + log(HD) + log(HF) + ND:log(HF) + log(HD):log(HF) + Delay + log(StimDur) + Tone + (1|Subject) + (1|Item)

(12) log(RT) ~ ND +log(NF) + log(HD) + log(HF) + ND:log(HF) + log(HD):log(HF) + Delay + log(StimDur) + Tone + (1|Subject) + (1|Item)

As shown in the model formulas above, in addition to RT, StimDur and all the frequency-related measures are also log-transformed. Furthermore, all numerical predictors were centered before being entering into the models in order to reduce collinearity. Effectiveness of different neighborhood schemes was compared by using neighborhood measures from the four different neighborhood schemes (see below for model trimming and selection detail). Table 6 gives the correlations among all numeric fixed-effects predictors for real monosyllables under each neighborhood scheme.

Models for the data from pseudo-syllables (in auditory lexical decision only) have much simply structure, since pseudo-syllables do not have HD and HF measures. Critical fixed-effects predictors include only ND and log(NF). Similar to models for real monosyllables, the models for pseudo-syllables also include log(StimDur) and Tone as control predictors, as well as an additional control predictor, ItemType, which indicates the type of lexical gaps (segmental gap vs. tonal gap). The models for pseudo-syllables also include random intercepts for Subject and Item to control for individual differences across subjects and items. The general initial model formula is given below in (13) and (14). Correlation coefficients among numeric fixed-effects predictors under each neighborhood scheme are given in Table 7.

(13) log(RT) ~ ND +log(NF) + ItemType + log(StimDur) + Tone + (1|Subject) + (1|Item)

(14) Accuracy ~ ND + log(NF) + ItemType + log(StimDur) + Tone + (1|Subject) + (1|Item)

Table 6: Correlation coefficients among numeric fixed predictors for real monosyllables, under each neighborhood scheme.

	One-segment/tone difference rule		One-segment difference rule		One-component/tone difference rule		One-component difference rule		log(HD)	log(StimDur)	Log(HF)
	ND_SegT	log(NF_SegT)	ND_Seg	log(NF_Seg)	ND_CompT	log(NF_CompT)	ND_Comp	log(NF_Comp)			
ND_SegT	-	0.41*	0.87*	-	0.70*	-	0.61*	-	0.05	-0.16*	0.01
log(NF_SegT)	-	-	-	0.62*	-	0.74*	-	0.45*	0.27*	-0.22*	0.14*
ND_Seg	-	-	-	0.46*	0.55*	-	0.71*	-	-0.06*	-0.11*	-0.01
log(NF_Seg)	-	-	-	-	-	0.45*	-	0.67*	0.20*	-0.11*	0.12*
ND_CompT	-	-	-	-	-	0.46*	0.79*	-	0.13*	-0.32*	0.03
log(NF_CompT)	-	-	-	-	-	-	-	0.60*	0.30*	-0.29*	0.03

ND_Comp	-	-	-	-	-	-	-	0.51*	-0.05	-0.26*	-0.01
log(NF_Comp)	-	-	-	-	-	-	-	-	0.22*	-0.25*	0.15*
log(HD)	-	-	-	-	-	-	-	-	-	-0.16*	0.16*
log(StimDur)	-	-	-	-	-	-	-	-	-	-	-0.02

Table 7: Correlation coefficient among numeric fixed predictors for pseudo-syllables, under each neighborhood scheme.

	One-segment/tone difference rule		One-segment difference rule		One-component/tone difference rule		One-component difference rule		log(StimDur)
	ND_SegT	log(NF_SegT)	ND_Seg	log(NF_Seg)	ND_CompT	log(NF_CompT)	ND_Comp	log(NF_Comp)	
ND_SegT	-	0.56*		-		-		-	-0.003
log(NF_SegT)	-	-	-		-		-		-0.18*
ND_Seg	-	-	-	0.57*					0.03
log(NF_Seg)	-	-	-	-					-0.04
ND_CompT	-	-	-	-	-	0.51*			-0.14*
log(NF_CompT)	-	-	-	-	-	-			-0.24*
ND_Comp	-	-	-	-	-	-	-	0.55*	-0.12*
log(NF_Comp)	-	-	-	-	-	-	-	-	-0.26*

3.3.2 Model trimming and selection

As mentioned above, data from auditory lexical decision and auditory naming are modelled separately. Furthermore, for auditory lexical decision, data for real monosyllables and pseudo-syllables are modelled separately; for auditory naming, which only included real monosyllables in the stimuli, data from instantaneous and delayed naming tasks are modelled separately. On top of this, for each sub-group of data, response accuracy (Accuracy) and reaction time (RT) are modelled separately.

For each line of modelling (i.e. Accuracy/RT on a specific sub-group of data), four initial models are built, each using a different neighborhood scheme, following the initial model formulas described in (9) – (14) in the Section above. Each initial model then undergoes a procedure of backward elimination in order to trim off non-significant predictors. In each round of backward elimination, the significance of one fixed-effects predictor is tested by comparing the model fit of the current model and a reference model where the predictor under testing is removed. If the change in model fit is not significant ($p > .05$ in the anova test that compares model fit), the predictor under testing will be deemed non-significant and trimmed out. The backward elimination procedure starts with testing the predictor with the lowest t value in the initial model, and proceeds until all remaining predictors are significant. Remaining predictors in general have t values no less than 2, which roughly corresponds to p values under 0.05 given a data size of more than 10000 data points (Krajewski & Matthews, 2010).

After applying backward elimination to all four initial models, each line of modelling has four final models. I then compare the model fit (AIC values) of the four final models, and select the model with the lowest AIC as the best-fit model for the specific line of modelling (Anderson & Burnham, 2004). In the following two chapters, I report the results from the final models of each neighborhood scheme, and the analysis is based on results from the best-fitting final model.

Chapter 4: Auditory lexical decision

This chapter brings forth the first experimental study of the current dissertation. This experiment examines the effect of phonological neighborhood and homophones using an auditory lexical decision task.

4.1 Auditory lexical decision experiment

To testify the aims of the current dissertation, auditory lexical decision was selected as one of the spoken word recognition task. In a regular auditory lexical decision task, participants hear speech stimuli and their task is to categorize the heard stimuli as words or non-words as quickly as possible. The concept of word is a bit vague in Chinese (see Section 2.4 Background on Mandarin). Monosyllabic morphemes form the building block of Mandarin, where each monosyllabic morpheme corresponds to a Chinese character (orthographic form). Therefore, the task of this experiment was not exactly the same as in a canonical auditory lexical decision task. The participants were asked to decide whether or not they could associate the heard monosyllable (stimulus) to at least one character in Chinese. The stimuli of this experiment consist of two groups: real monosyllables (i.e. syllables that can be associated with at least one real characters) and pseudo-syllables (i.e. syllables that cannot be associated with any real character).

4.2 Method

4.2.1 Participants

Seventy-eight participants took part in this experiment (49M, 29F, mean age = 23.4 years, SD = 4.26). All participants were recruited from the Hong Kong Polytechnic University and were native speakers of Mandarin, born and raised in Mainland China. All participants were right-handed and reported no speech and hearing problem.

4.2.2 Stimuli

1259 real monosyllables and 768 monosyllabic pseudo-syllables were used in the experiment. All pseudo-syllables were accidental gaps that are either tonal gaps (N = 356) or segmental gaps (N = 412). See Chapter 3, Section 3.1 Stimuli and recording for details.

4.2.3 Design and procedure

Due to the large number of stimuli (i.e. 1259 real monosyllables and 768 pseudo-syllables), it is impossible to include all the stimuli in one experimental session without making the experiment too demanding for the participant. Therefore, I divided the stimuli into 6 sub-lists, each with. 425 unique stimulus items (50% real monosyllables and 50% pseudo-syllables). Each stimulus item was presented twice in a sub-list, resulting in a total of 850 stimulus tokens per sub-list, with a completely randomized order of presentation. Each participant worked on only one sub-list, and each sub-list was presented to 13 participants. To evaluate cross-group differences, 12 stimulus items (6 real monosyllables and 6 pseudo-syllables) were shared by all sub-lists; apart from that, all other real monosyllables only appeared in one sub-list and all other pseudo-syllables appeared in one or two sub-lists.

The experiment was designed and conducted using E-Prime version 2.0 (Schneider, Eschman, & Zuccolotto, 2007) on a Lenovo laptop. The laptop was connected to a set of headphones and Chronos response box. The experiment was conducted in a sound-treated room. Each participant was comfortably seated before the start of the experiment. Each trial proceeded as follows: a fixation-cross appeared at the center at the beginning of the trial for 500 ms, followed by the auditory stimulus. Participants had 4000 ms from the onset of stimulus to respond by pressing the relevant key on the response box: if the stimulus could be associated with at least one Chinese character (i.e. real monosyllable), the participant would press the key “5”, the rightmost key on the response box; if the stimulus could not be associated with any character (i.e. pseudo-syllable), the participant would press the leftmost key “1”. If the participant did not respond within 4000 ms of the stimulus onset, the experiment will proceed to the next trial. Participants were instructed to make a judgment as quickly and accurately as possible, and reaction times were measured from the onset of the stimulus to the button-press response. Figure 15 shows a diagram of the experimental trial.

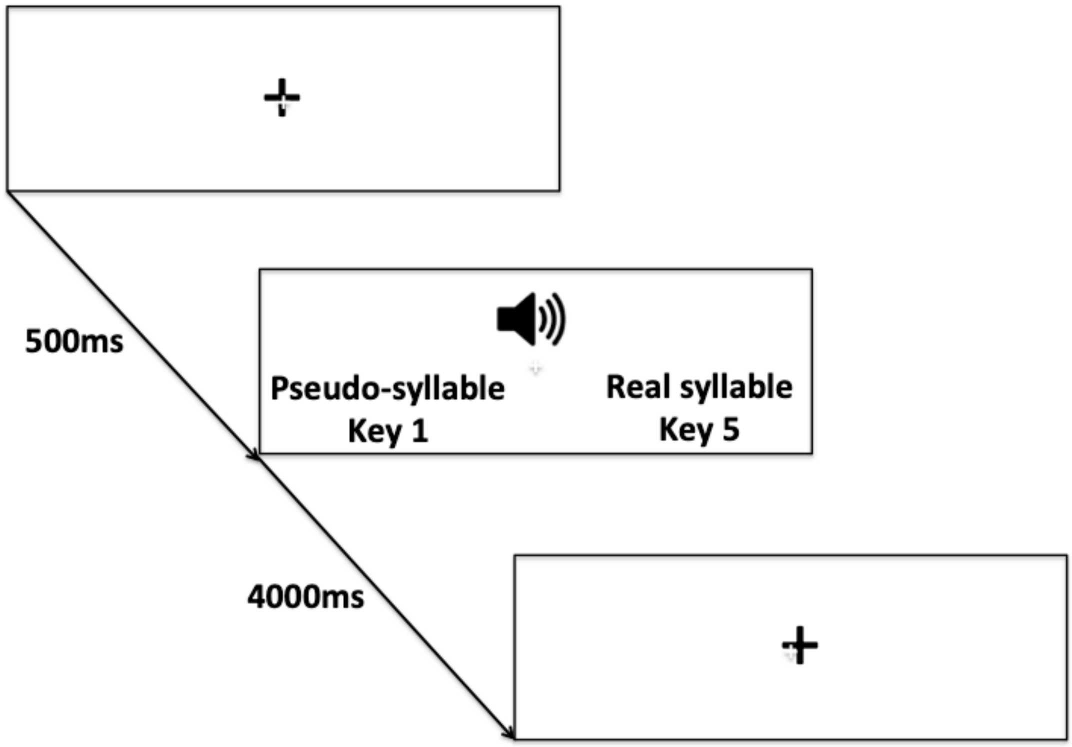

Figure 15: Diagram of an experimental trial of auditory lexical decision task.

Before the test session began, the participant would first have a practice session with 30 practice items. Feedback was provided for the practice trials only. None of the syllables presented during the practice trial of the experiment appeared in the main experiment. A complete experimental session lasted no more than 30 minutes for all the participants, with one mandatory break in between.

4.2.4 Analysis

As mentioned above, in this thesis, statistical analysis focused on the performance from 3-phoneme monosyllables (N = 677) and pseudo-syllables (N = 387). The experimental data were analyzed in terms of both accuracy (incorrect = "0", correct

= "1") and reaction time (RT; measured from the onset of the stimulus to the button-press response). Furthermore, the reaction time analysis is only applied to trials with correct responses.

General modeling methods are described in Chapter 3 Section 3.3. Neighborhood density (ND), neighborhood frequency (NF), homophone density (HD), homophone frequency (HF), interaction between HD and HF, and interaction between ND and HF were critical fixed-effects predictors in models for real monosyllables. In addition, the models for real monosyllables had stimulus duration (StimDur) and Tone as control fixed-effects predictors as well as intercepts for Subject and Item as random effects. Models for pseudo-syllables only had ND and NF as critical fixed-effects predictors (since pseudo-syllables do not have HF or HD measures). In addition to StimDur and Tone, the models for pseudo-syllables also had an additional control variable of item type (ItemType), which encoded the type of gap (tonal gap, segmental gap). The models for pseudo-syllables also had random intercepts for Subject and Item.

Four neighborhood schemes, each associated with a set of ND and NF measures, were tested in the models. Model trimming and selection followed the procedure described in Chapter 3 Section 3.3. In the next Section, I report the results of the final models with each set of ND and NF measures and focus the discussion on the best-fitting models selected from the four neighborhood schemes.

4.3 Results

4.3.1 Real monosyllables

The mean RT for the lexical decision task was 1023.22 ms (*SD* = 373.58 ms) and the mean accuracy rate was 84%. Before doing further analysis, data were trimmed for outliers. Two types of outliers were removed: (1) items with accuracy rate less than 20%, and (2) trials that were 2 SD above and below the mean RT. Altogether, 6.89% of the trials were excluded.

As mentioned in Section 2.5 Current study, here is the hypotheses for real monosyllables.

(1) Hypothesis 1A:

Given the independent representation model, homophone mates and phonological neighbors should have similar effects on the recognition of spoken real monosyllables in Mandarin. In other words, either both are inhibitory, or both are facilitative, or both have null effects on spoken word recognition. Given previous findings for both English and Mandarin (Luce & Pisoni, 1998; Wang et al., 2012), the most likely pattern is for both phonological neighbors and homophone mates to be inhibitory. Furthermore, the two should also have the same interaction patterns with word frequency. Given previous studies on Chinese homophone effects (Wang et al., 2012; Zhou, 2015), I predict that homophone mates have stronger influence for high frequency words than for low frequency words. Thus, phonological neighbors are predicted to show the same interaction with frequency.

Hypothesis 1B:

Given the shared representation model, homophone mates and phonological neighbors may not have similar effects on the recognition of spoken real monosyllables in Mandarin. In other words, either they will have opposite effects i.e, inhibitory for one and facilitatory for other, or one is inhibitory or facilitaotory while the other has null effects on spoken word recognition. Given previous findings in Mandarin (Wang et al., 2012; Zhou, 2015), the most likely pattern for homophone mates is inhibitory. So, phonological neighbors might have facilitative or null effects in Mandarin. Furthermore, the two should not have the same interaction patterns with syllable frequency. Given previous studies on Chinese homophone effects (Wang et al., 2012; Zhou, 2015), I predict that homophone mates have stronger influence for high frequency syllables than for low frequency syllables. Thus, phonological neighbors are predicted to have no interaction with syllable frequency.

4.3.1.1 Accuracy analysis

The dataset for the models on accuracy consisted of 15508 trials of 598 item types. Four separate models were constructed, each representing neighborhood measures from a different scheme: (1) models with neighborhood measures from one-segment/tone difference scheme (ND_SegT, NF_SegT); (2) models with neighborhood measures from one-segment difference scheme (ND_Seg, NF_Seg); (3) models with neighborhood measures from one-component/tone difference scheme (ND_CompT, NF_CompT); (4) models with neighborhood measures from one-component difference scheme (ND_Comp, NF_Comp). All the initial models follow the same formula (10) from chapter 3, repeated below), although the neighborhood measures were calculated using different schemes.

(10) Accuracy ~ ND + log(NF) + log(HD) + log(HF) + ND:log(HF) + log(HD):log(HF) + log(StimDur) + Tone + (1|Subject) + (1|Item)

All four initial models underwent a backward elimination procedure (as described in Chapter 3, Section 3.3.2 Model construction) so that only significant predictors remain. Only results from the final models are reported here. Table 8 summarizes the model fit (AIC) and all fixed effects in all four final models. AIC values were used to decide the best neighborhood scheme: the lower the AIC, the better the model fit (Anderson & Burnham, 2004). As shown in Table 8, all four models have the same AIC (9048.9) because none of the models showed significant effects of any of the neighborhood measures (ND and NF; $p > 0.05$). In all the four models, there was significant facilitative effect of HD ($\beta = 0.45$, $z = 6.06$, $p < 0.001$), and HF ($\beta = 0.27$, $z = 10.12$, $p < 0.001$) on accuracy. In other words, syllables with many homophones and higher frequency were responded more accurately than syllables with fewer homophones and lower frequency. Also, the interactions between HD and HF ($p = 0.7$) had no significant effect on accuracy.

