ABSTRACT

Title of Document:
THE EFFECTS OF PHONOLOGICAL NEIGHBORHOODS ON SPOKEN WORD RECOGNITION IN MANDARIN CHINESE

Pei-Tzu Tsai, Master of Arts, 2007

Directed By:
Professor Nan Bernstein Ratner
Department of Hearing and Speech Sciences

Associate Professor Rochelle Newman
Department of Hearing and Speech Sciences

Spoken word recognition is influenced by words similar to the target word with one phoneme difference (neighbors). In English, words with many neighbors (high neighborhood density) are processed more slowly or less accurately than words with few neighbors. However, little is known about the effects in Mandarin Chinese. The present study examined the effects of neighborhood density and the definition of neighbors in Mandarin Chinese, using an auditory naming task with word sets differing in density levels (high vs. low) and neighbor types (words with neighbors with a nasal final consonant vs. words without such nasal-final neighbors). Results showed an inhibitory effect of high neighborhood density on reaction times and a difference between nasal-final neighbors and vowel-final neighbors. The findings

suggest that neighbors compete and inhibit word access in Mandarin Chinese. Yet, other factors at the sublexical level may also play a role in the process.

THE EFFECTS OF PHONOLOGICAL NEIGHBORHOODS ON SPOKEN WORD
RECOGNITION IN MANDARIN CHINESE

By

Pei-Tzu Tsai

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Arts
2007

Advisory Committee:
Professor Nan Bernstein Ratner, Chair
Associate Professor Rochelle Newman
Assistant Professor Wei Tian

Dedication

This thesis is dedicated to my lovely family overseas, in Taiwan, my dearest grandma, mom, aunts, brother, sister-in-law and my new family in law. Their love is the greatest support of this research journey far away from home. This thesis is also dedicated to my dad, who, I believe, has been watching over me from above. He was the first researcher I knew, whom I followed into a lab for the first time in life.

This is my first work, dedicated to you, dad. Thank you.

Acknowledgements

It is a pleasure to thank my advisors and the members of the committee, Dr. Nan Bernstein Ratner, Dr. Rochelle Newman and Dr. Wei Tian for guiding my through the process with helpful discussions and constructive comments, providing resources and making this thesis possible. I would like to express my deep gratitude to Dr. Ratner, who has been encouraging, inspiring and very patient all along. Special thanks to Dr. Newman for allowing me to use her lab and promptly providing comments and answering all sorts of questions from research to technical issues. My gratitude also goes to her students who oriented me to the equipment at the beginning of the study.

All the participants are thanked for voluntarily making time to visit the lab and completing the study.

I feel grateful to all my friends here for their generous friendship, which supported me through the process of completing this thesis.

Last, but not the least, thanks to my beloved husband, who walked with me from the beginning to the end of this thesis with full support.

Table of Contents

List of Tables

List of Figures

List of Illustrations

Chapter 1: Introduction

When a spoken word is heard, a typical listener can usually recognize the word rapidly. Yet, some words are recognized more "easily" than others. For example, common words are easier to recognize than uncommon words. Recognizing a word means that the information about the word is successfully retrieved from memory. The speed and accuracy of retrieving such information from memory is believed to reflect the word's ease or difficulty of recognition. "Easy" words are recognized more quickly and/or more accurately than "difficult" words. The stored information about words in memory is often referred to as the mental lexicon. Research has focused on exploring the processing of word retrieval from the mental lexicon and the organization of words in the mental lexicon. The ease of word recognition is influenced by the properties of the target word itself, such as semantic concreteness and abstractness (Strain, Patterson, & Seidenberg, 1995), word frequency (Luce & Pisoni, 1998; Oldfield & Wingfield, 1965), word length (Lipinski & Gupta, 2005; Pitt & Samuel, 2006; Vitevitch & Luce, 1999) and also by the properties of similar-sounding words (Luce & Pisoni, 1998).

Most major theories of spoken word recognition propose that when a word is presented, acoustic-phonetic representations of words in the mental lexicon that are structurally similar to the target word are activated. In spoken word processing, speech signals are perceived together with all the background noise and speaker variability (e.g. coarticulation, segment reduction, deletion and so forth), and misperception of words happens. The imperfect, often ambiguous, speech signals

make it unlikely that spoken word recognition is completely based on the direct mapping of the phonetic information to the target lexical item (Luce & Pisoni, 1998). Therefore, rather than directly activating the target lexical item from perceived phonetic information, the processing system activates multiple words with similar sound patterns to provide choices upon speech signal input. It is proposed that the competition from these similar words is involved during lexical access (i.e., the activation-competition model of lexical access). This group of similar words is referred to by varying terms, such as cohort, competitors and neighbors, depending on the definition of similarity and the detailed process of word identification based on different theories (e.g. the Cohort Theory [Marslen-Wilson & Welsh, 1978], TRACE model [McClelland & Elman, 1986], MERGE model [Norris, McQueen, & Cutler, 2000], and Neighborhood Activation Model [Luce & Pisoni, 1998]).

The structural organization of the mental lexicon would be expected to have an influence on spoken word recognition because of the competition and identification process. The relative ease or difficulty of word identification would be influenced by the properties of the target word's competitors, and thus, it would reflect the structural organization of lexical items in the mental lexicon. However, few theories address how lexical items are organized in the mental lexicon. For example, Cohort Theory (Marslen-Wilson & Welsh, 1978) defines similarity by shared onsets (words sharing onsets are known as a cohort). It is hypothesized that acoustic-phonetic input is sequential in time, and that word recognition occurs when the target word is discriminated from its cohort at the point where it differs segmentally from others. This theory defines similarity among words and addresses

the temporal aspect of the word recognition process. Yet, it does not explicitly address the structural organization of lexical items in mental lexicon; moreover, it does not assume effects from the overall cohort structure on lexical access. The Neighborhood Activation Model (NAM) (Luce & Pisoni, 1998) explicitly addresses structural organization by proposing that words with similar sound patterns (i.e. neighbors) are stored close together and form a neighborhood. Neighbors are commonly defined as words that differ from the target word by one phoneme in any position rather than only initially, including substitution, addition and deletion. A word with many neighbors is said to be located in a dense neighborhood, whereas a word with few neighbors is located in a sparse neighborhood. The NAM model takes into account the effects of the neighborhood structure on spoken word recognition.