Table 8: Summary of fixed effects in the accuracy models for real-syllables across all neighborhood schemes. The model contains15508 trials from 598 item types.

Model with neighborhood measures based on the **one-segment/tone difference scheme**				
AIC = 9048.9				
Fixed effects				
Predictor	β	Standard Error	z-value	p-value
Intercept	2.733	0.13883	19.686	<2.00E-16
log(HD)	0.45088	0.07432	6.067	1.30E-09
log(HF)	0.27135	0.0268	10.126	<2.00E-16
log(StimDur)	-0.48844	0.46884	-1.042	0.297
Tone = T2	0.02834	0.15953	0.178	0.859
Tone = T3	0.17585	0.16935	1.038	0.299
Tone = T4	0.07533	0.16276	0.463	0.643
Model with neighborhood measures based on the **one-segment difference scheme**				
AIC = 9048.9				
Fixed effects				
Predictor	β	Standard Error	z-value	p-value
Intercept	2.733	0.13883	19.686	<2.00E-16
log(HD)	0.45088	0.07432	6.067	1.30E-09
log(HF)	0.27135	0.0268	10.126	<2.00E-16
log(StimDur)	-0.48844	0.46884	-1.042	0.297
Tone = T2	0.02834	0.15953	0.178	0.859

Tone = T3	0.17585	0.16935	1.038	0.299
Tone = T4	0.07533	0.16276	0.463	0.643
Model with neighborhood measures based on the **one-component /tone difference scheme**				
AIC = 9048.9				
Fixed effects				
Predictor	β	Standard Error	z-value	p-value
Intercept	2.733	0.13883	19.686	<2.00E-16
log(HD)	0.45088	0.07432	6.067	1.30E-09
log(HF)	0.27135	0.0268	10.126	<2.00E-16
log(StimDur)	-0.48844	0.46884	-1.042	0.297
Tone = T2	0.02834	0.15953	0.178	0.859
Tone = T3	0.17585	0.16935	1.038	0.299
Tone = T4	0.07533	0.16276	0.463	0.643
Model with neighborhood measures based on the **one-component difference scheme**				
AIC = 9048.9				
Fixed effects				
Predictor	β	Standard Error	z-value	p-value
(Intercept)	2.733	0.13883	19.686	<2.00E-16
log(HD)	0.45088	0.07432	6.067	1.30E-09
log(HF)	0.27135	0.0268	10.126	<2.00E-16

log(StimDur)	-0.48844	0.46884	-1.042	0.297
Tone = T2	0.02834	0.15953	0.178	0.859
Tone = T3	0.17585	0.16935	1.038	0.299
Tone = T4	0.07533	0.16276	0.463	0.643

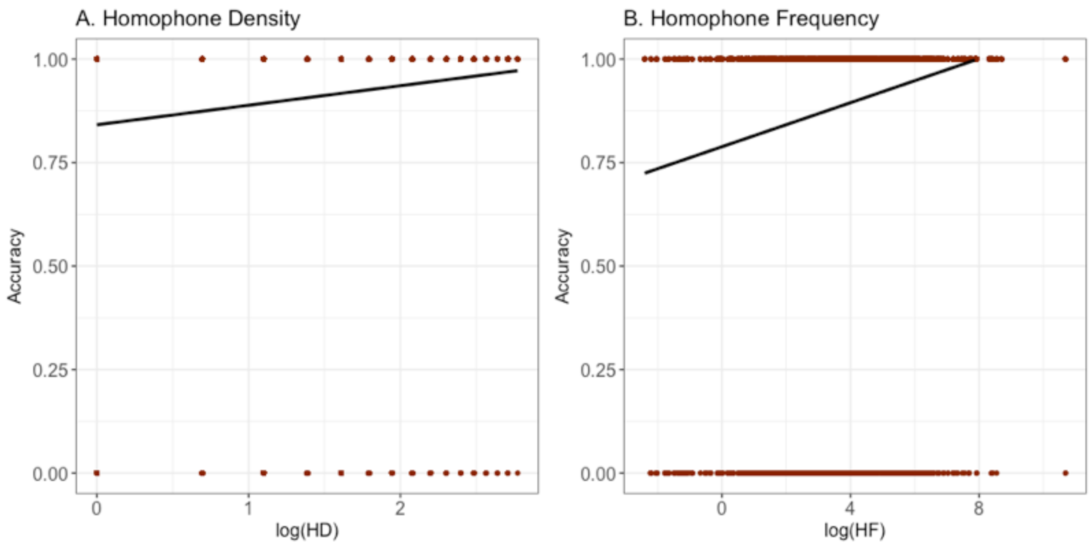

Figure 16: A. Scatterplot of log(HD) across accuracy; B. Scatterplot of log(HF) across accuracy.

Overall, the model on accuracy of real monosyllables showed significant effects of homophone measures (HD and HF) alone. Figure 16 shows scatterplot of log (HD) and log(HF) across accuracy. No significant effects of neighborhood measures (ND and NF) on accuracy of real monosyllables were observed. These results lend support to hypothesis 1B, which predicted that neighborhood effects and homophone effects based on shared representation model can be different. Having null effects for neighborhood measures and facilitatory effects for homophone measures shows that they do not have similar effects on spoken word recognition on accuracy of real monosyllables.

4.3.1.2 RT analysis

Analysis for RT was only conducted on correctly identified trials. The dataset for the models on RT consisted of 13797 trials from 598 item types. Four separate models were constructed, each representing neighborhood measures from a different scheme. All the initial models follow the formula (9) from Chapter 3, repeated below), although the neighborhood measures were calculated using different schemes.

(9) $\log(RT) \sim ND + \log(NF) + \log(HD) + \log(HF) + ND{:}\log(HF) + \log(HD){:}\log(HF) + \log(StimDur) + Tone + (1|Subject) + (1|Item)$

All four initial models underwent a backward elimination procedure (as described in Chapter 3, Section 3.3.2 Model construction) so that only significant predictors remain. Only results from the final models are reported here. Table 9 summarize the model fit (AIC) and all fixed effects in all four final models. Again, AIC values were used to decide the best neighborhood scheme. As shown in Table 9, the model with one-component/tone difference scheme has the lowest AIC (-5566.5) among all four models, and its AIC is lower than the second-best model, i.e. the model with one-segment/tone difference (AIC = -5562.0), by 4.5. According to Anderson and Burnham (2004) a difference of 2 in AIC indicates that the model with higher AIC is 0.368 times probable as the model with the lower AIC. Therefore, the model with one-component/tone difference scheme is significantly better than all the other three models.

Table 9: Summary of fixed effects in the RT models for real-syllables across all neighborhood schemes. The model contains 13797 trials from 598 item types.

Model with neighborhood measures based on the **one-segment/tone difference scheme**			
AIC = -5562.0			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.822379	0.011585	588.9
log(NF_SegT)	-0.015851	0.005213	-3
log(HD)	-0.009374	0.004402	-2.1
log(HF)	-0.013123	0.001898	-6.9
log(StimDur)	0.547436	0.027759	19.7
Tone = T2	0.008816	0.009642	0.9
Tone = T3	-0.028263	0.010263	-2.8
Tone = T4	0.061349	0.009985	6.1
HD:HF	0.004979	0.002376	2.1
Model with neighborhood measures based on the **one-segment difference scheme**			
AIC = -5559.8			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.825523	0.011541	591.4
log(NF_Seg)	-0.021313	0.008018	-2.7
log(HD)	-0.009647	0.004412	-2.2

log(HF)	-0.013279	0.001903	-7
log(StimDur)	0.548953	0.02782	19.7
Tone = T2	0.006028	0.009619	0.6
Tone = T3	-0.030816	0.010232	-3
Tone = T4	0.053993	0.009722	5.6
HD:HF	0.004664	0.002383	2
Model with neighborhood measures based on the **one-component /tone difference scheme**			
AIC = -5566.5			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.820454	0.011602	587.9
log(NF_CompT)	-0.021012	0.00565	-3.7
log(HD)	-0.007373	0.004458	-1.7
log(HF)	-0.013178	0.001891	-7
log(StimDur)	0.54332	0.02772	19.6
Tone = T2	0.0093	0.009603	1
Tone = T3	-0.024705	0.010339	-2.4
Tone = T4	0.064312	0.010044	6.4
HD:HF	0.005077	0.002368	2.1
Model with neighborhood measures based on the **one-component difference scheme**			
AIC = -5559.9			
Fixed effects			

Predictor	β	Standard Error	t-value
Intercept	6.825897	0.011549	591
log(NF_Comp)	-0.0209	0.007806	-2.7
log(HD)	-0.008512	0.004474	-1.9
log(HF)	-0.012832	0.00191	-6.7
log(StimDur)	0.538166	0.028306	19
Tone = T2	0.004499	0.009631	0.5
Tone = T3	-0.028782	0.010287	-2.8
Tone = T4	0.051825	0.009757	5.3
HD:HF	0.004812	0.002382	2

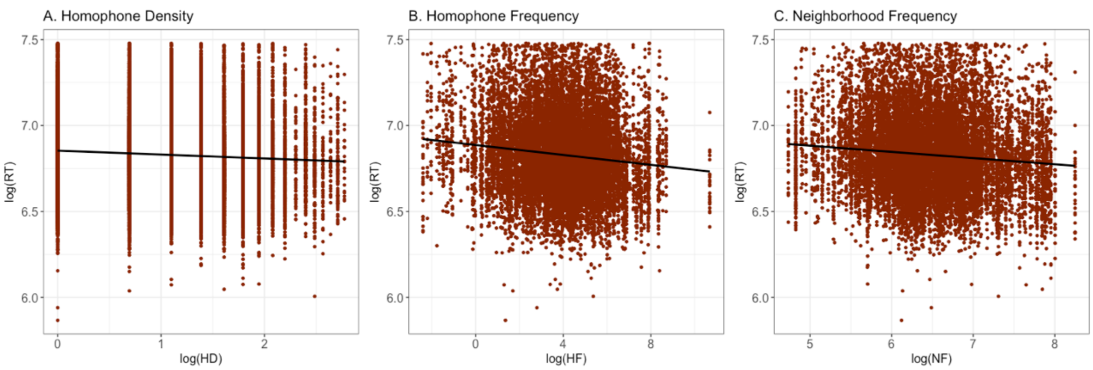

Figure 17: A. Scatterplot of log(HD) across RT; B. Scatterplot of log(HF) across RT; C. Scatterplot of log(NF) across RT.

The model with one-component/tone difference scheme showed significant facilitative effect NF (β = -0.02, t = -3.7), HD (β = -0.007, t = -1.7), and HF (β = -0.01, t = -7.0). In other words, syllables with high neighborhood frequency, were responded faster compared to syllables with low neighborhood frequency. Also, syllables with more homophones and higher frequency were responded faster than syllables with fewer homophones and lower frequency. Figure 17 shows scatterplot of log (HD), log(HF), and log(NF) across RT. In addition, there was a significant negative interaction between HD and HF (β = 0.005, t = 2.1). That is to say, a syllable with high homophone frequency which is associated with many homophones, is responded slower compared to syllable with high homophone frequency with fewer homophones. Figure 18 depicts the interaction between HD and HF on RT. There was no significant effect of ND on RT (p = 0.2).

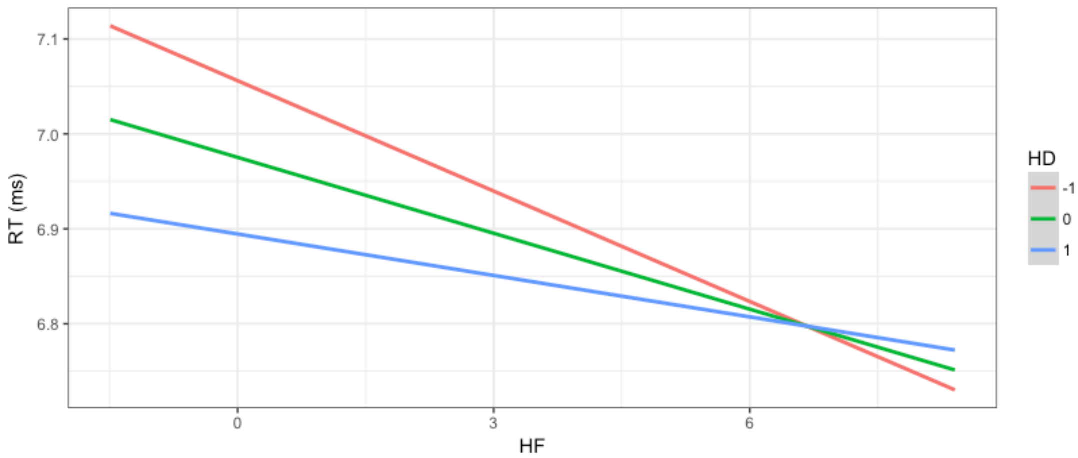

Figure 18: Interaction between log(HD) and log(HF) on RT.

In order to further investigate the results of interaction, the data were split into high-frequency dataset and low-frequency dataset using quantile-split. High frequency dataset comprised of data points from top 25% of the HF while low-frequency dataset comprised of data points from bottom 25% of the HF. The effect of HD was examined for high-frequency dataset and low-frequency dataset using separate mixed effect models. The model on low-frequency dataset showed significant facilitative effect (β = -0.02, t = -2.9) of HD. In other words, syllables with low HF but many homophone mates were responded to faster as compared to syllables with fewer homophone mates. Table 10 shows model results from low-frequency dataset.

The model on high-frequency dataset showed no significant effect of HD (β = 0.01, t = 1.1) but a trend towards inhibitory effects. In other words, in high-HF syllables, syllable with many homophone mates were responded slower to as compared to syllables with fewer homophone mates. Table 11 shows model results from high-frequency dataset.

Table 10: Summary of fixed effects in the RT models for real-syllables with low HF. The model contains 3449 trials from 169 item types.

Predictor	β	Standard Error	t-value
Intercept	6.86535	0.01603	428.2
log(HD)	-0.029	0.01004	-2.9
log(NF_CompT)	-0.01592	0.01156	-1.4
log(StimDur)	0.52062	0.0595	8.7
Tone = T2	-0.01201	0.0203	-0.6
Tone = T3	-0.01658	0.02126	-0.8
Tone = T4	0.07442	0.02023	3.7

Table 11: Summary of fixed effects in the RT models for real-syllables with high HF. The model contains 3453 trials from 143 item types.

Predictor	β	Standard Error	t-value
Intercept	6.785101	0.017406	389.8
log(HD)	0.010567	0.009472	1.1
log(NF_CompT)	-0.049187	0.012296	-4
log(StimDur)	0.551247	0.053783	10.2

Tone = T2	0.024827	0.021153	1.2
Tone = T3	-0.01722	0.02119	-0.8
Tone = T4	0.075578	0.022271	3.4

In addition to the critical effects, the model also shows significant effects from control variables: syllables with longer stimulus duration (StimDur) were responded slower (β = 0.54, t = 19.6); T3 syllables were responded faster than T2 and T4 (β = -0.02, t = -2.4), which could be attributed to the dipping tonal trajectory of T3 syllables that could have resulted in early recognition.

It should be noted that although the model with one-component/tone difference scheme produced the best model fit, the results of this model are highly similar to those from the other three models. Overall, the other three models also showed (1) significant facilitatory effects of NF (in the model with one-segment/tone difference scheme: β = -0.015, t = -3; one-segment difference scheme: β = -0.021, t = -2.7; one-component difference scheme: β = -0.020, t = -2.7), (2) significant facilitatory effects of HD (in the model with one-segment/tone difference: β = -0.009, t = -2.1; one-segment difference scheme: β = -0.009, t = -2.2; one-component difference scheme: β = -0.008, t = -1.9), (2) significant facilitatory effects of HF (in the model with one-segment/tone difference: β = -0.01, t = -6.9; one-segment difference scheme: β = -0.01, t = -7.0; one-component difference scheme: β = -0.01, t = -6.7), (4) significant inhibitory interaction of HD and HF (in the model with one-segment/tone difference: β = 0.004, t = 2.1; one-segment difference scheme: β = 0.004, t = 2.0; one-component difference scheme: β = 0.004, t = 2.0). In addition, all the alternative models show the same effects regarding control variables (StimDur, Tone) as the best-fitting model.

Overall, the model on RT of real monosyllables provide support to both hypothesis 1A and 1B. According to hypothesis 1A, both phonological neighbors and homophone mates should have similar effects on spoken word recognition. According to models on RT for real monosyllables, it can be observed that both neighborhood measures (NF) and homophone measures (HD and HF) showed significant facilitatory effects on spoken word recognition. Having similar facilitatory effects for both phonological neighbors and homophone mates lends evidence for hypothesis 1A. However, the model on RT also reported significant interaction between HD and HF. But no significant interactions were observed for ND and HF. This in turn provides an evidence for hypothesis 1B that suggests that the interaction between HD and frequency, and ND and frequency can be different. Overall, the model results on RT provides some evidence to hypothesis 1A and some evidence to hypothesis 1B.

4.3.2 Pseudo-syllables

The overall mean RT of pseudo-syllables was 1087.95 ms (*SD* = 447.95 ms) and mean accuracy rate was 86%. The mean RT of tonal gaps was 1142.97 ms (*SD* = 474.32) and mean accuracy rate was 84%. The mean RT of segmental gaps was 1011.89 ms (*SD* = 396.37 ms) and mean accuracy rate was 88%. As compared to segmental gaps, tonal gaps demonstrated slower RT and lower accuracy rates. This could be because tonal gaps are pseudo-syllables that exist with other lexical tones, which makes them difficult to recognize in a time-bound task. Therefore, participants took longer to recognize tonal gaps and the accuracy rates were also low. Segmental gaps are pseudo-syllables that do not exist with any lexical tone. This makes segmental gaps easier to recognize in comparison to tonal gaps resulting in lower RT and higher accuracy rates.

Before conducting further analysis, data were trimmed for outliers. Similar to the analysis of real monosyllables, two types of outliers were removed: (1) items with accuracy rate less than 20%, and (2) trials that were 2 SD above and below the mean RT. Altogether, 6.1 % of trials were excluded.

As mentioned in Section 2.5 (Current study), here is the hypotheses for pseudo-syllables.

(1) Hypothesis 2:

Following from Luce and Pisoni (1998), I predict inhibitory effects of phonological neighbors on the identification of pseudo-syllables. In other words, pseudo-syllables with many similar-sounding real monosyllables and high neighborhood frequency will be recognized as pseudo-syllables slower and less accurately than pseudo-syllables with fewer similar-sounding real monosyllables and low neighborhood frequency.

4.3.2.1 Accuracy analysis

The dataset for the models on accuracy consisted of 14881 trials of 358 item types. Again, all four-neighborhood schemes were tested in separate models. All the initial models follow the same formula (13) from Chapter 3, repeated below), although the neighborhood measures were calculated using different schemes.

(13) Accuracy ~ ND + log(NF) + ItemType + log(StimDur) + Tone + (1|Subject) + (1|Item)

All four initial models underwent a backward elimination procedure. Table 12 summarizes the model fit (AIC) and all fixed effects in all four final models. Based on the AIC values mentioned in Table 12, neighborhood measures based on one-segment/tone difference scheme best predicted the accuracy scores (AIC = 9587.6). The second-best model was with neighborhood measures from one-segment difference scheme (AIC = 9673.1) whose AIC differs from the best model by 85.5. Therefore, the model with one-segment/tone difference scheme is significantly better than all the other three models.

Table 12: Summary of fixed effects in the accuracy models for pseudo-syllables across all neighborhood schemes. The model contains14881 trials from 358 item types.

Model with neighborhood measures based on the **one-segment /tone difference scheme**				
AIC = 9587.6				
Fixed effects				
Predictor	β	Standard Error	z-value	p-value
Intercept	2.67933	0.21255	12.606	<2.00E-16
ND_SegT	-0.07194	0.01974	-3.644	0.000269

ItemType = Tonal gap	-0.48238	0.18291	-2.637	0.008356
log(StimDur)	1.92038	0.65619	2.927	0.003427
Tone = T2	0.65052	0.22101	2.943	0.003246
Tone = T3	-0.16032	0.25887	-0.619	0.53571
Tone = T4	0.25991	0.25297	1.027	0.304217
Model with neighborhood measures based on the **one-segment difference scheme**				
AIC = 9673.1				
Fixed effects				
Predictor	β	Standard Error	z-value	p-value
Intercept	2.74686	0.2133	12.878	<2.00E-16
ND_Seg	-0.03801	0.01533	-2.48	0.01314
log(NF_Seg)	0.32989	0.14541	2.269	0.023283
ItemType = Tonal gap	-0.67546	0.17669	-3.823	0.000132
log(StimDur)	2.09817	0.6616	3.171	0.001517
Tone = T2	0.81887	0.21891	3.741	0.000184
Tone = T3	-0.15152	0.25847	-0.586	0.557725
Tone = T4	0.22508	0.25171	0.894	0.371213
Model with neighborhood measures based on the **one-component /tone difference scheme**				
AIC = 9978.6				
Fixed effects				

Predictor	β	Standard Error	z-value	p-value
Intercept	2.7582	0.20051	13.756	<2.00E-16
ND_CompT	-0.07784	0.01452	-5.359	8.36E-08
ItemType = Tonal gap	-0.57144	0.16362	-3.492	0.000479
log(StimDur)	1.13258	0.63684	1.778	0.075329
Tone = T2	0.47737	0.21195	2.252	0.024308
Tone = T3	0.06604	0.24799	0.266	0.789999
Tone = T4	0.29776	0.24137	1.234	0.21734
Model with neighborhood measures based on the **one-component difference scheme**				
AIC = 9995.2				
Fixed effects				
Predictor	β	Standard Error	z-value	p-value
Intercept	2.79551	0.20714	13.496	<2.00E-16
ND_Comp	-0.03921	0.01345	-2.916	0.003546
log(NF_Comp)	0.42077	0.17007	2.474	0.013357
ItemType = Tonal gap	-0.67425	0.17577	-3.836	0.000125
log(StimDur)	1.79478	0.69062	2.599	0.009355
Tone = T2	0.71303	0.21269	3.352	0.000801
Tone = T3	-0.09844	0.2559	-0.385	0.700473
Tone = T4	0.16552	0.24858	0.666	0.505518

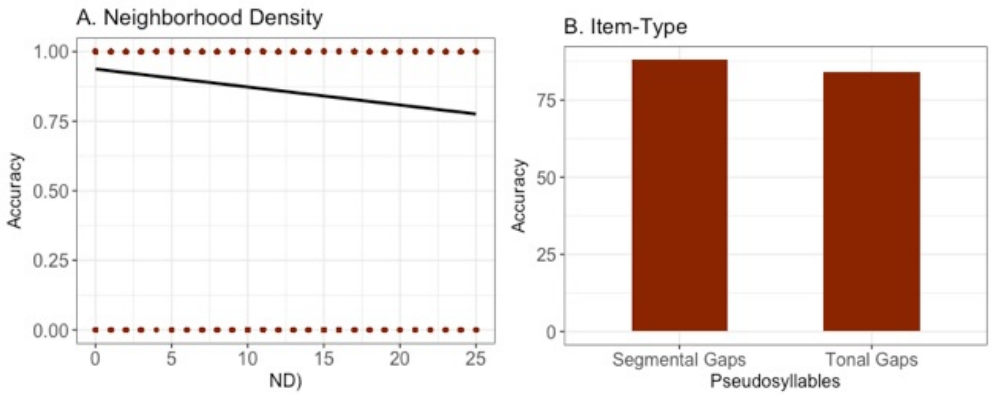

Figure 19: A. Scatterplot of ND across accuracy; B. Bar graph depicting accuracy of pseudosyllables.

In the model with one-segment/tone difference scheme, there was significant inhibitory effect of ND on accuracy (β = -0.07, z = -3.64, $p < 0.001$). In other words, pseudo-syllables from dense neighborhood were responded less accurately compared to pseudo-syllables from sparse neighborhood (see figure 19A for scatterplot of ND across accuracy). In addition, ItemType emerged as a significant predictor (β = -0.48, z = -2.63, $p < 0.01$) (See figure 19B). In other words, segmental gaps were responded more accurately than tonal gaps. NF did not appear as a significant predictor of accuracy (p = 0.5). In addition to the critical effects, the model also shows significant effects from control variables: syllables with longer stimulus duration were responded more accurately (β = 1.92, z = 2.92, $p < 0.01$); T2 syllables were responded more accurately than T3 and T4 (β = 0.65, z = 2.94, $p < 0.01$).

It should be noted that although the model with one-segment/tone difference scheme produced the best model fit, the results of this model are largely similar to those from the other three models. Overall, the other three models also showed (1) significant inhibitory effects of ND (one-segment difference scheme: $\beta = -0.03$, $z = -2.48$, $p = 0.01$; in the model with one-component/tone difference: $\beta = -0.07$, $z = -5.3$, $p < 0.001$; one-component difference scheme: $\beta = -0.03$, $z = -2.9$, $p = 0.003$), (2) significant facilitatory effects of NF on two out of three models (one-segment difference scheme: $\beta = 0.32$, $z = 2.26$, $p = 0.02$; one-component difference scheme: $\beta = 0.42$, $z = 2.47$, $p = 0.01$), (3) significant effects of ItemType (one-segment difference scheme: $\beta = -0.67$, $z = -3.82$, $p < 0.001$; in the model with one-component/tone difference: $\beta = -0.57$, $z = -3.49$, $p < 0.001$; one-component difference scheme: $\beta = -0.67$, $z = -3.83$, $p < 0.001$). In addition, all the alternative models show the same effects regarding control variables (StimDur, Tone) as the best-fitting model.

Overall, the model on accuracy of pseudo-syllables showed significant inhibitory effect of ND on spoken word recognition. This shows support to hypothesis 2, which predicts inhibitory effects for neighborhood measures.

4.3.2.2 RT analysis

Analysis of RT was only conducted for correctly identified trials. The models on RT were built on a dataset consisting of 12829 trials of 358 item types. All the four neighborhood measures were tested in separate models. Again, all the initial models follow the formula (12) from Chapter 3, repeated below), although the neighborhood measures were calculated using different schemes.

(12) log(RT) ~ ND +log(NF) + ItemType + log(StimDur) + Tone + (1|Subject) + (1|Item)

All four initial models underwent a backward elimination procedure. Table 13 summarizes the model fit (AIC) and all fixed effects in all four final models. Based on the AIC values mentioned in Table 13, neighborhood measures based on one-component/tone difference scheme best predicted the RT (AIC = -655.9). The second-best model was with neighborhood measures from one-component difference scheme (AIC = -627.9) whose AIC differs from the best model by 28.0. Therefore, the model with one-component/tone difference scheme is significantly better than all the other three models.

Table 13: Summary of fixed effects in the RT models for pseudo-syllables across all neighborhood schemes. The model contains12829 trials from 358 item types.

Model with neighborhood measures based on the **one-segment /tone difference scheme**			
AIC = -510.7			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.859919	0.016243	422.3
ND_SegT	0.00518	0.001014	5.1
log(NF_SegT)	-0.010451	0.004894	-2.1
ItemType = Tonal gap	0.055163	0.009437	5.8
log(StimDur)	0.325409	0.034354	9.5
Tone = T2	-0.007246	0.011314	-0.6

Tone = T3	0.001256	0.013424	0.1
Tone = T4	0.051952	0.013266	3.9
Model with neighborhood measures based on the **one-segment difference scheme**			
AIC = -473.4			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.8610346	0.0162784	421.5
ND_Seg	0.0030352	0.0007959	3.8
ItemType = Tonal gap	0.0603581	0.0089927	6.7
log(StimDur)	0.3366353	0.0346122	9.7
Tone = T2	-0.0171573	0.0114397	-1.5
Tone = T3	-0.0037205	0.013663	-0.3
Tone = T4	0.0532947	0.0134399	4
Model with neighborhood measures based on the **one-component /tone difference scheme**			
AIC = -655.9			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.8454032	0.0159609	428.9
ND_CompT	0.0052645	0.0007768	6.8
log(NF_CompT)	-0.0108241	0.00507	-2.1
ItemType = Tonal	0.0677787	0.0086291	7.9

gap			
log(StimDur)	0.3859443	0.0343656	11.2
Tone = T2	0.0017367	0.011334	0.2
Tone = T3	-0.0169294	0.0133855	-1.3
Tone = T4	0.0467728	0.0131231	3.6
Model with neighborhood measures based on the **one-component difference scheme**			
AIC = -627.9			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.8480849	0.0162029	422.6
ND_Comp	0.0025396	0.0007183	3.5
log(NF_Comp)	-0.023571	0.0093074	-2.5
ItemType = Tonal gap	0.0724307	0.0094269	7.7
log(StimDur)	0.3646805	0.0370731	9.8
Tone = T2	-0.0167993	0.0114636	-1.5
Tone = T3	-0.0116205	0.0138473	-0.8
Tone = T4	0.0538906	0.0136271	4

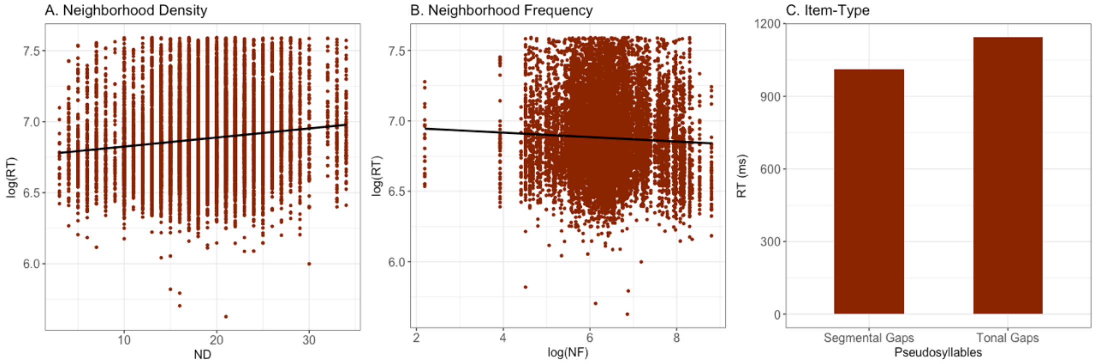

Figure 20: A. Scatterplot of ND across RT; B. Scatterplot of log(NF) across RT; C. Bar graph depicting RT of pseudosyllables.

In the model with one-component/tone difference scheme, there was significant inhibitory effect of ND on RT ($\beta = 0.005$, $t = 6.8$) (see figure 20A). In other words, pseudo-syllables from dense neighborhood were responded slowly compared to pseudo-syllables from sparse neighborhood. In addition, significant facilitative effect of NF ($\beta = -0.01$, $t = -2.1$) on RT was also observed (see figure 20B). Pseudo-syllables with high neighborhood frequency were responded faster compared to pseudo-syllables with low neighborhood frequency. Also, ItemType emerged as significant predictors (($\beta = 0.06$, $t = 7.9$) (see figure 20C). In other words, segmental gaps were responded faster than tonal gaps. Apart from critical effects, the model also shows significant effects from control variables: syllables with shorter stimulus duration were responded faster ($\beta = 0.38$, $t = 11.2$); T4 syllables were responded more slowly than T3 and T4 ($\beta = 0.04$, $t = 3.6$).

Overall, the model on RT of pseudo-syllables supports the prediction that ND has an inhibitory effect on RT. However, the model does not support the prediction for NF.

It should be noted that although the model with one-component/tone difference scheme produced the best model fit, the results of this model are largely similar to those from the other three models. Overall, the other three models also showed (1) significant inhibitory effects of ND (in the model with one-segment/tone difference scheme: $\beta = 0.005$, $t = 5.1$; one-segment difference scheme: $\beta = 0.003$, $t = 3.8$; one-component difference scheme: $\beta = 0.002$, $t = 3.5$), (2) significant facilitative effects of NF on two out of three models (in the model with one-segment/tone difference scheme: $\beta = -0.01$, $t = -2.1$; one-component difference scheme: $\beta = -0.02$, $t = -2.5$), (3) significant effects of ItemType (in the model with one-segment/tone difference scheme: $\beta = 0.05$, $t = 5.8$; one-segment difference scheme: $\beta = 0.06$, $t = 6.7$; one-component difference scheme: $\beta = 0.07$, $t = 7.7$). In addition, all the alternative models show the same effects regarding control variables (StimDur, Tone) as the best-fitting model.

Overall, the model on RT of pseudo-syllables showed significant inhibitory effect of ND but also facilitatory effect of NF on spoken word recognition. Specifically, the inhibitory effects of ND support hypothesis 2, but the facilitative effects of NF do not. It is unclear why there is faciltitative effect of NF, I will discuss more in Chapter 6.

4.4 Summary

The lexical decision task was aimed at investigating the effects of phonological neighbors and homophones on spoken word recognition. From the results, it can be inferred that phonological neighbors as well as homophones predict participants' performance in an auditory lexical decision task.

To summarize results from real monosyllables, both neighborhood measures and homophone measures significantly affect processing of spoken words. Among neighborhood measures, ND had no significant effect on RT and accuracy. However, NF showed a significant facilitative effect on RT alone. Syllables with higher NF were responded faster as compared to syllables with lower NF. Among homophone measures, HD and HF showed significant effects on RT and accuracy of lexical decision task. Both HD and HF exhibited significant facilitative effect on spoken word recognition. In other words, syllables with more homophone and higher homophone frequency were responded faster and more accurately compared to syllables with less homophones and lower homophone frequency. In addition to fixed effects, HD and HF showed significant interactions on RT alone. Interestingly, the interactions were in opposite direction of the fixed effects of HD and HF. Based on fixed effects, if a syllable has high HD, RT decreases. Also, when a syllable is associated with a high HF, RT reduces. However, there is a bit of adjustment because of the interactions between HD and HF. If a spoken syllable with high HF is also associated with many homophones, then the RT gets longer. In general, neighborhood measures mainly affect the speed while homophones affect speed and accuracy of processing real monosyllables.