Phonological Neighborhood

Lexical items can be organized based on similarity at different levels of lexical information, from higher-level semantic, conceptual and morphological features to lower-level acoustic-phonetic information. With the assumption that acoustic input activates representations from lower to higher (i.e. bottom-up priority), the low-level acoustic-phonetic stage represents one of the earlier stages at which structural organization could exert an influence. Thus, the NAM hypothesizes that lexical items are organized based on their similarity at the acoustic-phonetic level, forming lexical (phonological) neighborhoods, and assumes that structural relations among lexical items at the acoustic-phonetic level would effect the ease of word recognition through the process of discrimination.

There are several ways to define phonological neighbors. The most common definition is the one-phoneme difference; that is, neighbors are words that differ from a target word in one phoneme, including substitution, deletion and addition (Luce & Pisoni, 1998). For example, a target word /kæt/ ("cat") would have neighbors such as /bæt/ ("bat"), /æt/ ("at"), /kaɪt/ ("kite"), /kæʃ/ ("cash", "cache"), /kæst/ ("cast"), /kætɪ/ ("catty"), and so forth. A modified definition of neighbor expands the one-phoneme difference from words matching two out of three phonemes in CVC words to words matching two-thirds of the phonemes in longer words (i.e., neighbors are words matching 66% of the phonemes within the target word) (Frisch, Large, & Pisoni, 2000).

The one-phoneme difference rule for defining neighbors would suggest that all neighbors are equally different from the target word. For example, words differing in onset phoneme would have the same value as words differing in final phoneme (e.g., for the target word "bat", rime neighbor "mat" and onset neighbor "bang" are equal neighbors); words differing by a phoneme substitution are equivalent to words differing in phoneme deletion/addition (e.g. for the target word "bat", substitution neighbor "b_i_t" is the same as deletion neighbor "at" and addition neighbor "battle"); and finally, phoneme difference with shared phonetic features would be equivalent to phoneme difference with different phonetic features (e.g. /p/-/b/ substitution "pat" and /s/-/b/ substitution "sat" are equal neighbors of "bat"). However, according to linguistic theory, phonemes differ on the basis of a small set of phonetic features,

such as place of articulation, manner of articulation, voicing, and so forth. Therefore, these features have been used not only to characterize phonemes but also to measure phoneme similarity. The more features the two phonemes share (or do not share) could influence the phoneme's confusability, and thus influence a word's confusability in spoken word recognition (Bailey & Hahn, 2005; Hahn & Bailey, 2005). Vowels and consonants have also been found to differ in their influence on lexical access; an altered consonant appeared to have a greater influence on lexical access than an altered vowel (van Ooijen, 1996).

Lexical neighborhood is generally discussed in terms of two measures, the number of neighbors (neighborhood density) and the average frequency of neighbors (neighborhood frequency). For example, "cat" has higher neighborhood density than "elephant" because the former has 35 neighbors (e.g. "at" "bat" "that" "mat" and "cattle") while the latter only has one neighbor ("element") (see Illustration 1); "forth" has higher neighborhood frequency than "daily" because the former has a log-adjusted neighborhood frequency of 2.95, while the latter has a log-adjusted neighborhood frequency of 1.35 (see Illustration 2). Since English words differ substantially along these two dimensions, it has been hypothesized that they would mediate spoken word recognition (Luce & Pisoni, 1998).
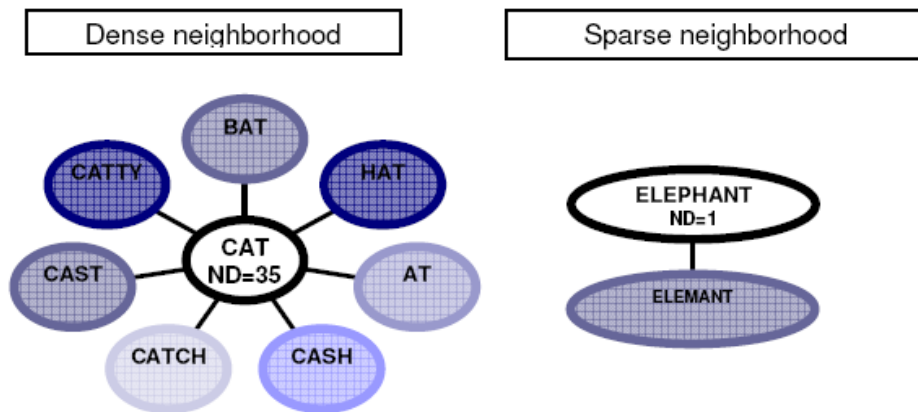
Illustration 1. Neighborhood density
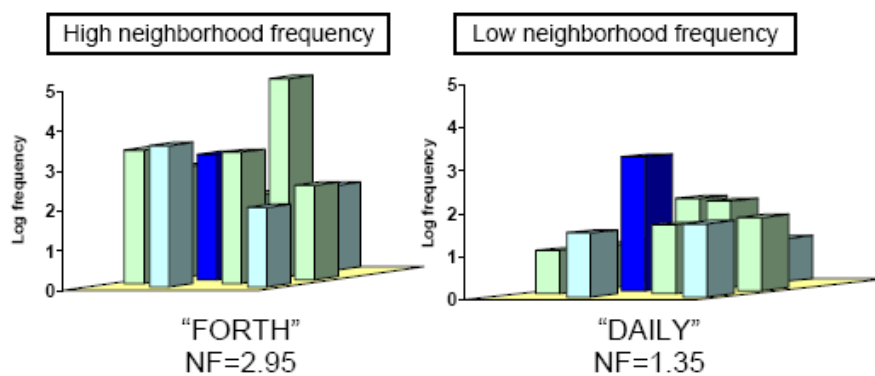
Illustration 2. Neighborhood frequency

The effects of lexical neighborhoods on word access have been examined using various methodologies, such as lexical decision tasks, auditory naming tasks, same-difference matching tasks, and phoneme identification tasks. In a lexical decision task, participants decide whether the perceived target stimulus is a real word or not (i.e. lexical status of the stimulus) as quickly as possible. The auditory naming task, also known as a repetition task, is a speeded task in which participants verbally

repeat the perceived target stimulus; word or non-word repetition have been used for different purposes. In a same-different matching task, participants respond to pairs of stimuli by reporting, usually by button/key press, whether the presented stimuli are the same or different. In a phoneme identification task, participants listen to a series of words and identify the target phonemes, typically involving words in noise or words with a phonetically ambiguous segment. These tasks vary in their demands, with lexical decision involving explicit decision making on lexical status, auditory naming requiring precise identification of acoustic information for verbal production, same-difference matching and phoneme identification involving decision making on acoustic-phonetic information. These tasks have been used in combination to examine the role of lexical neighborhood in spoken word recognition across modalities.