Pseudo-syllables were also analyzed to investigate the effects neighborhood measures and pseudo-syllable type (segmental gap or tonal gap). Among neighborhood measures, ND showed significant inhibitory effects on RT and accuracy. However, significant effect of NF was observed in RT model alone where NF showed a facilitatory effect on RT. Pseudo-syllables with more neighbors face more competition due to the activation of a greater number of neighbors resulting in longer RT and lower accuracy scores compared to pseudo-syllables with lesser number of neighbors. However, pseudo-syllables with high NF result in shorter RT as compared to pseudo-syllables with low NF. ItemType also emerged as a significant predictor for both RT and accuracy models. As expected, segmental gaps were responded significantly faster compared to tonal gaps. Overall, the inhibitory effects of ND are stronger as it is seen in both models on accuracy and RT. But the facilitation of NF is only seen in model on RT.

Neighborhood measures based on one-component/tone difference scheme best predicts the RT of real monosyllables. However, for pseudo-syllables, neighborhood measures based on one-component/tone difference scheme best predicts the RT while neighborhood measures based on one-segment/tone difference scheme best predicts the accuracy of pseudo-syllables. Even though the neighborhood scheme for accuracy of pseudo-syllables is different but the results from neighborhood scheme based on one-segment/tone difference is similar to the results from neighborhood scheme based on one-component/tone difference. In any case, the result emphasizes on the importance of tone in spoken word recognition.

Chapter 5: Auditory naming

This chapter brings forth the second experimental study of this dissertation. The aim of this experiment was to confirm the results obtained with the auditory lexical decision experiment (Chapter 4: Auditory lexical decision). This task is different from auditory lexical decision task as it examines the phenomenon of spoken word recognition from a different perspective. While auditory lexical decision task involves more complicated processing, including distinction between words and nonwords, auditory naming is a less complicated task that involves only words. In addition, auditory lexical decision task may be more affected by frequency while auditory naming task is less affect by frequency (Balota, 1980).

In an auditory naming task, participants hear real monosyllables and repeat them as quickly as possible. In this experiment, two tasks were conducted: (1) Instantaneous naming, and (2) Delayed naming. As the name suggests, in the instantaneous naming task, the participants repeat the stimuli as soon as possible from the time they hear the stimuli. In comparison, in the delayed naming, a delay (or a pause) is provided after stimulus presentation for the participants to prepare their response. The participants are prompted to repeat the stimulus after the delay, as quickly as possible.

The purpose of using a delayed naming task was to control for possible effects of articulatory planning on auditory naming performance. In an instantaneous naming task, the participants are doing both perception and production. Therefore, the reaction time (RT i.e. the time between hearing the onset of the stimulus to the moment of beginning the articulation) includes the time for both perceiving the stimulus and preparing for the articulation. By contrast, in a delayed naming task, the RT (i.e. the time between the onset of prompt to the onset of beginning the articulation) only includes the time for preparing the articulation. By using both instantaneous naming and delayed naming, we can control for possible effects of articulatory effect. The focus of the current study is on lexical effects on spoken word recognition. So, if there are lexical effects (neighborhood effects or homophone effects) on spoken word perception, they should be evident in instantaneous naming but not delayed naming.

In the current experiment, delay of 600 ms and 1200 ms were used. These delays were chosen based on the experimental results of the auditory lexical decision task (see Chapter 4). The mean RT for the auditory lexical decision task was 1023.22 ms (SD = 373.58 ms). To ensure that the lexical processing has taken place, a delay of 1200 ms was used. Another delay was 600 ms, which is half the time of 1200 ms, was also included so that participants cannot anticipate the delay.

5.1 Method

5.1.1 Participants

One hundred thirty Mandarin speakers, born and raised in mainland China, participated in this study. None of the participants reported any speech and hearing problems. Written informed consents were obtained from all the participants prior to the experiment. A group of 130 participants, separated from those who participated in the auditory lexical decision experiment, participated in the auditory naming experiment. Specifically, 65 (35M, 30F, mean age 20.65; SD= 2.78) subjects participated in the instantaneous naming task, and the other 65 (13M, 52F, mean age = 23.46 years; SD = 4.16) participated in the delayed naming task.

5.1.2 Stimuli

1259 real monosyllables were used in the experiment. See Chapter 3, Section 3.1 (Stimuli and recording) for details.

5.1.3 Design and procedure

It is impossible to include all the stimuli (n= 1259) in one experimental session without making the experiment too demanding for the participant. Therefore, the stimuli were divided into 5 sub-lists, each with 256 unique stimulus items. Each stimulus item was presented twice in a sub-list, resulting in a total of 512 stimulus tokens per sub-list, with a completely randomized order of presentation. Each participant worked on only one sub-list, and each sub-list was presented to 13 participants. To evaluate cross group differences, 5 stimulus items were shared by all sub-list; apart from that, all other real monosyllables appeared in one sub-list.

The experiment was designed and conducted using Opensesame version 3.1 (Mathôt, Schreij, & Theeuwes, 2011) on a Philips desktop connected to M-audio interface. A multi-channel recorder TASCAM DR-44WL was connected to M-audio interface. Channel 1 of the multi-channel recorder was connected to the headphones through which the stimulus was presented as well as recorded for later analysis. Channel 2 was internally connected to microphone. Channel 2 recorded participants responses. The experiment was conducted in a quiet room. Each participant was seated comfortably before beginning the experiment. Each experiment session consisted of 15 practice trials followed by a block of test items presented in a random order. None of the practice items were repeated in the main experiment. No feedback was provided in this experiment. Figure 21 shows a diagram of the experimental setup.

5.1.3.1 Instantaneous naming task

In an instantaneous naming task, 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 quickly and accurately as possible. After participants' response, the experiment proceeded to next trial with a button press. Each session lasted for not more than 30 minutes for all the participants, including one mandatory break in between. Figure 22 shows a diagram of the instantaneous naming task trial.

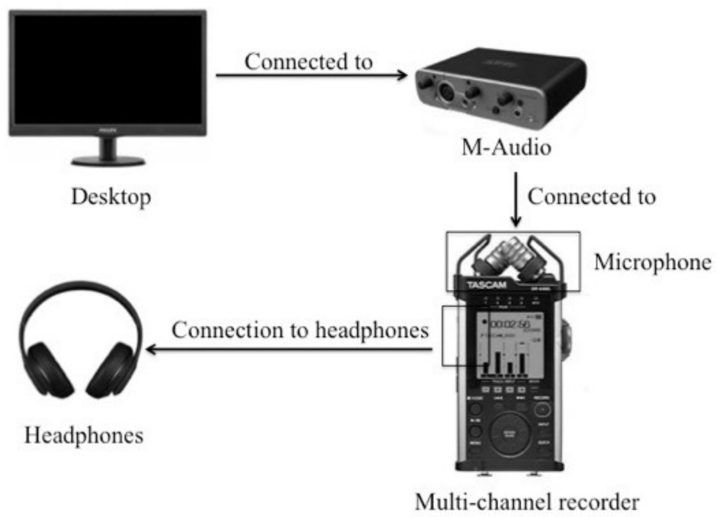

Figure 21: Experimental set up for auditory naming experiment.

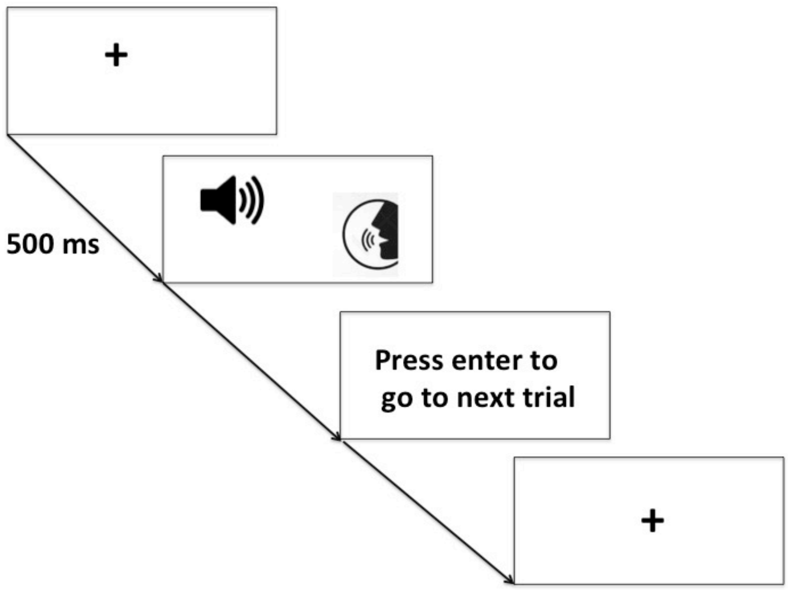

Figure 22: Diagram of the instantaneous naming task trial.

5.1.3.2 Delayed naming task

In the delayed naming task, each trial started with a button-press followed by a fixation-cross for 500 ms followed by the auditory stimulus. After the stimulus appeared, there was a delay of 600 ms or 1200 ms. After the delay, a 1000 Hz pure-tone was presented for 250 ms to alert the participants to respond. After participants responded, the experiment went to next trial with a button press. The participants were instructed that they will hear Mandarin monosyllables. After the stimulus, there was a delay (or pause) so that they can prepare the heard stimulus. They were asked to repeat the stimuli as quickly and accurately as possible, as soon as they heard a pure-tone. Each item was presented with a delay of 600 ms and 1200 ms, randomly. Each session lasted for not more than 40 minutes, including a mandatory break in between. Figure 23 shows a diagram of the delayed naming task trial.

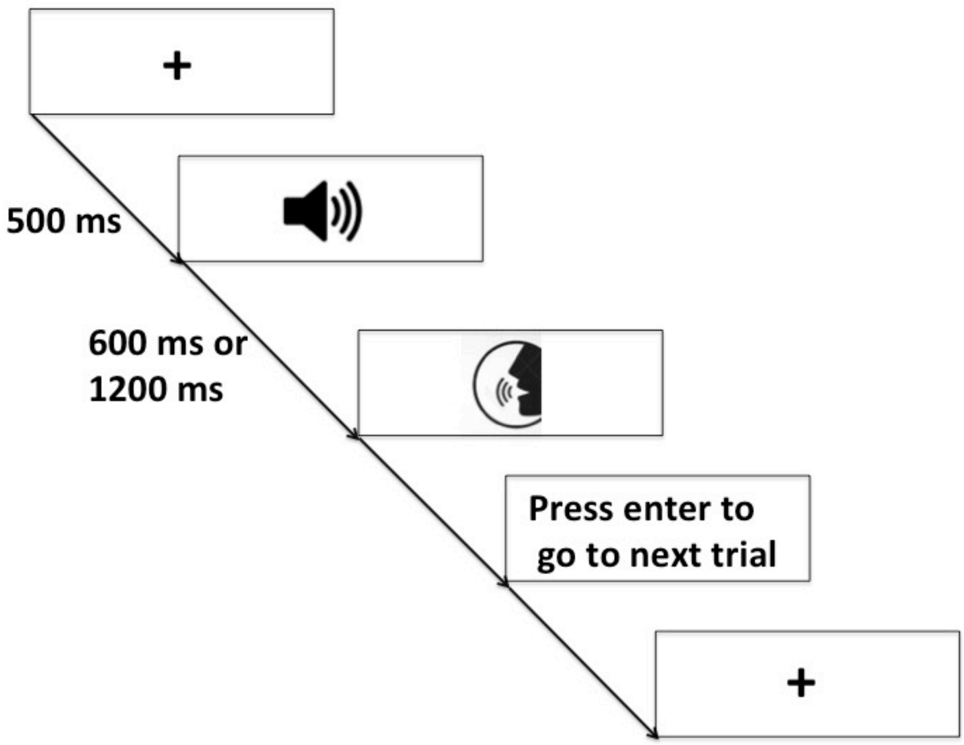

Figure 23: Diagram of delayed naming task trial.

5.2 Analysis

Data extracted from each participant contained 2 audio files: (1) Stimuli from Channel 1, and (2) Participant's production from Channel 2. The two files were combined as a stereo recording using audacity freeware (Mazzoni & Dannenberg, 2000). Further, Praat scripts were used to put onset boundaries at the beginning of the stimulus and at the onset of participants' production. Each file was manually corrected for any errors in placing onset boundaries. Once all the files were corrected, another Praat script was used to extract the duration between the onset of the stimuli and the onset of the participants' production. This duration was referred to as RT (ms) in instantaneous naming task. In case of delayed naming, the RT was measured from the onset of 1000 Hz pure-tone to the onset of participants' production. For accuracy, 5 audio files were randomly selected from the naming experiment to calculate accuracy scores. As expected, the mean accuracy score was at ceiling (~99%). As a result, accuracy data were not considered for further analyses and statistical models were only constructed with RT as the outcome variable.

General modeling methods as described in Chapter 3, Section 3.3 were used. Separate mixed-effect models were built with RT as the outcome variable for instantaneous naming and delayed naming task. In both instantaneous naming and delayed naming task, neighborhood density (ND), neighborhood frequency (NF), homophone density (HD), homophone frequency (HF), interaction between HD and HF, and interaction between ND and HF were critical fixed-effects predictors in models. In addition, the models had stimulus duration (StimDur) and Tone as control fixed-effects predictors as well as intercepts for Subject and Item as random effects.

In addition to StimDur and Tone, the models for delayed naming task also had an additional control variable of delay (Delay), which encoded the type of delay (600 ms, 1200 ms).

To summarize, all four measures of neighborhood schemes, each associated with a set of ND and NF were tested in the models. Model trimming and selection followed the procedure described in Chapter 3, Section 3.3. In the following Section, I report result from the best-fitted model selected from the four neighborhood schemes.

5.3 Results

5.3.1 Instantaneous naming task

The mean RT for the instantaneous naming task was 701.56 ms (*SD* = 157.95). The data were trimmed for trials that were 2 SD above and below the mean RT. As a result, 7.22% data were excluded from the analysis.

As already mentioned in Section 2.5, I repeat the hypotheses below.

(1) Hypothesis 1A:

Given the independent representation model, homophone mates and phonological neighbors should have similar effects on the recognition of spoken real monosyllables in Mandarin. In other words, either both are inhibitory, or both are facilitative, or both have null effects on spoken word recognition. Given previous findings for both English and Mandarin (Luce & Pisoni, 1998; Wang et al., 2012), the most likely pattern is for both phonological neighbors and homophone mates to be inhibitory. Furthermore, the two should also have the same interaction patterns with word frequency. Given previous studies on Chinese homophone effects (Wang et al., 2012; Zhou, 2015), I predict that homophone mates have stronger influence for high frequency words than for low frequency words. Thus, phonological neighbors are predicted to show the same interaction with frequency.

Hypothesis 1B:

Given the shared representation model, homophone mates and phonological neighbors may not have similar effects on the recognition of spoken real monosyllables in Mandarin. In other words, either they will have opposite effects i.e, inhibitory for one and facilitatory for other, or one is inhibitory or facilitatory while the other has null effects on spoken word recognition. Given previous findings in Mandarin (Wang et al., 2012; Zhou, 2015), the most likely pattern for homophone mates is inhibitory. So, phonological neighbors might have facilitative or null effects in Mandarin. Furthermore, the two should not have the same interaction patterns with syllable frequency. Given previous studies on Chinese homophone effects (Wang et al., 2012; Zhou, 2015), I predict that homophone mates have stronger influence for high frequency syllables than for low frequency syllables. Thus, phonological neighbors are predicted to have no interaction with syllable frequency.

The dataset for the models on RT consisted of 14908 trials and 602 items types. Four separate models were constructed with (log-transformed) RT as the outcome variable, each representing neighborhood measures from a different neighborhood scheme : (1) model with neighborhood measures from one-segment/tone difference scheme (ND_SegT, NF_SegT); (2) model with neighborhood measures from one-segment difference scheme (ND_Seg, NF_Seg); (3) model with neighborhood measures from one-component/tone difference scheme (ND_CompT, NF_CompT); (4) model with neighborhood measures from one-component difference scheme (ND_Comp, NF_Comp). Similar to auditory lexical decision, all the initial models for naming experiment follow the same formula (9) from Chapter 3, repeated below), although the neighborhood measures were calculated using different schemes:

(9) log(RT) ~ ND +log(NF) + log(HD) + log(HF) + ND:log(HF) + log(HD):log(HF) + log(StimDur) + Tone + (1|Subject) + (1|Item)

All four initial models underwent a backward elimination procedure. Table 14 summarizes the model fit (AIC) and all fixed effects in all four final models. AIC values were used to decide the best neighborhood scheme. As shown in Table 14, all four models have the same AIC (-24339.4) because none of the models showed significant effects of any of the neighborhood measures (ND and NF; $p > 0.05$) or homophone measures (HD and HF; $p > 0.05$). In other words, none of the neighborhood measures and homophone measures showed any significant effect in predicting RT of an instantaneous naming task. Among the control variables, StimDur and Tone showed significant effects on RT. Syllables with longer stimulus duration were responded slowly.

Table 14: Summary of fixed effects in the RT models for real-syllables across all neighborhood schemes. The model contains 14908 trials from 602 item types.

Model with neighborhood measures based on the **one-segment /tone difference scheme**			
AIC = -24339.4			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.512706	0.017077	381.4
log(StimDur)	0.365775	0.013118	27.9
Tone = T2	0.012082	0.004464	2.7
Tone = T3	0.044618	0.0048	9.3
Tone = T4	0.030795	0.00455	6.8
Model with neighborhood measures based on the **one-segment difference scheme**			
AIC = -24339.4			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.512706	0.017077	381.4
log(StimDur)	0.365775	0.013118	27.9
Tone = T2	0.012082	0.004464	2.7
Tone = T3	0.044618	0.0048	9.3
Tone = T4	0.030795	0.00455	6.8
Model with neighborhood measures based on the **one-component /tone difference scheme**			
AIC = -24339.4			
Fixed effects			

Predictor	β	Standard Error	t-value
Intercept	6.512706	0.017077	381.4
log(StimDur)	0.365775	0.013118	27.9
Tone = T2	0.012082	0.004464	2.7
Tone = T3	0.044618	0.0048	9.3
Tone = T4	0.030795	0.00455	6.8
Model with neighborhood measures based on the **one-component difference scheme**			
AIC = -24339.4			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	6.512706	0.017077	381.4
log(StimDur)	0.365775	0.013118	27.9
Tone = T2	0.012082	0.004464	2.7
Tone = T3	0.044618	0.0048	9.3
Tone = T4	0.030795	0.00455	6.8

Overall, the model on RT of instantaneous naming task showed no significant effects of neighborhood measures or homophone measures. The results from this experiment provides no evidence to any of the hypothesis (1A and 1B).

5.3.2 Delayed naming task

The mean RT for the delayed naming task was 446.35 ms (*SD* = 271.04 ms). The data were trimmed for trials 2 SD above the mean RT. As a result, 1.42 % of the data was rejected. The dataset for the model on RT consisted of 14403 trials and 602 items types. Similar to instantaneous naming task, all four-neighborhood schemes were tested in separate models. All the initial models for delayed naming task follow the same formula (11) from Chapter 3 (repeated below), although the neighborhood measures were calculated using different schemes.

(11) log(RT) ~ ND +log(NF) + log(HD) + log(HF) + ND:log(HF) + log(HD):log(HF) + Delay + log(StimDur) + Tone + (1|Subject) + (1|Item)

All four initial models underwent a backward elimination procedure. Table 15 summarizes the model fit (AIC) and all fixed effects in all four final models. AIC values were used to decide the best neighborhood scheme. As shown in Table 15, all four models have the same AIC (858.2) because none of the model showed significant effect of any of the neighborhood measures (ND and NF; $p > 0.05$) or homophone measures (HD and HF; $p > 0.05$). In other words, none of the neighborhood measures and homophone measures showed any significant effect in predicting RT of delayed naming task. Among the control variables, StimDur, Delay, and Tone had significant effects. Surprisingly, StimDur showed a facilitative effect on RT, i.e. syllables with longer stimulus duration were responded quickly compared to syllables with shorter stimulus duration. It is unclear as to why syllables with longer duration were responded faster. Delay also significantly affected the RT i.e.

longer the delay, longer was the RT. In addition, T3 and T4 were produced with longer RT as compared to other tones.

Table 15: Summary of fixed effects in the RT models for real-syllables across all neighborhood schemes. The model contains 14403 trials from 602 item types.

Model with neighborhood measures based on the **one-segment /tone difference scheme**			
AIC = 858.2			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	5.922488	0.044995	131.62
Delay (600)	0.110531	0.004032	27.41
log(StimDur)	-0.45802	0.026168	-17.5
Tone = T2	-0.015538	0.008907	-1.74
Tone = T3	0.03331	0.009571	3.48
Tone = T4	-0.052883	0.009079	-5.82
Model with neighborhood measures based on the **one-segment difference scheme**			
AIC = 858.2			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	5.922488	0.044995	131.62
Delay (600)	0.110531	0.004032	27.41
log(StimDur)	-0.45802	0.026168	-17.5
Tone = T2	-0.015538	0.008907	-1.74

Tone = T3	0.03331	0.009571	3.48
Tone = T4	-0.052883	0.009079	-5.82
Model with neighborhood measures based on the **one-component /tone difference scheme**			
AIC = 858.2			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	5.922488	0.044995	131.62
Delay (600)	0.110531	0.004032	27.41
log(StimDur)	-0.45802	0.026168	-17.5
Tone = T2	-0.015538	0.008907	-1.74
Tone = T3	0.03331	0.009571	3.48
Tone = T4	-0.052883	0.009079	-5.82
Model with neighborhood measures based on the **one-component difference scheme**			
AIC = 858.2			
Fixed effects			
Predictor	β	Standard Error	t-value
Intercept	5.922488	0.044995	131.62
Delay (600)	0.110531	0.004032	27.41
log(StimDur)	-0.45802	0.026168	-17.5
Tone = T2	-0.015538	0.008907	-1.74
Tone = T3	0.03331	0.009571	3.48
Tone = T4	-0.052883	0.009079	-5.82

Overall, the model on RT of delayed naming task showed no significant effects of neighborhood measures or homophone measures.

5.4 Summary

The auditory naming experiment was aimed at investigating the effects of phonological neighborhood and homophones on spoken word recognition. In this experiment there were to tasks: (1) Instantaneous naming, and (2) Delayed naming. The delayed naming task was conducted to control for any articulatory effects that might occur due to production in a naming experiment. Results from instantaneous naming task showed no significant effect of neighborhood measures (ND, NF, and interaction between ND and HF) or homophone measures (HD, HF, and interaction between HD and HF) on spoken word recognition. Among the control variables, both StimDur and Tone showed significant effects on RT. Syllables with longer duration were responded slowly. Results from delayed naming task were similar to instantaneous naming task. None of the critical predictors emerged as significant predictor of RT. The absence of any significant effects of the critical predictors on the RT of delayed naming task indicate that the effects observed in instantaneous task were based on perception and not production. Among the control variables, both StimDur and Tone showed significant effects on RT. Contrary to instantaneous naming task, syllables with longer duration were responded faster compared to syllables with shorter duration in delayed naming task.

Chapter 6: Discussion and conclusion

The main goal of the current dissertation is to investigate the effects of phonological neighbors and homophone mates on spoken word recognition in Mandarin, focusing on whether phonological neighbors and homophone mates have similar effect. A secondary goal is to compare different ways of modeling phonological neighborhoods in Mandarin in terms of the effectiveness of predicting phonological neighborhood effects. To achieve these goals, two experimental studies were designed and conducted: an auditory lexical decision experiment and an auditory naming experiment. Experimental results were analyzed in a series of mixed effects models, where the critical fixed-effects predictors include neighborhood density (ND), neighborhood frequency (NF), homophone density (HD), homophone frequency (HF), the interactions between ND and HF and the interaction between HD and HF. To compare the effectiveness of different neighborhood schemes, separate models were built with neighborhood measures from four neighborhood schemes defined by (1) one-segment/tone difference rule, (2) one-segment difference rule, (3) one-component/tone difference rule, and (4) one-component difference rule, respectively.

6.1 Discussion of results from auditory lexical decision

6.1.1 Results from real monosyllables

The mean accuracy in an auditory lexical decision experiment is 84%, and the average reaction time (RT) in correctly identified trials is 1023.33 ms. Previous research on auditory lexical decision in Mandarin (Wang et al., 2012) typically reported an accuracy rate of over 90% and a mean RT in correctly identified trials around 800 ms. Compared with the previous results, performance in the current study were significantly less accurate and slower. This discrepancy can be explained by two differences between the current study and previous research: (1) difference in usage frequency of the stimuli, and (2) difference in stimulus duration. The monosyllabic stimuli in Wang et al.'s study was all high in usage frequency, whereas the monosyllabic stimuli in the current study cover a wide range of syllable frequency (see Table 4 for range of frequency used in current dissertation). It is well established that high frequency items are recognized faster and more accurately than low frequency items. Therefore, the higher accuracy and shorter RT observed in Wang et al.'s study could be attributed to the selective use of high frequency syllables as stimuli. Secondly, in both the current study and previous research, RT is measured from the onset of the stimuli to the input of the response. However, the auditory stimuli used in the current study were much longer (mean duration = 626 ms) than that of the stimuli in Wang et al.'s study (mean duration = 430 ms), which may have led to a later recognition point (therefore longer RT) in the current study.

It should also be noted that when compared with the performance in auditory lexical decision in other languages such as English (e.g. Luce and Pisoni (1998) reported a mean RT around 400ms), the performance in the current study seems to be even more impoverished. I argue that the seemingly large difference in performance is mainly due to the difference in the nature of the task. In a typical auditory lexical decision (such as the one for English), participants are asked to categorize an auditory stimulus as either a word or a non-word. This paradigm may not be viable for Mandarin monosyllabic morphemes, which are the target items used in the current study, both because each Mandarin monosyllable may be mapped to multiple lexical items and because a Mandarin monosyllabic morpheme may not be understood as a “word” by Mandarin speakers (the concept of “word” in Chinese languages is highly ambiguous). As a result, in the auditory lexical decision experiment of the current study, participants were instead asked to decide whether or not they could associate the auditory stimulus to at least one Chinese character (representing a monosyllabic morpheme). In other words, if the participant can think of at least one character that can be associated with the auditory stimulus they have heard, they should give a “real” monosyllable response; otherwise, the participant should give a “non-exists” pseudo-syllable response. Compared to a typical lexical decision task, the task used in the current study is much more demanding, which in turn leads to much longer RT.

In this study, four different neighborhood schemes (See Table 3), based on different neighborhood definitions, were tested and examined. Each neighborhood scheme gives a set of ND and NF measures. As mentioned in Chapter 3, neighborhood measures across schemes have a sizable correlation (r is between 0.4 -0.8, see Table 6), and their effectiveness is tested in separate models. Overall, neighborhood measures from different neighborhood schemes produce highly similar modeling results regarding neighborhood effects. As shown in Table 16, none of the neighborhood measures had significant effects on the accuracy of identifying real monosyllables, but HD and HF both had facilitative effect on Accuracy. Meanwhile, all the models on RT showed significant facilitatory effects of NF, HD, and HF and an inhibitory interaction of HD and HF, in addition to highly similar control effects of StimDur and Tone. Based on the comparison of AIC values, the RT models with neighborhood measures from the one-component/tone difference neighborhood scheme had the best model fit (AIC = -5566.5). The second-best model fit was based on one-segment/tone difference scheme (AIC = -5562.0). The difference between the best and the second-best model is 4.5. A difference of 2 in AIC indicates that the model with lower AIC is 0.368 times more probable as the model with the higher AIC (Anderson & Burnham, 2004). Therefore, it can be concluded that the neighborhood measures based on one-component/tone difference scheme best predicts the model results and its model fit is significantly better than the other alternative models.

Taken together, the models for real monosyllables in auditory lexical decision showed that other things being equal, real monosyllables with many homophone mates and higher-frequency homophone mates were identified more accurately and faster than those with fewer homophone mates or lower-frequency homophone mates; real monosyllables with higher-frequency phonological neighbors are responded to faster than those with lower-frequency neighbors. Furthermore, the facilitation effect of homophone density was reduced for real monosyllables with high frequency homophone mates.

Table 16: Summary result from real monosyllables on auditory lexical decision experiment with significant predictors only. Here + indicates a positive coefficient and - indicates a negative coefficient.

Neighborhood Scheme	Model on Accuracy	Model on RT
One-segment/tone difference rule	HD (+) HF (+)	NF (-), HD (-), HF (-), HD:HF (+), StimDur (+),

		Tone
One-segment difference rule		NF (-), HD (-), HF (-), HD:HF (+), StimDur (+), Tone
One-component/tone difference rule		NF (-), HD (-), HF (-), HD:HF (+), StimDur (+), Tone
One-component difference rule		NF (-), HD (-), HF (-), HD:HF (+), StimDur (+), Tone

How to account for the observed effects in the identification of real monosyllables in the current auditory lexical decision experiment? Obviously, the finding of inhibitory effects of phonological neighborhoods (mostly from neighbor frequency in the current study) do not agree with previously documented inhibitory effects of neighborhood density and frequency in English spoken word recognition. However, it should be noted that effects of phonological neighbors may very well be language-specific. The research by Vitevitch and Rodríguez (2005) found significant facilitatory effects of both ND and NF on Spanish spoken word recognition. The authors suggested that the facilitatory effects of phonological neighborhood seen in Spanish could have been due to language-specific properties of the lexicon.

In the current study, not only do we observe contrast with previous studies in terms of neighborhood effects, we also see differences from previous research in the findings regarding homophone mates (HD and HF). As mentioned earlier, Wang et al. (2012) reported inhibitory effects of HD on Mandarin in auditory lexical decision (i.e. monosyllables with many homophone mates are more difficult to identify than those with fewer homophone mates), which is contrary to the present findings. I argue that this difference could be due to the difference in stimuli selection. As mentioned earlier, Wang et al.'s study only used high frequency monosyllables whereas the current experiment used both high- and low frequency syllables. As suggested in Zhou (2015), there is probably an interaction of homophone density and overall frequency, in the sense that the inhibitory effect of homophone mates is reduced for low frequency items than for high frequency items. In the current study, we also observe this interaction (HD:HF, where HF is a proxy for syllable frequency, since the two are highly correlated) in the same direction. That is to say, it is very likely that the inhibitory effects of HD observed in Wang et al.'s study is limited to high frequency items only. When the range of frequency is extended—as in the current study, the main effect of HD is no longer inhibitory.

According to the current results, there is an overall facilitative effect of HD but a negative interaction of HD and HF on RT of spoken word recognition. In other words, the facilitation of HD is reduced when HF is high. How to interpret these findings? One possible explanation is to allude to the difference between strong co-activated items and weak co-activated items. Chen and Mirman (2012) conducted a series of simulations for single word processing to investigate the dynamics of interactive activation and competition that could account for both facilitative and inhibitory effects of lexical neighbors. In their simulations, they used one of the general principles in the theories of cognition i.e. selection of the right candidate via competition among parallelly activated multiple similar representations. They found that co-activated representations could have either facilitative or inhibitory effects, depending on the magnitude of their co-activations. Specifically, there may be a net inhibitory effect if the representations are strongly activated or a net facilitative effect if they are weakly activated. In other words, strongly activated items results in competition among the co-activated items resulting in inhibitory effects, while weakly activated items help each other when co-activated resulting in facilitatory effects.

We can extend the contrast of strong vs. weak co-activated items to the current findings. In my dataset, HF is highly correlated with the presence of high-frequency homophone mates ($r > 0.9$). In other words, if a homophone family has overall high frequency, it is very likely that there are high frequency homophone mates in the family. In the current auditory lexical decision task, participants were asked to associate the heard stimuli with a character. Given the nature of task, the strongest competition is probably between the high-frequency homophone mates. According to Chen and Mirman (2012), strong co-activated items compete with each other. Therefore, high frequency homophone mates were strongly activated and competed with each other, resulting in a trend towards inhibition. This is most evident from the items with the top 25% of homophone frequency, which showed a trend of an inhibitory effect of higher HD. Conversely, items with the lowest 25% homophone frequency showed an overall facilitative effect of HD.

Following the same reasoning, it is possible that in this task where participants were asked to associate the auditory stimulus to a character, phonological neighbors were even more weakly activated compared to low-frequency homophone mates. This could have resulted in an overall minuscule effect of phonological neighbors on spoken word processing which was observed from the facilitative effects of NF on RT model alone.

An alternative explanation for the facilitation of NF may also be attributed to possible false identification that inadvertently yielded correct responses. If a real monosyllable has some high frequency phonological neighbors, it is likely that the participant wrongly associates the auditory stimulus with a similar-sounding syllable, and produce the "real" monosyllable response, which is technically the correct answer although it arises from a false identification. Having high frequency neighbors may promote this type of responses, and thus contribute to the overall observation of "facilitation" of phonological neighborhoods.

The results from auditory lexical decision experiment on real monosyllables provide some support to both hypothesis 1A and 1B. According to hypothesis 1A, phonological neighbors and homophone mates might have similar effects on spoken word recognition based on independent representation model. There is evidence for similar effects of phonological neighbors and homophone mates from the current results. Both phonological neighbors (NF) and homophone mates (HD and HF) have similar facilitative effects on spoken word recognition on RT model. According to hypothesis 1B, phonological neighbors and homophone mates might have different effects on spoken word recognition based on shared representation model. There is evidence for different effects of phonological neighbors and homophone mates from the current results. Both HD and HF showed significant facilitative effects on spoken word recognition on RT and accuracy models while ND showed no effect on spoken word recognition on both RT and accuracy models. Also, NF showed no effect on accuracy model. Along with fixed effects, HD and HF showed significant interaction on RT model whereas ND and HF showed no interaction on RT and accuracy models. This indicates that the effects of phonological neighbors and homophone mates are not entirely the same. These differences provide some evidence for hypothesis 1B. Our findings lend more support to hypothesis 1B that supports shared representation model which suggests that phonological neighbors and homophone mates are represented differently in the mental lexicon resulting in different effects on spoken word recognition.