Neighborhood density is predicted to have an inhibitory effect on spoken word recognition (i.e., words with high neighborhood density are more "difficult" to recognize than words with low neighborhood density), based on the assumption that neighbors of a target word are activated upon stimulus presentation and would compete with the target word in the process of identification for lexical access. Studies examining the effects of neighborhood density on spoken word recognition in typical English-speaking adults have shown the predicted pattern of inhibitory effects. Inhibitory effects of neighborhood density or frequency-weighted neighborhood density, which takes into account the effects of frequencies of neighbors, have been found not only in the speed of response (i.e. response latency or reaction time) but also in accuracy rate; that is, words with higher neighborhood density, or frequency-weighted neighborhood density, are responded to more slowly and less accurately

7

than words with lower neighborhood density. Neighborhood density also influences phonetic perception in non-words, with a shift of phonetic perception towards non-words with high neighborhood density. The effects of phonological neighborhood on word recognition and phonetic perception have been found in auditory word naming tasks (Luce & Pisoni, 1998; Vitevitch & Luce, 1998), auditory lexical decision tasks (Luce & Pisoni, 1998; Vitevitch & Luce, 1999), same-different judgment tasks (Vitevitch & Luce, 1999) and phoneme identification tasks (Newman, Sawusch, & Luce, 1997; Newman, Sawusch, & Luce, 2005). Yet, a neighborhood density effect pattern is not consistently found in studies conducted in other languages and with non-words. Studies have found either opposite or no effect of neighborhood density on non-word repetition (Lipinski & Gupta, 2005; Vitevitch & Luce, 1999), opposite effects of neighborhood density in Spanish (Vitevitch & Rodriquez, 2005), and inconsistent effects of neighborhood density in Japanese (Amano & Kondo, 2000). The details of these studies will be discussed later.

Neighborhood frequency, similar to neighborhood density, has a predicted inhibitory effect on spoken word recognition also as a result of competition between neighbors and the target word in lexical access. If neighbors are more common, they will compete more strongly with the target word. In typical English-speaking adults, inhibitory effects of neighborhood frequency on spoken word recognition have been found in auditory lexical decision tasks, but not in auditory naming tasks (Luce & Pisoni, 1998). It was suggested that task-specific and higher-level process demands might be responsible for this difference. Lexical decision is biased by lexical

frequency because it involves higher-level information for decision making, whereas auditory naming response only requires isolation of acoustic-phonetic representation.

Additional Factors in Lexical Access

Besides neighborhood effects, several other lexical properties also have important effects on lexical access. The frequency with which word occurs in a language, and thus presumably is encountered in lexical access (word frequency) is one strong factor influencing lexical access. Studies show that high frequency words are responded to faster and more accurately than low frequency words in a lexical decision task, suggesting that word frequency biases the processing system towards choosing high frequency words over low frequency words (Dupoux & Mehler, 1990; Luce & Pisoni, 1998; Oldfield & Wingfield, 1965). Yet, similar to neighborhood frequency, no word frequency effect has been found in reaction time in auditory naming tasks. Under the interactive activation theory in lexical access, activation of representations spreads in both directions between levels. If word frequency affects the access of a target word, it should also affect the activation of its neighbors. Thus, the lack of both word frequency effects and neighborhood frequency effects in auditory repetition tasks is likely the result of task demand. Response generation in a repetition task does not involve lexical status decision making or require higher-level lexical information to optimize performance of the processing system (Luce & Pisoni, 1998).

Age of acquisition is a factor highly correlated with word frequency; that is, common words are generally acquired earlier in life. Some suggest that words learned early in life are processed more quickly than late-learned words, and argue that

9

frequency effects are actually effects of age of acquisition. Age of acquisition was found to be a better predictor of the speed of word access than word frequency using regression analyses (Morrison & Ellis, 1995), while others suggest that both factors have independent effects on spoken word processing (Ellis & Morrison, 1998; Gerhand & Barry, 1999).

A sublexical property that has been separated from lexical neighborhood in word recognition is the relative frequencies of segments or sequences of segments in a word, its phonotactic probability. A word with high phonotactic probability contains segments or sequences of segments occurring frequently in other words in the language. This generally indicates that such a word has many similarly sounding neighbors and is high in neighborhood density. It is proposed that the neighborhood density effect has its locus at the lexical level, which is the competition among lexical items for lexical access, and that the phonotactic probability effect has its locus at the sublexical level, which facilitates sublexical representation activation and articulation planning. Opposite effects of the two factors on spoken word recognition have been found at the two hypothesized levels of processing, an inhibitory effect of neighborhood density with the primary level of processing at the lexical level and a facilitative effect of phonotactic probability with the primary level of processing at the sublexical level (Vitevitch & Luce, 1998, 1999, 2005; Pÿlkkanen, Stringfellow, & Marantz, 2002).

Word length is another factor mediating spoken word processing. Word length can be measured by stimulus duration or by the amount of information conveyed in a lexical item such as number of phonemes or syllables. Long word length has found to

be inhibitory in spoken word recognition response time (Yoneyama, 2002), but facilitative in lexical activation as a result of decreased competition in long words (Pitt & Samuel, 2006). Word length may potentially influence processing strategies, with a more important role of phonotactic probability in processing long words than short words (Vitevitch & Luce, 1999). Furthermore, lexical activation of short words appears easier to manipulate than long words through task instructions and task demands (Pitt & Samuel, 2006).

Nature of Neighbors

A large amount of research has supported the effects of lexical neighborhood using the rough estimation of neighbors, considering neighbors to be words with one phoneme difference from the target. Studies have further examined the nature of neighbors and neighborhood structure. Neighborhoods may differ not only by density and frequency but also by the relations among neighbors. For example, majority of phonological neighbors in English are neighbors sharing rimes. Some have speculated that there might be differential influences from different types of neighbors on spoken word recognition (De Cara & Goswami, 2002).