6.1.2 Results from pseudo-syllables

The mean accuracy of pseudo-syllables is 86% and the average RT incorrectly identified trials is 1087.95 ms. Although the difference was not much between RT and accuracy rate of real and pseudo-syllables (difference in mean RT = 64.62 ms, difference in mean accuracy rate = 2%), RT was longer, and the accuracy rate was higher in pseudo-syllables than real monosyllables. As discussed in the above Section, this could be attributed to the type of task involved. In this task, participants were asked to associate the heard stimulus to at least one character in Mandarin. Therefore, when the participant hears a stimulus, participant keeps searching for lexical units that match with the stimulus and finally decide that the stimulus does not match with anything in case of a pseudo-syllables. Two types of pseudo-syllables were used: (1) Tonal gaps, and (2) Segmental gaps. Tonal gaps were responded slowly and less accurately as compared to segmental gaps (see Table 17).

Table 17: Summary of mean RT and mean accuracy rate of pseudo-syllables.

ItemType	Mean RT (ms)	Mean Accuracy rate (%)
Tonal gaps	1142.97	84%
Segmental gaps	1011.89	88%

As described in Chapter 3, tonal gaps are pseudo-syllables that exist with other tones while segmental gaps do not exist with any tone. Therefore, when a stimulus is a tonal gap, the syllables that exist with other tones might get activated. In order to correctly identify the stimuli as pseudo-syllable, participants need to reject all the activated real monosyllables. This makes the process more competitive resulting in longer RT and lower accuracy rates. However, when a stimulus is a segmental gap, there are no real monosyllables that exist with other tones. So, the process of rejection and deciding it as a pseudo-syllable becomes easier and faster resulting shorter RT and higher accuracy. These results are similar to findings in (Neergaard, 2018) where he reported that segmental gaps (mean RT = 1086 ms) were responded faster than tonal gaps (mean RT = 1176 ms).

Since, pseudo-syllables are gaps in Mandarin that do not exist. Therefore, they do not have homophone mates or frequency. However, they can have phonological neighbors (i.e. real monosyllables) that sound similar to them. Thus, pseudo-syllables were analyzed for neighborhood effects alone. Four different schemes of neighborhood based on different definitions of neighborhood were tested. Overall, neighborhood measures across different schemes produced largely similar modelling results. As shown in Table 18, models of RT showed significant inhibitory effects of ND on all the four neighborhood schemes. NF, on the other hand, showed facilitatory effects on three out of four schemes. In accuracy models, ND showed significant inhibitory effect on all four neighborhood schemes. While, NF showed significant facilitatory effect on two out of four neighborhood schemes.

Table 18: Summary result from pseudo-syllables on auditory lexical decision experiment with significant predictors only. Here + indicates a positive coefficient and - indicates a negative coefficient.

Neighborhood Scheme	Model on accuracy	Model on RT
One-segment/tone difference rule	ND (-), ItemType: Tonal gap (-), StimDur (+), Tone	ND (+), NF (-), ItemType: Tonal gap (+), StimDur (+), Tone
One-segment difference rule	ND (-), NF (+), ItemType: Tonal gaps (-), StimDur (+), Tone	ND (+), ItemType: Tonal gap (+), StimDur (+), Tone
One-component/tone difference rule	ND (-), ItemType: Tonal gaps (-), StimDur (+), Tone	ND (+), NF (-), ItemType: Tonal gap (+), StimDur (+), Tone
One-component difference rule	ND (-), NF (+), ItemType: Tonal gaps (-), StimDur (+), Tone	ND (+), NF (-), ItemType: Tonal gap (+), StimDur (+), Tone

Based on the comparison of AIC values, accuracy model with neighborhood measures from one-segment/tone difference neighborhood scheme had the best model fit (AIC = 9587.6). The second-best model was with neighborhood measures from one-segment difference scheme (AIC = 9673.1) whose AIC differs from the best model by 85.5. Therefore, the model with neighborhood measures from one-segment/tone difference scheme is significantly better than all the other three models. For RT models, neighborhood measures based on one-component/tone difference scheme had the best model fit (AIC = -655.9). The second-best model was with neighborhood measures from one-component difference scheme (AIC = -627.9) whose AIC differs from the best model by 28.0. Therefore, the model with one-component/tone difference scheme was found to be significantly better than all the other alternative models of RT.

Overall, pseudo-syllables with many real monosyllable neighbors were identified less accurately and slowly compared to pseudo-syllable with fewer real monosyllable neighbors. The inhibitory effects of ND are in line with the previously documented ND effects in English (Luce & Pisoni, 1998; Vitevitch & Luce, 1999) and Mandarin (Tsai, 2007). Similar inhibitory effects of ND were seen on a different task in Mandarin (Tsai, 2007). When a pseudo-syllable with many neighbors is heard, a greater number of real monosyllables those are neighbors to the target pseudo-syllable are co-activated. In order to identify the target syllable as pseudo-syllable, participants need to reject all the strongly activated real monosyllables. This makes the process difficult, resulting in longer RT. Also, when there are more real monosyllable neighbors, it is more likely that the participants wrongly associate the target pseudo-syllable with a real monosyllable resulting in an incorrect response that leads to lower accuracy rates.

NF, on the other hand, showed a significant facilitative effect on RT. In simple words, pseudo-syllable with high frequency neighbors (real monosyllables) were responded faster compared to pseudo-syllables with low frequency neighbors. It is unclear, as to why NF has a facilitatory effect on RT in pseudo-syllables. In case of pseudo-syllables when high frequency real monosyllables are co-activated, an incorrect association with real monosyllables could result in incorrect responses. Thus, there should not be any effects of NF for pseudo-syllables because incorrect responses will be excluded from RT models. So, no effects should be found which is clearly not the case. Three out of four schemes of neighborhood showed significant faciliatory effects of NF on RT and two out of four schemes showed significant faciliatory for accuracy models. Therefore, it makes it hard to explain these consistent yet conflicting effects with the previous findings on NF. At this stage, it is difficult to say whether these effects were language-specific until more research is devoted to look at the NF effects. Is the effect specific to certain gap type (segmental gaps or tonal gaps) or only existent for certain neighborhood size (when ND is high vs. low)? More research is needed for further investigation.

6.2 Discussion of results from auditory naming

In the auditory naming experiment, two tasks were conducted: (1) Instantaneous naming, and (2) Delayed naming. The purpose of delayed naming task was to control for any articulatory effects due to production of the syllables. In a naming task, along with perception of stimuli, production of the stimuli was also conducted. To make sure that the effects seen were solely from perception and not affected by the production, delayed naming task was conducted. When a delay is used, it is assumed that lexical access has already taken place and what is left are the effects that might have arisen from planning articulation of the stimuli. For the purpose of delay, 600 ms and 1200 ms were used. The mean RT for instantaneous naming task was 701.56 ms and for delayed naming task was 446.35 ms. Delay condition of 600 ms had a mean RT of 463.62 ms and a delay condition of 1200 ms had a mean RT of 429.07. RT in “no delay” was longer than RT in delay conditions, and RT in 600ms delay was longer to that in 1200ms condition. However, the difference in the RT is not huge (difference = 34.55 ms).

Accuracy of five audio files randomly selected from the instantaneous naming task were calculated. The mean accuracy rate was 99%. Accuracy rate from a randomly selected sample was very high, suggesting ceiling effect in accuracy. As a result, accuracy data were not considered for further analyses and only RT data were further analyzed.

The overall performance of the participants in the current task with no delay was slower (mean RT: 701.56 ms) as compared to a previous study (Tsai, 2007), that reported a mean RT of around 440 ms. But the accuracy rate was comparable to previous study (Tsai, 2007) that reported a mean accuracy rate of 98%. The slower performance in the current task can be explained on the basis of stimulus parameters mainly syllable frequency, and stimulus duration. In Tsai’s study (Tsai, 2007), only high frequency syllables were used, however, in the current study, a wide range of syllable frequency was used that consisted of both high and low frequency syllables. Therefore, based on the previous findings (Luce & Pisoni, 1998) on word frequency, it is known that high frequency words are recognized earlier as compared to low frequency words. Therefore, the shorter RT in Tsai's study can be attributed to the use of high frequency syllables as stimuli. However, in the current study, a wide range of syllable frequency has been used that could have resulted in longer RT. Secondly, the mean stimulus duration in Tsai's study ranged from 332 ms -358 ms. However, the mean stimulus duration in the current study was 626 ms. A shorter stimulus duration in Tsai's study could have resulted in earlier recognition resulting in shorter RT. In comparison, the stimulus duration in the current study was longer compared to Tsai's study that would have contributed to longer RT.

RT models were separated by delay condition (no delay vs. delay). Under no delay condition, none of the neighborhood measures and homophone measures showed any significant effect on the processing. All the four neighborhood scheme yielded same results as none of the neighborhood measures showed any significant effects, resulting in same AIC values. Similarly, in the delay condition, none of the neighborhood measures and homophone measures exhibited any significant effect (see Table 19). As a result, all four neighborhood scheme showed same result with same AIC values for model fit.

Table 19: Summary result from auditory naming experiment with significant predictors only. Here + indicates a positive coefficient and - indicates a negative coefficient.

Neighborhood Scheme	Model on RT in instantaneous naming	Model on RT in delayed naming
One-segment/tone difference rule	StimDur (+), Tone	Delay: 600ms (+), StimDur (-), Tone

One-segment difference rule		
One-component/tone difference rule		
One-component difference rule		

Overall, null effects of neighborhood density, neighbor frequency, homophone density and frequency were observed. The findings from auditory naming experiment is in contradiction to earlier reported findings by Tsai (2007). Tsai (2007) reported inhibitory neighborhood density effect in auditory naming experiment. However, there were concerns with the method of the study. The significant inhibitory effect of neighborhood density was seen in condition where the difference in high and low neighborhood density was relatively low (difference between mean high and mean low ND = 1.81). Surprisingly, no significant effect of neighborhood density was found when the difference in high and low neighborhood density was large (difference between mean high and mean low ND = 7.88). Assuming the effect of neighborhood density exists, it should also be visible in condition with large density difference, which was not the case.

In general, it can be noted the effects are reduced in naming experiments. Balota and Chumbley (1984) compared naming task with lexical decision and found that the frequency effects were reduced in naming task. They also suggested that the effects were more reduced when the task modality is auditory. Therefore, finding no effects on instantaneous naming task might be due to the reason that the effects were reduced due to the modality of task.

Luce and Pisoni (1998) used both auditory lexical decision and auditory naming experiment to investigate the effects of phonological neighborhood and found that not just frequency effects were reduced, significant NF effects observed in auditory lexical decision were also absent in auditory naming experiment. Therefore, the absence of effects could be attributed to task differences. In general, an auditory lexical decision is more difficult task that involves accessing the lexicon for making a decision, compared to a naming task where the participants need to repeat the heard stimuli, that might not involve the access to the lexicon. This difference in the task might result in having reduced effects for naming task that showed no significant effects for any of the critical predictors.

Having no effects of critical variables on delayed naming suggests that the results of instantaneous naming were not affected by the production rather the results were solely based on perception. No significant effects of any of the critical variables on delayed naming task suggests the absence of neighborhood effects and homophone effects due to production.

6.3. Theoretical implications

In this dissertation, I conducted two experiments, namely, auditory lexical decision and auditory naming for investigating the effects of lexical neighborhoods, consisting of both phonological neighbors and homophone mates, on spoken word recognition in Mandarin. In the auditory lexical decision task, seventy-two participants were recruited, who had to hear a Mandarin syllable and associate it with a Chinese character to categorize it as a real monosyllable or pseudo-syllable as quickly and as accurately as possible. 1259 real monosyllables and 758 pseud-syllables were used as stimuli. Both RT and accuracy were measured, and mixed-models were built for RT and accuracy. Four neighborhood schemes were tested in separate models with neighborhood measures from each scheme. For real monosyllables, models on RT reflected significant facilitatory effects of homophone measures (HD and HF) and neighbor frequency (NF). In addition, significant inhibitory interaction between HD and HF were found. Neighborhood scheme based on one-component/tone scheme gave the best model fit. Models on accuracy showed significant facilitative effects of homophone measures (HD and HF) alone. None of the neighborhood measures nor the interactions between homophone measures showed any significant effects on accuracy models on real monosyllables. In case of pseudo-syllables, both accuracy and RT models showed significant inhibitory effects of ND. However, significant facilitative effect of NF was observed in the model on RT. AIC values gives the best fit on RT models for neighborhood measures based on one-component/tone difference scheme and best fit on accuracy models for neighborhood measures based on one-segment/tone scheme. In auditory naming experiment, 130 participants were recruited, who were instructed to repeat the heard real monosyllable as quickly and accurately as possible. Under naming experiment, 65 participants took part in no delay condition and rest 65 participants took part in delayed condition. The purpose of delayed condition was to account for any articulatory effects related to production that might have occurred in the naming experiment. Neither naming task showed any significant effect of neighborhood measures and homophone measures.

My general hypothesis was regarding the similarity and differences between phonological neighbors and homophone mates in terms of their influence on spoken word recognition. Specifically, I hypothesized that phonological neighbors and homophone mates would have similar effects on the recognition of real monosyllables based on independent representation model. Based on this hypothesis, either the effects could be inhibitory, or facilitative, or null effects for both homophone mates and phonological neighbors on spoken word recognition. In addition, both homophone mates and phonological neighbors were predicted to reveal similar interaction patterns with frequency (see Hypothesis 1A). However, based on shared representation model, phonological neighbors and homophone mates were hypothesized to have dissimilar effects on the recognition of real monosyllables in Mandarin. Based on this hypothesis, the effects were predicted to be either completely opposite i.e. inhibitory for one and facilitatory for other, or the effects would not be similar. Further, according to this hypothesis, the two (homophone mates and phonological neighbors) would not have the same interaction patterns with frequency (see Hypothesis 1B).

Given the current results, there is some evidence for both hypotheses 1A and 1B. More specifically, the evidence emerges from the results from auditory lexical decision experiment for real monosyllables. Both neighborhood measures (NF) and homophone measures (HD and HF) showed significant facilitative effects on RT models of real monosyllables supporting hypothesis 1A. However, the significant interaction between homophone measures (HD and HF) and the absence of interaction between neighborhood measures (ND and HF) lends support to hypothesis 1B. Also, there was an absence of significant effects of ND in real monosyllables and presence of significant inhibitory effects of ND in pseudo-syllables. However, for NF, the effects were in opposite direction i.e. facilitatory for both real and pseudo-syllables. This further supports hypothesis 1B.

Can the current results be completely accommodated by either hypothesis? According to Hypothesis 1A - independent representation model, homophone mates and phonological neighbors share the same status in the lexicon as a result of which the two can be thought to have the similar effects in spoken word recognition. However, if we further enrich this model with the idea of strong versus weak activated items from Chen and Mirman (2012), it is possible to distinguish homophone mates as stronger lexical neighbors than phonological neighbors as homophone mates are more similar in the phonological form than phonological neighbors. In other words, it may be possible that while both phonological neighbors and homophone mates have the same representation/status in the lexicon, they may not exactly have the same effects on spoken word recognition. Due to the higher phonological similarity between homophone mates, they are more likely to exhibit competition effects (as in the inhibitory interaction of homophone density and homophone frequency) than phonological neighbors, even though both show overall facilitatory effects.

Can the shared representation model accommodate all current results? According to Hypothesis 1B shared model, homophone mates and phonological neighbors are represented at different levels of the lexicon. While homophone mates are distinguished at the lemma level and share a representation at the lexeme level, phonological neighbors are distinguished at both the lexeme level and the lemma level. The shared representation model does not exclude the possibility of phonological neighbors and homophone mates having similar effects; neither does it exclude the possibility of phonological neighbors and homophone mates having opposite effects. Overall, our findings show more support to shared representation model i.e. phonological neighbors and homophone mates have different representation in the mental lexicon. Further, it should be noted that the two hypotheses (hypothesis 1A and 1B) may not be totally mutually exclusive since hypothesis 1B does not totally exclude the possibility of phonological neighbors and homophone mates to behave in a similar manner.