Studies have looked at the definition of structural similarity among word forms (the nature of the competitor set). As the Cohort Theory placed emphasis on words sharing the same onsets (Marslen-Wilson & Welsh, 1978), the definition of cohort has been examined in the context of neighborhood structure. It has been found that neighbor (using the one-phoneme difference definition) serves as a better definition for similarity structure than cohort (words sharing onsets), and that words sharing initial phonemes do not exert a greater effect compared to other neighbors

(Amano & Kondo, 2000; Newman, Sawusch, & Luce, 2005). Onset and post-onset neighbors may differ in the time course of representation activation, but not in the overall activation for competition and identification in lexical access (with neighbors sharing the same onset being activated earlier than neighbors differing at the onset phoneme) (Newman, Sawusch, & Luce, 2005).

A different measure of neighborhood structure was proposed by Vitevitch (2007), by calculating the number of possible phoneme position differences among neighbors (spread). For example, "mop" and "mob" have the same neighborhood density, but they have different spreads of neighborhood, as "mop" has neighbors that differ in any of the three phoneme positions, spread of 3 (e.g. "<u>h</u>op" "m<u>a</u>p" and "mo<u>ck</u>"), and "mob" has neighbors that differ in only two of the three phoneme positions, spread of 2 (e.g. "<u>r</u>ob" and "mo<u>ck</u>"). By matching words in overall neighborhood density, the study found inhibitory effect of neighborhood spreads; words with greater spread were responded to more slowly than words with smaller spread.

Bashford, Warren and Lenz (2006) examined the definition of neighbor and neighborhood structure by *generating* neighbors rather than examining pre-defined neighbors. In the study, participants listened to the same stimulus repeatedly. In this situation, listeners frequently experience an illusory change in the stimulus, known as the verbal transformation effect. For example, hearing "police" repeated, listeners may report hearing words such as "please" "fleece" or "fleas". In the study, listeners were instructed to call out the perceived word when they heard a change in stimulus. Monosyllabic target words were played over and over until illusionary lexical items

12

were perceived secondary to the adaptation to salient phonological features. The assumption was that the competitor set of a target word would be perceived following the verbal transformation effect. The findings were consistent with the one-phoneme difference definition for words with high frequency-weighted neighborhood density (that is, more lexical neighbors were reported than non-neighbors following the verbal transformation effect), but not for words with low frequency-weighted neighborhood density (that is, listeners reported more lexical words that differed by two or more phonemes than words with one phoneme difference in the low frequency-weighted neighborhood density condition), suggesting that words with more than one-phoneme difference from the target word could also be activated in the word recognition process.

Neighbors of a target word differing in any position contribute to the effects of neighborhood. The number of possible positional difference from the target in a neighborhood also plays a role in competition and identification during lexical access. Further, there was no evidence showing greater influence from onset neighbors than from other neighbors on lexical access. Compared to cohort, the one-phoneme difference definition of neighbors has been relatively successful in English without specifying the position of phoneme difference, but it remains unclear whether the type of difference (i.e. addition, deletion and substitution) plays a role in neighborhood structure.

Neighborhood Effects across Languages

With numerous research findings supporting the effects of lexical neighborhood in English, researchers have looked at effects in other languages. So

far, patterns of neighborhood effects in spoken word recognition were found to be similar in French (Ziegler, Muneaux, & Grainger, 2003), but varied in two other languages examined.

In a study on phonological neighborhood in Spanish using an auditory lexical decision task, findings showed that phonological neighborhoods have an effect on word recognition in Spanish, but in the opposite directions from that found in English; neighborhood density and neighborhood frequency were found to be facilitative in spoken word recognition in Spanish. Words with higher neighborhood density were recognized more quickly and accurately than words with lower neighborhood density. The same was found for neighborhood frequency; words with higher neighborhood frequency were responded to more quickly and accurately than words with lower neighborhood frequency. The authors suggest that linguistic system differences may play a role in word processing strategy. Spanish is dominated by two- to three-syllable words, while English has more one- to two-syllable words. As noted earlier, research in English showed some indication of different processing strategies for short and long words (Vitevitch & Luce, 1999). With the difference in word length, it is suspected that the typical word processing strategy used in Spanish speakers may vary from that of English speakers, and that neighborhood density plays a different role in spoken word recognition in Spanish (Vitevitch & Rodriguez, 2005).

Neighborhood effects have also been examined in Japanese word recognition, using speeded lexical decision and unspeeded word identification in noise that required participants to type the word they heard. An effect of neighborhood density was found only in the unspeeded word identification. This effect was inhibitory on

response accuracy. No effect was found in the speeded lexical decision task (Amano & Kondo, 2000). In another study, neighborhood density had an inhibitory effect on response accuracy in a speeded task. A consistent facilitative effect of neighborhood density was found in reaction time across tasks, with many responses initiated prior to stimulus offset. The inhibitory effect of neighborhood density on reaction time, which is typical in English, was only found in a semantic categorization task in Japanese, with neighbors defined by similarity in auditory features (Yoneyama, 2002). The lack of inhibitory neighborhood effect in reaction time using the one-phoneme difference and the relatively fast reaction time seen in speeded naming tasks could suggest that adult Japanese speakers might not have fully accessed the lexical representations at the time of response generation; instead, participants might have repeated words with similar strategies used in non-word repetition, with a more important role of phonotactic probability in fast naming. Since the word recognition processing is believed to be efficient, it is likely that speakers of different languages would show varying word recognition strategies as a result of the varying structures among languages, such as long word length in Spanish, pitch accent information across moras in Japanese words, simple syllable structure and the lexical tone associated with each syllable in Mandarin, and so forth.

In summary, both neighborhood density and neighborhood frequency have shown effects on spoken word recognition in English (Luce & Pidoni, 1998; Newman, Sawusch & Luce, 1997; Newman, Sawusch, & Luce, 2005; Vitevitch & Luce, 1998, 1999). Words with high neighborhood density generally show more competition in lexical access compared to words with low neighborhood density

and/or low neighborhood frequency. However, it was suggested that task demand difference mediates the presence of frequency bias, with lexical decision task requiring higher-level lexical information and auditory naming task requiring lexical access yet not higher-level lexical information (Luce & Pisoni, 1998). It was found that neighbors and neighborhoods may differ in nature; that is, the one-phoneme difference definition might not include all neighbors (Bashford, Warren, & Lenz, 2006), and there are other ways to measure neighborhoods (Vitevitch, 2007). Furthermore, lexical and neighborhood properties may vary among languages, and languages that differ substantially from English in certain linguistic aspects can result in different strategies used by their speakers in spoken word recognition, as neighborhood showed different effects among languages (Amano & Kondo, 2000; Vitevitch & Rodriguez, 2005; Yoneyama, 2002). Lexical neighborhood has been well supported as a factor in word recognition in English, and examining this factor in other languages may facilitate our understanding of the nature of lexical neighborhood, its relation with language structures and spoken word processing strategies, especially in a language that differs from English in many aspects. One such language that would be worth investigating is Mandarin Chinese.

Mandarin Chinese Structure

Mandarin Chinese is a language composed entirely of monosyllabic morphemes. The language has a limited number of legal syllables (approximately four hundred without counting tonal contrasts). Each monosyllable has a simple structure consisting of a single consonant or zero consonant (called an initial), followed by a rime (called a final), and associated with a tone. The rime includes a

nucleus and a possible coda. The syllable nucleus is a set of one to three vowel(s), and the coda is one of the two legal final consonants in Mandarin Chinese, /n/ and /ŋ/. Mandarin has 22 initial consonants, 22 vowel finals, and 15 nasal finals (Li & Thompson, 2003). The tone system includes four relative, contrastive pitch patterns. Tone 1 is a high-level tone, Tone 2 is a high-rising tone, Tone 3 is a dipping tone, and Tone 4 is a high-falling tone. Tone, similar to phoneme, contrasts meanings in Mandarin Chinese (a concept known as lexical tone). For example, "ma1" and "ma3" are words with the same phoneme combination but different tones, and they differ in meanings ("mother" and "horse" respectively).

The morpheme is an abstract unit, the smallest meaningful element in a language (Li & Thompson, 2003). In Mandarin, each phonological syllable corresponds to several tone-specific syllables, with the number of legal tones varying among syllables (e.g., the syllable /hua/ is associated with Tone 1, Tone 2 and Tone 4, whereas syllable /sə/ is associated with Tone 4 only). Each tone-specific syllable corresponds to several morphemes (homophones), resulting in many morphemes sharing one syllable. For example, the tone-general syllable "shi" (/ʂɨ/) takes four tones, with each tone-specified syllable corresponding to several homophones: "shi1" (詩 "poem," 濕 "wet," 師 "teacher," etc.), "shi2" (石 "stone," 實 "solid," 十 "ten," etc.), "shi3" (使 "make," 史 "history," 始 "start," etc.) and "shi4" (是 "yes," 市 "market," 事 "event," etc.). Though morphemes are roughly related to characters, many characters have multiple semantic-syntactic functions and represent different

17

morphemes. For example, 信 ("xin4" /ɕɪn/) can be a noun and a verb, meaning

"letter" and "trust" In addition, the general Chinese orthographic system is shared

among dialects with different phonological systems (e.g., Mandarin Chinese and

Cantonese). The orthographic system itself includes two subsystems, the simplified

and the traditional Chinese characters. Some other dialects lack a mature writing

system. For example, Taiwanese shares characters partially with Mandarin but no

mature writing system is currently in use, and government policy aims to revive the

"language" (dialect) by developing its unique writing system (Hsiau, 1997)

Mandarin Neighborhood Studies

There has been substantial lexical neighborhood research conducted in

Chinese, but the primary focus has been on orthographic neighborhoods. In contrast

to the alphabetic writing system of English, Chinese is a logographic system with

each character composed of different portions called radicals. A semantic radical

reflects some level of semantic information of the word, and a phonetic radical

reflects some level of information about the word's pronunciation. For example, the

characters "媽" (meaning "mother") and "姐" (meaning "sister") both have a

semantic radical "女" in the left portion of the characters, meaning "female"; the

word "媽" (pronounced "ma1") has a phonetic radical of "馬" (pronounced "ma3")

while the character "姐" ("pronounced "jie3") has a phonetic radical "且"

(pronounced "qie3"). About 97% of characters are composed of a semantic radical

and a phonetic radical (DeFrancis, 1989). When the pronunciation of a character is

consistent with its phonetic radical, it is a regular character (e.g. the character "媽",

pronounced "ma1" with a phonetic radical "ma3"). When the pronunciation is

inconsistent with its phonetic radical, it is an irregular character (e.g. "抽",

pronounced "chou1" with a phonetic radical "you2". Each character corresponds to

one, or occasionally several, phonetic syllable(s). Studies on Chinese orthographic

neighborhoods define neighbors as words sharing one of the two constituents with the

target word. Thus, neighbors of a single phonetic-compound character share one of

the two radicals with the target word and neighbors of a two-character compound

word share one of the two characters with the target word. It was found that neighbors

sharing the same first character (similar to the cohort definition of neighbors in

English) have a stronger influence on access than neighbors sharing the second

character. Different patterns of neighborhood density effects were found in one-

character and two-character compound word processing, with inhibitory effects on

single-character word recognition (Bi, Hu, & Weng, 2006) but facilitative effect on

two-character word recognition (Huang et al., 2006; Tsai, Lee, Lin, Tzeng & Hung,

2006). The effect of relative neighborhood frequency to a target word was inhibitory

on two-character word recognition (Huang et al., 2006).

In English, there is an overall facilitative orthographic neighborhood effect on

word recognition (Andrews, 1989; Peereman & Content, 1995). Yet, it has been

argued that orthographic neighborhood is an artifact of phonological neighborhood

because of the positive correlation between orthography and phonology in English. A

recent study showed only a phonological neighborhood effect and no orthographic

neighborhood effect in English using a read-aloud task (Mulatti, Reynolds, & Besner,

2006). In contrast with English, emerging evidence in recent years shows that

Chinese visual word recognition is influenced more by semantic and graphic factors

and less so by phonological factors (Chen & Shu, 2001; Li, Liu & Shu, 2006; Wu & Chen, 2000; Wu & Chou, 2000), suggesting that the orthographic neighborhood effects found in Chinese might not be related to phonology. The role of phonological neighborhood in Mandarin spoken word recognition remains unclear.