Another important question that was investigated in the current dissertation was regarding the role of lexical tone in defining phonological neighbors in tone languages. Neighborhood schemes that included tone, emerged as the best fitted models out of all the four neighborhood schemes that were tested. The current study is in agreement with the findings of Neergaard and Huang (2016) who tested 14 schemes of defining phonological neighborhood, 7 included lexical tones and 7 without lexical tones. They also found that the best-fitted model results were obtained for the schemes that included lexical tone. Therefore, it can be concluded that tone is an important unit in Mandarin that cannot be ignored. However, at this stage, it is difficult to say whether the segmentation of phonological neighbors is based on segments or components (onset, rime and tone) in Mandarin. According to the current results, one-component/tone rule is the best fit measure in RT model for real monosyllables and pseudo-syllables and one-segment/tone difference rule is the best fit measure in accuracy model for pseudo-syllables. Segmentation based on components of syllables received more support from the experimental results compared to segmentation based on segments of syllables. Based on these results, it is evident that both one-component/tone rule and one-segment/tone rule provide a good fit for the results in comparison to the other two rules (one-component difference rule and one-segment difference rule). However, there is not enough evidence to ascertain whether it is the one-component/tone rule or one-segment/tone rule that provides the best fit.

6.4 Limitations of the study

In this dissertation, only one-unit difference rule was used to define phonological neighbors. One-unit difference rule states that any two words that differ by one unit, either by addition or deletion or substitution are phonological neighbors. The unit used in this dissertation was either segment or component. One-unit difference rule is a discrete measure to quantify phonological neighborhood based on perceptual similarity of words. All neighbors were treated equivalent according to this rule. But, all phonemes are not equally similar or confusable. All neighbors of a target word vary depending on the amount of confusability with the target word. Another limitation to this rule is that words that differ from the target word by two or more units were not included. There is a possibility that words that differ by multiple units are sometimes more confusing than words that differ by one unit.

6.5 Future directions

In future, to tease apart shared versus independent representation, experiments that specifically examine whether different homophone mates from the same family behave similarly or differently in lexical processing. For example, comparing the high frequency homophone mates with low frequency homophone mates from the same family on identification tasks. In order to tease apart whether segment or component define neighborhood in Mandarin, experiments that allow participants to rate which one of the candidate syllables is more similar to the target stimulus, can be conducted. In order to define strong versus weak items, computational models can be designed in future studies. In future, other measures of perceptual similarity can be used to test the neighborhood measures can also be investigated. For example, phi square density (Iverson et al., 1998) can be calculated based on perceptual similarity of phonemes and can be tested along with one-unit difference measure to evaluate the best model-fit. Other spoken word recognition tasks can be further employed to further confirm the results derived in the current dissertation. Further, neural mechanisms of the effects of neighborhood on spoken word recognition can be explored using event related potentials. In addition, more research is needed to understand the effects of neighborhood frequency on spoken word recognition.

References

Academia Sinica (1997). *Academia Sinica Balanced Corpus of Modern Chinese [version 3.0]*. Retrieved from http://www.sinica.edu.tw/ftms-bin/kiwi1/mkiwi.sh

Amano, N., & Kondo, K. (2003). NTT detabesu shirizu: Nihongo no goi-tokusei dai 1-ki CD-ROM-ban [NTT database series: Lexical properties of Japanese, Vol. 1, CD-ROM version]. *Tokyo, Japan: Sanseido*.

Amano, S., & Kondo, T. (2000). Neighborhood and cohort in lexical processings of japanese spoken words. *ISCA Tutorial and Research Workshop (ITRW) on Spoken Word Access Processes*. Retrieved from http://www.isca-speech.org/archive_open/swap/swap_091.html

Anderson, D., & Burnham, K. (2004). Model selection and multi-model inference. *Second. NY: Springer-Verlag*, *63*.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459.

Bates, D., Maechler, M., Bolker, B., Walker, S., & others. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, *1*(7). Retrieved from http://keziamanlove.com/wp-content/uploads/2015/04/StatsInRTutorial.pdf

Boersma, P., & Weenink, D. (2010). *Praat: Doing phonetics by computer [Computer program], Version 5.1. 44*. Retrieved from http://scholar.google.com/scholar?cluster=834095044295694369&hl=en&oi=scholarr

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0010729

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1430.

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, *17*(1), 205–215.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417.

Content, A., Mousty, P., & Radeau, M. (1990). Brulex. Une base de données lexicales informatisée pour le français écrit et parlé. *L'année Psychologique*, *90*(4), 551–566.

Cutler, A., & Chen, H.-C. (1995). Phonological similarity effects in Cantonese word recognition. *The Thirteenth International Congress of Phonetic Sciences*, 106–109. Retrieved from http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:77054

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313–349.

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Dufour, S., & Frauenfelder, U. H. (2010). Phonological neighbourhood effects in French spoken-word recognition. *The Quarterly Journal of Experimental Psychology*, *63*(2), 226–238.

Fang, J., Li, X., & Luo, W. (2014). Influence of Cumulative Frequencies of Chinese Syllables on the Activation of Homophone Representation in Auditory Lexical Access. *Acta Psychologica Sinica*, *46*(4), 467–480.

Francis, W. N., & Kucera, H. (1967). Computational analysis of present-day American English. *Providence, RI: Brown University Press. Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, AB (2014). Emotion and Language: Valence and Arousal Affect Word Recognition. Journal of Experimental Psychology: General*, *143*, 1065–1081.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496.

Iverson, P., Bernstein, L. E., & Auer Jr, E. T. (1998). Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition. *Speech Communication*, *26*(1–2), 45–63.

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824.

Jiauyubu Guoyu Cidian Jianbianben Bianji Shiautzu (Ministry of Education National Language Concise Lexicon Editorial Team) (1997). *Guoyu cidian jianbianben bianjitzliau tztspin tungji baugau (Word frequency statistic report of the database for national language concise lexicon)* [Web version, 3rd ed.]. Taipei: Ministry of Education, Republic of China.

Kirby, J. P., & Yu, A. C. (2007). Lexical and phonotactic effects on wordlikeness judgments in Cantonese. *Proceedings of the International Congress of the Phonetic Sciences XVI*, 1389–1392. Retrieved from https://www.researchgate.net/profile/James_Kirby9/publication/228379459_Lexical_and_phonotactic_effects_on_wordlikeness_judgments_in_Cantonese/links/53edf16a0cf2981ada173d92.pdf

Krajewski, G., & Matthews, D. (2010). RH Baayen, Analyzing linguistic data: A practical introduction to statistics using R. Cambridge: Cambridge University Press, 2008. Pp. 368. ISBN-13: 978-0-521-70918-7. *Journal of Child Language*, *37*(2), 465–470.

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, *53*(2), 157–165.

Li, X. J., Fang, J., & Lou, J. (2011). Interaction effect of homophone family size and specific-word frequency on the auditory lexical access of Chinese characters. *Journal of Psychological Science*, *34*, 43–47.

Li, X. J., Wang, W. N., & Li, X. Q. (2011). Auditory word frequency effect within homophone families and the activation of homophone representations. *Acta Psychologica Sinica*, *43*, 749–762.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1), 1–36.

Luce, Paul A. (1986). *Neighborhoods of Words in the Mental Lexicon. Research on Speech Perception. Technical Report No. 6.* Retrieved from http://eric.ed.gov/?id=ED353610

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, *50*(8), 2032–2043.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71–102.

Mathôt, S., Schreij, D., & Theeuwes, J. (2011). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mazzoni, D., & Dannenberg, R. (2000). Audacity (software). *The Audacity Team, Pittsburg, PA, USA*.

McBride-Chang, C., Tong, X., Shu, H., Wong, A. M.-Y., Leung, K., & Tardif, T. (2008). Syllable, phoneme, and tone: Psycholinguistic units in early Chinese and English word recognition. *Scientific Studies of Reading*, *12*(2), 171–194.

McClelland, J. L., & Elman, J. L. (1986). Interactive processes in speech perception: The TRACE model. *Parallel Distributed Processing*, *2*(58), 121.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165.

Neergaard, K. D., Xu, H., & Huang, C.-R. (2016). Database of Mandarin Neighborhood Statistics. *Language Resources and Evaluation Conference LREC 2016*. Retrieved from https://www.researchgate.net/profile/Chu-Ren_Huang/publication/303288164_Database_of_Mandarin_Neighborhood_Statistics/links/573b27d208ae9f741b2d78d2.pdf

Neergaard, K., & Huang, C.-R. (2016). Graph theoretic approach to Mandarin syllable segmentation. *Fifteen International Symposium on Chinese Language and Linguistics (IsCLL). Hsinchu: Hsinchu University of Education*.

Neergaard, Karl David. (2018). *Phonological segmentation neighborhoods* (PhD Thesis). The Hong Kong Polytechnic University.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*(3), 189–234.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon. *Research on Spoken Language Processing Report No*, *10*, 357–76.

Peereman, R. (1999). LEXOP: A lexical database providing orthography-phonology statistics for French monosyllabic words. *Behavior Research Methods, Instruments, & Computers*, *31*(2), 376–379.

Peramunage, D., Blumstein, S. E., Myers, E. B., Goldrick, M., & Baese-Berk, M. (2011). Phonological neighborhood effects in spoken word production: An fMRI study. *Journal of Cognitive Neuroscience*, *23*(3), 593–603.

Sadat, J., Martin, C. D., Costa, A., & Alario, F.-X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33–58.

Schneider, W., Eschman, A., & Zuccolotto, A. (2007). E-Prime 2.0 Professional. *Pittsburgh: Psychology Software Tools*.

Sebastián-Gallés, N. (2000). *LEXESP: Léxico informatizado del español*. Edicions Universitat Barcelona.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating Phonotactic Probability and Neighborhood Density in Adult Word Learning. *Journal of Speech, Language, and Hearing Research : JSLHR*, *49*(6), 1175–1192. https://doi.org/10.1044/1092-4388(2006/085)

Strand, J. F. (2014). Phi-square Lexical Competition Database (Phi-Lex): An online tool for quantifying auditory and visual lexical competition. *Behavior Research Methods*, *46*(1), 148–158.

Strand, J. F., & Sommers, M. S. (2011). Sizing up the competition: Quantifying the influence of the mental lexicon on auditory and visual spoken word recognition. *The Journal of the Acoustical Society of America*, *130*(3), 1663–1672.

Sun, C. C., Hendrix, P., Ma, J., & Baayen, R. H. (2018). Chinese lexical database (CLD). *Behavior Research Methods*, *50*(6), 2606–2629.

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

Tsai, P.-T. (2007). *The effects of phonological neighborhoods on spoken word recognition in Mandarin Chinese*. Retrieved from https://www.google.com/books?hl=en&lr=&id=UNld_LOYB6MC&oi=fnd&pg=PR2&dq=tsai+et+al+2007+phonological+neighborhood&ots=3PDtRmoB7X&sig=Y-1LB3kvWxncPmHZmml7LMWktq4

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735.

Vitevitch, M. S., Armbrüster, J., & Chu, S. (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.

Vitevitch, M. S., & Luce, P. A. (1998). When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, *9*(4), 325–329. https://doi.org/10.1111/1467-9280.00064

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, *40*(1), 47–62.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, Neighborhood Activation, and Lexical Access for Spoken Words. *Brain and Language*, *68*(1–2), 306–311. https://doi.org/10.1006/brln.1999.2116

Vitevitch, M. S., & Rodríguez, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*, *3*(1), 64–73.

Vitevitch, M. S., & Sommers, M. S. (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, M. S., Stamer, M. K., & Sereno, J. A. (2008). Word Length and Lexical Competition: Longer is the Same as Shorter. *Language and Speech*, *51*(4), 361–383. https://doi.org/10.1177/0023830908099070

Wang, W., Li, X., Ning, N., & Zhang, J. X. (2012). The nature of the homophone density effect: An ERP study with Chinese spoken monosyllable homophones. *Neuroscience Letters*, *516*(1), 67–71. https://doi.org/10.1016/j.neulet.2012.03.059

Webster's Seventh Collegiate Dictionary. Library Reproduction Service; Los Angeles: 1967. - Google Search. (n.d.). Retrieved January 23, 2019, from https://www.google.com/search?source=hp&ei=FBdJXOHeD-SGjwSO1Yq4CA&q=Webster%27s+Seventh+Collegiate+Dictionary.+Library+Reproduction+Service%3B+Los+Angeles%3A+1967.&btnK=Google+Search&oq

=Webster%27s+Seventh+Collegiate+Dictionary.+Library+Reproduction+Service%3B+Los+Angeles%3A+1967.&gs_l=psy-ab.3...794.794..1422...0.0..0.119.196.1j1......0....2j1..gws-wiz.....0.1ozZXpBGsZQ

Weiner, F. M., Miller, G. A. (1946). Transmission and reception of sounds under combat conditions. *Summary Technical Report of Division 17, 58-68*. Washington: *National Defense Research Committee*.

Yuan, L., Ci, C. Y., & Dian, P. C. (1990). *Dictionary of Frequently Used Words.* Beijing: Astronautics Publishing House.

Yip, M. (2002). *Tone*. Retrieved from https://www.google.com/books?hl=en&lr=&id=KFv2lojXjpwC&oi=fnd&pg=PR12&dq=yip+2002&ots=maIFLDBMDN&sig=PfsXqI03tBFqQ2CHU5C-N2_Mqew

Yoneyama, K. (2002). *Phonological neighborhoods and phonetic similarity in Japanese word recognition* (The Ohio State University). Retrieved from https://etd.ohiolink.edu/pg_10?0::NO:10:P10_ACCESSION_NUM:osu1302192053

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*(7), 1761–1770.

Zhou, W. (2015). *The Homophone Effect in Mandarin Word Recognition* (PhD Thesis). The Ohio State University.

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes*, *9*(3), 393–422.

Zhou, X., & Marslen-Wilson, W. D. (2009). Pseudohomophone effects in processing Chinese compound words. *Language and Cognitive Processes*, *24*(7–8), 1009–1038.

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, *48*(4), 779–793.