Lexical Access in Mandarin

Although little is known about the role of phonological neighborhoods in Mandarin Chinese spoken word recognition, the few neighborhood studies and other studies on lexical access in spoken word processing could provide some information about the phonological structure in the Mandarin mental lexicon.

Studies examining the processing of homophones showed no effect of tone density on spoken word recognition in a lexical decision task and gating experiment in Mandarin Chinese and Cantonese; that is, syllables allowing more tones were not responded to differently from syllables allowing fewer tones (Yip, 2000; Zeng & Mattys, 2004). Studies also suggest that tone information is used later in processing than segmental information, as tone mismatch was responded to more slowly and less accurately than segment mismatch in lexical decision tasks, same-different matching tasks and monitoring tasks (Cutler & Chen, 1997; Ye & Connine, 1999). In studies examining the time course of spoken word recognition using priming tasks, it was found that primes with matching tone and mismatching syllable did not cause any interference in naming tasks; in fact, none of the sublexical phonological components, including syllable onset, rime, rime plus tone, or tone alone primes, showed priming effects in Mandarin Chinese speech production. It has been suggested that in Mandarin Chinese, syllable, rather than segment, is the stored unit or the most

efficient processing unit in the mental lexicon (Chen, Chen, & Dell, 2002; Zhang & Yang, 2004, 2005; Zhou & Zhuang, 2000).

Zeng and Mattys (2004) examined morpheme neighbors that are phonologically similar yet not necessarily semantically related in Mandarin Chinese. Their study defined morphemes by their written forms, that is, each character is a separate morpheme. They used a lexical decision task, with two definitions of morpheme neighbors: 1) tone-general morphemes neighbors are morphemes sharing the same syllable (phonological word form) with the target morpheme, but they may differ in tone, or 2) tone-specific morpheme neighbors are characters sharing the same syllable and the same tone with the target morpheme (homophonic morphemes). Using the example of syllable "shi" mentioned earlier, all characters with the same word form of "shi1" are tone-specific (or homophonic) morpheme neighbors of each other, but all characters with the same word form of "shi" (without tonal contrasts), including "shi1" "shi2" "shi3" "shi4", are tone-general morpheme neighbors of each other. The authors found a reliable inhibitory morpheme density effect under both definitions of neighbors and no tone density effect; that is, words with low morpheme density were responded to more quickly and accurately than words with high morpheme density, suggesting that morphemes sharing the same syllable compete for lexical access, and such discrimination is not affected by the number of legal tones associated with the syllable. However, this inhibitory effect of morpheme density does not address lexical identification processing at the acoustic-phonetic level, because all morpheme neighbors share the same phonological syllable structure. Since information at the phonological level is not sufficient for identifying a target

morpheme from a group of morpheme neighbors sharing the same syllable, their study would imply that higher-level information is required to identify a specific lexical item (lexical access). Their finding does, however, support that lexical items are organized at the phonological level with morphemes grouped in terms of the same phonological syllable structure. Even though their study assumed a phonologically-based organization of lexical items in the mental lexicon, the authors did not control for the phonological neighborhoods of the selected syllable stimuli. With the bottom-up priority assumption for spoken word recognition, the processing system would first activate phonological representations prior to activation of morphemic representations, and discrimination among phonological neighbors would be part of the process. In addition, their definition of morpheme suggests that morphemic representations are equivalent to, or at least correlated to orthographic representations. Orthography could potentially influence the spoken word processing because some morphemes/characters could be consistent or inconsistent between their pronunciation and their phonetic radicals (regular or irregular characters). Some morpheme neighborhoods might have more morphemes with regular characters than others.

So far, our understanding of the organization of lexical items (or morphemes) in Mandarin is vague at the phonological level (simply suggesting that morphemes with the same syllable are stored close together). The finding that morpheme density (the number of morphemes sharing the same tone-general syllable) has an inhibitory effect on spoken word recognition implies that identifying a target word from its morpheme neighbors requires higher-level lexical information. Further, tone

neighborhood does not appear to influence this process (Yip, 2000; Zeng & Mattys, 2004). The findings are weakened by the fact that, even though organization is assumed at the phonological level, phonological neighborhoods of the selected syllables were not addressed, nor controlled in prior research. In terms of spoken word processing, studies have suggested that syllabic representations are more important than segmental or tonal representations in lexical access (Chen, Chen, & Dell, 2002; Zhang & Yang, 2004, 2005; Zhou & Zhuang, 2000). In addition, there have been inconsistent findings for visual word processing of single word and compound words (Bi, Hu, & Weng, 2006; Huang et al., 2006; Tsai, Lee, Lin, Tzeng, & Hung, 2006), which may suggest a potential difference in processing strategies, similar to the finding in English that different strategies may be used for processing short and long spoken words (Vitevitch & Luce, 1999).

Without information on the effects of phonological neighborhood on Mandarin Chinese spoken word recognition, all studies on Mandarin Chinese spoken word processing naturally can not take phonological neighbors into account. Yet, with the clear effect of phonological neighborhood on spoken word recognition in English, it is important to examine this factor. If a similarly consistent neighborhood effect exists in Mandarin Chinese, it should be controlled as with any other lexical factor in studies of spoken word recognition.

The purpose of the present study is to examine the effects of phonological neighborhoods and the definition of phonological neighbors in Mandarin Chinese. Words varying in neighborhood density and words with neighbors varying in types were constructed to address these questions. Based on the large evidence found in

English and some supportive findings in other languages for the effect of neighborhood on spoken word recognition, it is hypothesized that an inhibitory neighborhood density effect exists in spoken word processing in Mandarin as well. Also, based on the common definition of neighbors as words with one-phoneme difference from the target word, it is hypothesized that types of neighbors are functionally the same (that CVC and CV syllables are both equivalent neighbors to a CV target). Therefore, it is predicted that syllables with high neighborhood density will be responded to more slowly and/or less accurately than syllables with low neighborhood density. It is also hypothesized that there will be no difference in neighborhood density effect between vowel-final neighbor mismatch and nasal-final neighbor mismatch, when the size of density difference is matched between the two pairs of high and low density word sets.

Chapter 2: Method

Participants

Twenty-six adult native speakers of Taiwan Mandarin Chinese (TMC), 14 male and 12 female, with a mean age of 31 (range: 27 - 56) participated in the study. All participants were recruited from the University of Maryland at College Park campus area through a Taiwanese graduate student mail list, posting on Taiwanese graduate student association websites and flyer distribution at Taiwanese student events. The 26 selected participants were originally from Taiwan, with their first language being either TMC or Taiwanese, and had spoken mainly TMC since elementary education. All participants completed all or part of their college education in Taiwan before coming to the United States. All participants reported that their longest place of residency is Taiwan. The daily use of language averaged 60% TMC, 34% English and 6% Taiwanese. All participants were free of known hearing and/or speech-language disorders.

Stimuli

To determine Mandarin Chinese phonological neighborhoods, the *Guoyu Cidian Jianbianben Bianjitzliau Tztspin Tungji Baugau (Word Frequency Statistic Report of the Database for National Language Concise Lexicon)* (Jiauyubu Guoyu Cidian Jianbianben Bianji Shiautzu [Ministry of Education National Language Concise Lexicon Editorial Team], 1997) along with the Frequency Statistics of the

Academia Sinica Balanced Corpus of Modern Chinese (Cheng, et al., 2005; Academia Sinica, 1997) were used. The word frequency report provides statistical analysis of a database that contains approximately two million words collected from written materials of magazine, books and newspapers, covering various topics for adults and children. The Academia Sinica Balanced Corpus is an online database which contains about five million words, tagged with part-of-speech, across various topics and sources, and allows frequency search of individual words.

Prior to determining phonological neighbors based on the definition of a one-phoneme difference, it would be necessary to first define phoneme. A phoneme is the underlying representation of speech sounds, commonly defined as the minimal unit that contrasts meanings between words. The phonological system of Mandarin Chinese creates a challenge for determining vowel phonemes. The limited number of legal monosyllable structures in Mandarin Chinese leaves little opportunity for contrast and comparison among vowel environments, allowing a large number of vowel allophones. For example, there are about five mid-vowels, which are phonetically distinct but in complementary distribution in the language; linguists have proposed a single underlying representation for all those mid-vowels, but have not agreed on the representing phoneme and the collection of its allophones (Wan & Jaeger, 2003). As Mandarin Chinese has several regional dialects/accents (e.g., Beijing Mandarin and Taiwan Mandarin), the present study focuses on Taiwan Mandarin Chinese (TMC) to control for potential phonetic-articulatory differences among dialects. Wan and Jaeger (W&J) (2003) examined TMC vowel phonemes

through speech errors and distribution analysis, and proposed the following five-vowel system for TMC:

/i/ → [i, ɨ, j]
/y/ → [y, ɥ]
/u/ → [u, w]
/ə/ → [e, ɛ, ə, ɔ, o, ɤ]
/a/ → [a, ɑ]

The W&J five-vowel system of TMC was adopted but modified for the purpose of this study, which will be discussed later. A combination of two to three vowels (e.g. /uaɪ/) at the nucleus position was considered as one vowel unit for the calculation of neighbors in this study, as multiple vowels turn into diphthongs or glides in a single syllable, similar to diphthongs in English (e.g. /aɪ/). Thus, /ou/, /au/, /uaɪ/, /ueɪ/, /iau/ are all single phonemes.

The word frequency report provides the frequency for each tone-general syllable (the syllable without considering tonal contrasts), and unless specified, the term *syllable* in this section refers to tone-general syllable. Each syllable is a combination of an initial and a final represented in Mandarin Phonetic Symbol I (MPS-I), a widely used phonetic system to facilitate literacy and pronunciation in early elementary education in Taiwan. In order to calculate phonological neighbors, MPS-I in the frequency table were first transcribed into International Phonetic Alphabet (IPA), including initial consonant, vowel unit (vowel/s) and final nasal consonant; this set of transcription represents the surface form; the vowel units were then coded into the underlying representations based on W&J's five-vowel system.

27

However, Wan & Jaeger's five vowel system was adopted and modified based on one criterion: when the representation of a vowel unit based on W&J did not distinguish between two separate MPS-I vowel units, the vowel transcription based on MPS-I is retained as a separate vowel representation. For example, [a] (in vowel combinations only) and [ɑ] (as single vowel only) are in complimentary distribution, and transcribing both into /a/ based on W&J would still distinguish the vowel units among the MPS-I symbols (e.g. "ㄚ" [ɑ], "ㄠ" [au] and "ㄞ" [aɪ]), but transcribing [o] and [ə] into /ə/ would not distinguish between two vowel units of the MPS-I symbols "ㄛ" [o] and "ㄜ" [ə], and transcribing [i] and [ɨ] into /i/ would not distinguish between the "empty rime (no symbol)" [ɨ] and "一" [i]. The modification discussed above is used to compromise between linguistic and phonetic-articulatory characteristics, avoiding large number of vowel allophones. As a result of the modification, the present study defines /i, ɨ, y, e, ɛ, ə, ɔ, o, u, a/ as separate vowels. The five-vowel system and the modified 10 vowel system appear different, but when used to define neighbors, the two systems showed no difference for 90% of the stimuli selected in this study.

Finally, neighbors in Mandarin Chinese in this study are defined as syllables differing from the target syllable in one of the three phoneme positions (initial, nucleus and coda). Neighborhood density (ND) is the number of neighbors, and neighborhood frequency (NF) is the mean syllable frequency of neighbors. All

frequencies in the study were computed into log-adjusted values (log of 10 times the syllable frequency per one million words); this transformation avoids a log value of zero for words occurring once or less per one million words. ND and NF were calculated for all syllables.

Final Materials

Ninety-one CV syllables were selected to form three pairs of word sets that contrasted in neighborhood densities (see Table 1 for summary). Pair 1 was the nasal-final mismatch condition. This pair of word sets tested neighborhood density using word sets that did and did not have neighbors with a nasal final consonant (nasal-final neighbors). In this condition, the high and low word sets matched in the number of neighbors with a vowel final (vowel-final neighbors) within .35 (average vowel-final neighborhood density of 26.29 and 25.94 respectively), but the high density word set also had nasal-final neighbors in addition, while the low density word set did not. The two word sets thus mismatched in the number of nasal-final neighbors. For example, the syllable "lu" /lu/ has no nasal-final neighbor (/lun/ and /luŋ/ are not legal syllables); in contrast "li" /li/ has nasal-final neighbors, because "lin" /lin/ and "ling" /liŋ/ are both legal syllables. This pair of word sets contained high and low density word sets of 17 items each. Including the nasal-final neighbors, the nasal-high density set had an average neighborhood density of 27.94 (range between 23 and 33) and the nasal-low density set had an average neighborhood density of 25.94 (range between 16 and 31). The two word sets had a small density difference of 2 if nasal-

final syllables are considered as neighbors, but if not, the two word sets are similar in neighborhood density.