Appendices

A. List of target monosyllables in pinyin(n=1259).

a1, a2, a3, a4, ai1, ai2, ai3, ai4, an1, an3, an4, ang1, ang2, ang4, ao1, ao2, ao3, ao4, e1, e2, e3, e4, ei2, ei3, ei4, en1, en4, er2, er3, er4, o1, o2, o4, ou1, ou3, ou4, ba1, ba2, ba3, ba4, bai1, bai2, bai3, bai4, ban1, ban3, ban4, bang1, bang3, bang4, bao1, bao2, bao3, bao4, bei1, bei3, bei4, ben1, ben3, ben4, beng1, beng2, beng3, beng4, bi1, bi2, bi3, bi4, bian1, bian3, bian4, biao1, biao3, biao4, bie1, bie2, bie3, bie4, bin1, bin4, bing1, bing3, bing4, bo1, bo2, bo3, bo4, bu1, bu2, bu3, bu4, ca1, ca3, cai1, cai2, cai3, cai4, , can1, can2, can3, can4, cang1cang2, cao1, cao2, cao3, ce4, cen1, cen2, ceng1, ceng2, ceng4, cha1, cha2, cha3, cha4, chai1, chai2, chai4, chan1, chan2, chan3, chan4, chang1, chang2, chang3, chang4, chao1, chao2, chao3, chao4, che1, che3, che4, chen1, chen2, chen3, chen4, cheng1, cheng2, cheng3, cheng4, chi1, chi2, chi3, chi4, , chong1, chong2, chong3, chong4, chou1, chou2, chou3, chou4, chu1, chu2, chu3, chu4, chuai1, chuai3, chuai4, chuan1, chuan2, chuan3, chuan4, chuang1, chuang2, chuang3, chuang4, chui1, chui2, chun1, chun2, chun3, chuo1, chuo4, ci1, ci2, ci3, ci4, cong1, cong2, cou4, cu1, cu2, cu4, cuan1, cuan2, cuan4, cui1, cui3, cui4, cun1, cun2, cun3, cun4, cuo1, cuo2, cuo3, cuo4, da1, da2, da3, da4, dai1, dai3, dai4, dan1, dan3, dan4, dang1, dang3, dang4, dao1, dao2, dao3, dao4, de2, dei3, deng1, deng3, deng4, di1, di2, di3, di4, dia3, dian1, dian3, dian4, diao1, diao3, diao4, die1, die2, ding1, ding3, ding4, diu1, dong1, dong3, dong4, dou1, dou3, dou4, du1, du2, du3, du4, duan1, duan3, duan4, dui1, dui4, dun1, dun3, dun4, duo1, duo2, duo3, duo4, fa1, fa2, fa3, fa4, fan1, fan2, fan3, fan4, fang1, fang2, fang3, fang4, fei1, fei2, fei3, fei4, fen1, fen2, fen3, fen4, feng1, feng2, feng3, feng4, fo2, fou3, fu1, fu2, fu3, fu4, ga1, ga2, ga3, ga4, gai1, gai3, gai4, gan1, gan3, gan4, gang1, gang3, gang4, gao1, gao3, gao4, ge1, ge2, ge3, ge4, gei3, gen1, gen2, gen3, gen4, geng1, geng3, geng4, gong1, gong3, gong4, gou1, gou3, gou4, gu1, gu2, gu3, gu4, gua1, gua3, gua4, guai1, guai3, guai4, guan1, guan3, guan4, guang1, guang3, guang4, gui1, gui3, gui4, gun3, gun4, guo1, guo2, guo3, guo4, ha1, ha2, ha3, ha4, hai1, hai2, hai3, hai4, han1, han2, han3, han4, hang1, hang2, hang4, hao1, hao2, hao3, hao4, he1, he2, he4, hei1, hen2, hen3, hen4, heng1, heng2, heng4, hong1, hong2, hong3, hong4, hou2, hou3, hou4, hu1, hu2, hu3, hu4, hua1, hua2, hua4, huai2, huai4, huan1, huan2, huan3, huan4, huang1, huang2, huang3, huang4, hui1, hui2, hui3, hui4, hun1, hun2, hun4, huo1, huo2, huo3, huo4, ji1, ji2, ji3, ji4, jia1, jia2, jia3, jia4, jian1, jian3, jian4, jiang1, jiang3, jiang4, jiao1, jiao2, jiao3, jiao4, jie1, jie2, jie3, jie4, jin1, jin3, jin4, jing1, jing3, jing4, jiong1, jiong3, jiu1, jiu3, jiu4, ju1, ju2, ju3, ju4, jun1, jun4, juan1, juan3, juan4, jue1, jue2, jue3, jue4, ka1, ka3, kai1, kai3, kai4, kan1, kan3, kan4, kang1, kang2, kang4, kao1, kao3, kao4, ke1, ke2, ke3, ke4, ken3, ken4, keng1, kong1, kong3, kong4, kou1, kou3, kou4, ku1, ku3, ku4, kua1, kua3, kua4, kuai3, kuai4, kuan1, kuan3, kuang1, kuang2, kuang3, kuang4, kui1, kui2, kui3, kui4, kun1, kun3, kun4, kuo4, la1, la2, la3, la4, lai2, lai4, lan2, lan3, lan4, lang1, lang2, lang3, lang4, lao1, lao2, lao3, lao4, le1, le4, lei1, lei2, lei3, lei4, leng1, leng2, leng3, leng4, li1, li2, li3, li4, lia3, lian2, lian3, lian4, liang2, liang3, liang4, liao1, liao2, liao3, liao4, lie1, lie3, lie4, lin2, lin3, lin4, ling1, ling2, ling3, ling4, liu1, liu2, liu3, liu4, long1, long2, long3, long4, lou1, lou2, lou3, lou4, lu1, lu2, lu3, lu4, luan2, luan3, luan4, lue3, lue4, lun1, lun2, lun4, luo1, luo2, luo3, luo4, lv3, lv4, m2, ma1, ma2, ma3, ma4, mai2, mai3, mai4, man1, man2, man3, man4, mang2, mang3, mao1, mao2, mao3, mao4, mei2, mei3, mei4, men1, men2, men4, meng1, meng2, meng3, meng4, mi1, mi2, mi3, mi4, mian2, mian3, mian4, miao1, miao2, miao3, miao4, mie1, mie4, min2, min3, ming2, ming3, ming4, miu4, mo1, mo2, mo3, mo4, mou1, mou2, mou3, mu2, mu3, mu4, n2, n3, n4, na1, na2, na3, na4, nai3, nai4, nan1, nan2, nan3, nan4, nang1, nang2, nang3, nao1, nao2, nao3, nao4, ne2, ne4, nei3, nei4, nen4, neng2, ni1, ni2, ni3, ni4, nian1, nian2, nian3, nian4, niang2, niang4, niao3, niao4, nie1, nie4, nin2, ning2, ning3, ning4, niu1, niu2, niu3, niu4, nong2, nong4, nou4, nu2, nu3, nu4, nuan3, nue4, nuo2, nuo4, nv3, nv4, pa1, pa2, pa4, pai1, pai2, pai3, pai4, pan1, pan2, pan4, pang1, pang2, pang3, pang4, pao1, pao2, pao3, pao4, pei1, pei2, pei4, pen1, pen2, pen4, peng1, peng2, peng3, peng4, pi1, pi2, pi3, pi4, pian1, pian2, pian3, pian4, piao1, piao2, piao3, piao4, pie1, pie3, pin1, pin2, pin3, pin4, ping1, ping2, po1, po2, po3, po4, pou1, pou2, pou3, pu1, pu2, pu3, pu4, qi1, qi2, qi3, qi4, qia1, qia3, qia4, qian1, qian2, qian3, qian4, qiang1, qiang2, qiang3, qiang4, qiao1, qiao2, qiao3, qiao4, qie1, qie2, qie3, qie4, qin1, qin2, qin3, qin4, qing1, qing2, qing3, qing4, qiong2, qiu1, qiu2, qiu3, qu1, qu2, qu3, qu4, quan1, quan2, quan3, quan4, que1, que2, que4, qun1, qun2, ran2, ran3, rang1, rang2, rang3, rang4, rao2, rao3, rao4, re3, re4, ren2, ren3, ren4, reng1, reng2, ri4, rong2, rong3, rou2, rou4, ru2, ru3, ru4, ruan3, rui2, rui3, rui4, run4, ruo4, sa1, sa3, sa4, sai1, sai4, san1, san3, san4, sang1, sang3, sang4, sao1, sao3, sao4, se4, sen1, seng1, sha1, sha3, sha4, shai1, shai3, shai4, shan1, shan3, shan4, shang1, shang3, shang4, shao1, shao2, shao3, shao4, she1, she2, she3, she4, shei2, shen1, shen2, shen3, shen4, sheng1, sheng2, sheng3, sheng4, shi1, shi2, shi3, shi4, shou1, shou2, shou3, shou4, shu1, shu2, shu3, shu4, shua1, shua3, shua4, shuai1, shuai3, shuai4, shuan1, shuan4, shuang1, shuang3, shui2, shui3, shui4, shun3, shun4, shuo1, shuo4, si1, si3, si4, song1, song3, song4, sou1, sou3, sou4, su1, su2, su4, suan1, suan4, sui1, sui2, sui3, sui4, sun1, sun3, suo1, suo3, ta1, ta3, ta4, tai1, tai2, tai3, tai4, tan1, tan2, tan3, tan4, tang1, tang2, tang3, tang4, tao1, tao2, tao3, tao4, te4, tei1, teng2, ti1, ti2, ti3, ti4, tian1, tian2, tian3, tian4, tiao1, tiao2, tiao3, tiao4, tie1, tie3, tie4, ting1, ting2, ting3, ting4, tong1, tong2, tong3, tong4, tou1, tou2, tou3, tou4, tu1, tu2, tu3, tu4, tuan1, tuan2, tuan3, tuan4, tui1, tui2, tui3, tui4, tun1, tun2, tun3, tun4, tuo1, tuo2, tuo3, tuo4, wa1, wa2, wa3, wa4, wai1, wai3, wai4, wan1, wan2, wan3, wan4, wang1, wang2, wang3, wang4, wei1, wei2, wei3, wei4, wen1, wen2, wen3, wen4, weng1, weng3, weng4, wo1, wo2, wo3, wo4, wu1, wu2, wu3, wu4, xi1, xi2, xi3, xi4, xia1, xia2, xia4, xian1, xian2, xian3, xian4, xiang1, xiang2, xiang3, xiang4, xiao1, xiao2, xiao3, xiao4, xie1, xie2, xie3, xie4, xin1, xin2, xin4, xing1, xing2, xing3, xing4, xiong1, xiong2, xiu1, xiu3, xiu4, xu1, xu2, xu3, xu4, xuan1, xuan2, xuan3, xuan4, xue1, xue2, xue3, xue4, xun1, xun2, xun4, ya1, ya2, ya3, ya4, yan1, yan2, yan3, yan4, yang1, yang2, yang3, yang4, yao1, yao2, yao3, yao4, ye1, ye2, ye3, ye4, yi1, yi2, yi3, yi4, yin1, yin2, yin3, yin4, ying1, ying2, ying3, ying4, yo1, yong1, yong2, yong3, yong4, you1, you2, you3, you4, yu1, yu2, yu3, yu4, yuan1, yuan2, yuan3, yuan4, yue1, yue4, yun1, yun2, yun3, yun4, za1, za2, za3, zai1, zai3, zai4, zan1, zan2, zan3, zan4, zang1, zang3, zang4, zao1, zao2, zao3, zao4, ze2, ze4, zei2, zen3, zen4, zeng1, zeng4, zha1, zha2, zha3, zha4, zhai1, zhai2, zhai3, zhai4, zhan1, zhan3, zhan4, zhang1, zhang3, zhang4, zhao1, zhao2, zhao3, zhao4, zhe1, zhe2, zhe3, zhe4, zhei4, zhen1, zhen3, zhen4, zheng1, zheng3, zheng4, zhi1, zhi2, zhi3, zhi4, zhong1, zhong3, zhong4, zhou1, zhou2, zhou3, zhou4, zhu1, zhu2, zhu3, zhu4, zhua1, zhua3, zhuai1, zhuai3, zhuai4, zhuan1, zhuan3, zhuan4, zhuang1, zhuang3, zhuang4, zhui1, zhui4, zhun1, zhun3, zhuo1, zhuo2, zi1, zi3, zi4, zong1, zong3, zong4, zou1, zou3, zou4, zu1, zu2, zu3, zuan1, zuan3, zuan4, zui1, zui3, zui4, zun1, zun3, zuo1, zuo2, zuo3, zuo4.

B. List of pseudo-syllables in pinyin.

Pseudo-syllables (n=758)	
Segmental gaps (n = 412)	Tonal gaps (n = 356)
bia1, bia2, bia3, bia4, biang1, biang2, biang3, biang4, biong1, biong2, biong3, biong4, biu1, biu2, biu3, biu4, bong1, bong2, bong3, bong4, bou1, bou2, bou3, bou4, cei1, cei2, cei3, cei4, chei1, chei2, chei3, chei4, cua1, cua2, cua3, cua4, chua1, chua2, chua3, chua4, cuai1, cuai2, cuai3, cuai4, cuang1, cuang2, cuang3, cuang4, den1, den2, den3, den4, dia1, dia2, dia3, dia4, diang1, diang2, diang3, diang4, din1, din2, din3, din4, diong1, diong2, diong3, diong4, dua1, dua2, dua3, dua4, duai1, duai2, duai3, duai4, duang1, duang2, duang3, duang4, dv1, dv2, dv3, dv4, due1, due2, due3, due4, eng1, eng2, eng3, eng4, fe1, fe2, fe3, fe4, fai1, fai2, fai3, fai4, fao1, fao2, fao3, fao4, fi1, fi2,	an2, ang3, ben2, ban2, bang2, bei2, biao2, bian2, bin2, bin3, bing2, ce1, ce2, ce3, che2, cen3, cen4, ceng3, ca2, ca4, chai3, cang3, cang4, cong3, cong4, cou1, cou2, cou3, cu3, chuai2, cuan3, cui2, chui3, chui4, chun4, chuo2, chuo3, de1, de3, de4, deng2, dai2, dan2, dang2, dei1, dei2, dei4, dia1, dia2, dia4, diao2, die3, die4, dian2, ding2, diu2, diu3, diu4, dong2, dou2, duan2, dui2, dui3, dun2, ei1, en2, en3, er1, fo1, fo3, fo4, fou1, fou2, fou4, gai2, kai2, gan2, gang2, gao2, gei1, gei2, gei4, geng2, gong2, gou2, gua2, guai2, guan2, guang2, gui2, gun1, gun2, he3, hen1, heng3, hang3, hei2, hei4, hei3, hou1, hua3, huai1, huai3, hun3, jiang2, jian2, jin2, jing2, jiong2, jiong4, jiu2,

fi3, fi4, fia1, fia2, fia3, fia4, fiang1,	juan2, jun2, jun3, ken1, ken2, keng2,
fiang2, fiang3, fiang4, fiao1, fiao2, fiao3,	keng3, keng4, ka2, ka4, kan2, kang3,
fiao4, fie1, fie2, fie3, fie4, fian1, fian2,	kao2, kong2, kou2, ku2, kua2, kuai1,
fian3, fian4, fin1, fin2, fin3, fin4, fing1,	kuai2, kuan2, kuan4, kun2, kuo1, kuo2,
fing2, fing3, fing4, fiong1, fiong2, fiong3,	kuo3, le2, le3, lai1, lai3, lan1, lia1, lia2,
fiong4, fiu1, fiu2, fiu3, fiu4, fong1, fong2,	lia4, liang1, lie2, lian1, lin1, luan1, lun3,
fong3, fong4, gi1, gi2, gi3, gi4, gv1, gv2,	lv1, lv2, lue1, lue2, men3, mai1, mang1,
gv3, gv4, gue1, gue2, gue3, gue4, hi1,	mang4, mei1, mie2, mie3, mian1, min1,
hi2, hi3, hi4, hv1, hv2, hv3, hv4, hue1,	min4, ming1, miu1, miu2, miu3, mou4,
hue2, hue3, hue4, je1, je2, je3, je4,	mu1, ne1, ne3, nen1, nen2, nen3, neng1,
juang1, juang2, juang3, juang4, kei1,	neng3, neng4, nai1, nai2, nang4, nei1,
kei2, kei3, kei4, ki1, ki2, ki3, ki4, kv1,	nei2, niang1, niang3, niao1, niao2, nie2,
kv2, kv3, kv4, kue1, kue2, kue3, kue4,	nie3, nin1, nin3, nin4, ning1, nong1,
len1, len2, len3, len4, liong1, liong2,	nong3, nou1, nou2, nou3, nu1, nuan1,
liong3, liong4, lua1, lua2, lua3, lua4,	nuan2, nuan4, nuo1, nuo3, nv1, nv2,
luai1, luai2, luai3, luai4, luang1, luang2,	nue1, nue2, nue3, ou2, pen3, pa3, pan3,
luang3, luang4, lui1, lui2, lui3, lui4, mia1,	pei3, pie2, pie4, ping3, ping4, pou4, qia2,
mia2, mia3, mia4, miang1, miang2,	qiong1, qiong3, qiong4, qiu4, que3, qun3,
miang3, miang4, miong1, miong2,	qun4, re1, re2, ren1, reng3, reng4, ri1, ri2,
miong3, miong4, mong1, mong2, mong3,	ri3, ran1, ran4, rao1, rong1, rong4, rou1,
mong4, nia1, nia2, nia3, nia4, niong1,	rou3, ru1, ruan1, ruan2, ruan4, rui1, run1,
niong2, niong3, niong4, nun1, nun2,	run2, run3, ruo1, ruo2, ruo3, se1, se2, se3,
nun3, nun4, nua1, nua2, nua3, nua4,	sen2, sen3, sen4, seng2, seng3, seng4, si2,

nuai1, nuai2, nuai3, nuai4, nuang1, nuang2, nuang3, nuang4, nui1, nui2, nui3, nui4, ong1, ong2, ong3, ong4, pia1, pia2, pia3, pia4, piang1, piang2, piang3, piang4, piong1, piong2, piong3, piong4, pong1, pong2, pong3, pong4, piu1, piu2, piu3, piu4, qe1, qe2, qe3, qe4, quang1, quang2, quang3, quang4, ra1, ra2, ra3, ra4, rai1, rai2, rai3, rai4, rei1, rei2, rei3, rei4, rua1, rua2, rua3, rua4, ruai1, ruai2, ruai3, ruai4, ruang1, ruang2, ruang3, ruang4, rv1, rv2, rv3, rv4, sei1, sei2, sei3, sei4, shong1, shong2, shong3, shong4, sua1, sua2, sua3, sua4, suai1, suai2, suai3, suai4, suang1, suang2, suang3, suang4, ten1, ten2, ten3, ten4, tei1, tei2, tei3, tei4, tia1, tia2, tia3, tia4, tiang1, tiang2, tiang3, tiang4, tin1, tin2, tin3, tin4, tiong1, tiong2, tiong3, tiong4, tiu1, tiu2, tiu3, tiu4, tua1, tua2, tua3, tua4, tuai1, tuai2, tuai3, tuai4, tuang1, tuang2, tuang3, tuang4, tv1, tv2, tv3, tv4, tue1, tue2, tue3, tue4, xe1, xe2, xe3, xe4, xuang1, xuang2,	sa2, sha2, sai2, sai3, shai2, san2, sang2, shan2, shang2, sao2, shei1, shei3, shei4, song2, sou2, su3, shua2, shuai2, suan2, suan3, shuan2, shuan3, shuang2, shuang4, shui1, sun2, sun4, shun1, shun2, suo2, suo4, shuo2, shuo3, te1, te2, te3, teng1, teng3, teng4, ta2, tei2, tei3, tei4, tie2, weng2, wai2, xia3, xin3, xiong3, xiong4, xiu2, xun3, yue2, yue3, ze1, ze3, zen1, zen2, zeng2, zeng3, zhen2, zheng2, zi2, za4, zai2, zang2, zhan2, zhang2, zei1, zei3, zei4, zhei1, zhei2, zhei3, zong2, zhong2, zou2, zu4, zhua2, zhua4, zhuai2, zuan2, zhuan2, zhuang2, zui2, zhui2, zhui3, zun2, zun4, zhun2, zhun4, zhuo3, zhuo4

xuang3, xuang4, zua1, zua2, zua3, zua4, zuai1, zuai2, zuai3, zuai4, zuang1, zuang2, zuang3, zuang4	