Pair 2 was the vowel-final mismatch condition. This pair of word sets tested neighborhood density using word sets that did not have any nasal-final neighbors. The high and low density word sets contained 21 items each, and had a small neighborhood density difference of 1.81. This pair of word sets was constructed to have a small density difference that matched the density difference in the previously described nasal-final neighbor mismatch condition. The vowel-high density set had an average neighborhood density of 28 (range between 22 and 31), and the vowel-low density set had an average neighborhood density of 26.19 (range between 22 and 28). The two word sets overlapped in the number of neighbors (neighborhood density) because this pair of word sets was constructed to match the word sets in the nasal-final condition, in which the two word sets also overlapped in the number of neighbors (matching number of vowel-final neighbors and mismatching number of nasal-final neighbors).

Pair 3 was the large difference condition. This pair of word sets tested neighborhood density with a larger density difference than the two pairs mentioned above. The high and low density word sets contained 16 items each. The two word sets had a large density difference of 7.88; unlike previously described word sets, the high and low density word sets in this condition did not overlap in the number of neighbors. The large-high density set had an average neighborhood density of 24.63 (range between 20 and 31), and the low density set had an average neighborhood density of 16.75 (range between 14 and 19).

Table 1. Neighborhood densities of the final materials

	Pair 1 Nasal-final neighbor mismatch		Pair 2 Vowel-final neighbor mismatch	Pair 3 Large density difference
Neighbors included	N+V	(V-only)	V-only	V-only
Neighborhood density				
High density	27.94	(26.29)	28.00	24.63
Low density	25.94	(25.94)	26.19	16.75
Density difference	2	(0.35)	1.81	7.88

In the parentheses, ND includes vowel-final neighbors (V) only, excluding nasal-final neighbors (N).

Among the 91 monosyllables selected, 11 items occurred in two word sets, 3 occurred in three word sets, and none of the items in the nasal-high word set (these items all had nasal-final neighbors) occurred in another word set. In addition, another 10 monosyllables that were not among the 91 selected targets were randomly selected as practice items.

All procedures for selecting syllables mentioned above did not consider tonal contrasts (all tone-general syllables) since tonal information by itself had little influence on spoken word processing. Yet, a real word in Mandarin Chinese consists of initial and final components (a monosyllabic combination of phonemes) plus tone; thus, the selected syllables were not real words in the language until a legal tone was determined. In Mandarin Chinese, each tone-general syllable corresponds to many lexical items, and in this present study, no contextual information was available as

31

monosyllabic stimuli were presented in isolation. It was assumed that word frequency would bias lexical access once a phonological word form is identified. Therefore, for each selected syllable in the present study, the word with the highest word frequency was selected out of all the words sharing the same syllable structure. Specific word frequencies were searched, compared and selected from the Academia Sinica Balanced Corpus (Academia Sinica, 1997; Cheng, et al., 2005). Log-adjusted values of frequency were calculated for the selected words. The log word frequency and syllable log frequency of the 91 selected syllables were positively correlated ($r= .70$).

Each pair of word sets were matched in neighborhood frequency, word form/syllable frequency, word frequency (log frequency differences within .1 for the above mentioned factors), initial consonants (i.e., each initial consonant occurred the same number of times in both word sets, and thus the number of initial-consonant neighbors was also matched), and number of items (see Table. 2 for summary).

Table 2. Matched word sets

	Nasal-final neighbor mismatch	Vowel-final neighbor mismatch	Large density difference
Neighborhood frequency (log-adjusted)			
High density	3.98	4.00	3.95
Low density	3.96	4.01	3.89
Syllable/word form frequency (log-adjusted)			
High density	4.14	4.26	3.75
Low density	4.07	4.25	3.68
Word frequency (log-adjusted)			
High density	3.31	3.28	2.77
Low density	3.28	3.36	2.74
Stimuli duration (msec)			
High density	341.82	342.24	342.81
Low density	340.00	357.29	331.81
Initial consonants	/ l, k, kʰ, n, p, ʂ, tʂ, tʂʰ, t, tʰ, X/	/kʰ k, l, m, n, ʂ, ts, tʂ, tʂʰ, t, tʰ, X/	/f, k, kʰ l, m, n, pʰ, p, ɹ, ts, tʂʰ, t, X/
Number of initial-consonant neighbors	10.76	10.43	9
Number of items	17	21	16

However, it was difficult to match vowels or the number of vowels in a vowel unit between word sets. Among the six word sets, both vowel-high and vowel-low sets had mostly diphthongs (62% and 76% respectively); the nasal-high set had mostly single vowels (88%) and nasal-low mostly diphthongs (82%); the large-high set had mostly diphthongs (88%) while the large-small set had mostly three vowels (50%). However, no difference was found in word length in terms of stimulus duration, measured from stimulus signal onset to offset, between word sets.

The 91 test stimuli and 10 practice stimuli were recorded by an adult female native speaker of TMC using a Shure SM81 microphone and a Mackie Micro Series 1202-VLZ 12 Channel Mic/Line Mixer onto a computer at 44.1 kHz sampling rate and 16 bits precision in a sound booth. The recorded stimuli were separated into 109

single digital files and edited with Cool Edit 2000 by deleting silent sections before and after each stimulus. The volume of each stimulus recording was normalized so that all syllables reached the same peak amplitude.

Test items and practice items were stored on a Macintosh computer. Stimuli were randomly presented to the participants using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) in two separate sections; the 10 practice items were presented in the practice section and the 91 test items in the test section. Participants heard speech signals through Koss SB-40 headphones, which were connected to a Symetrix headphone amplifier interfaced to the computer. They responded by speaking into the microphone of the headset placed immediately next to the mouth. The microphone was connected to the Button Box of PsyScope, which was interfaced to the computer for response (reaction time) measurement and recording. For response accuracy check, a clip-on microphone, connected to an audiotape recorder, was placed about 30 cm from the participant's mouth. To record both the stimuli and responses in case of response recording failure of the computer, a set of speakers was connected to the headphone amplifier as an additional output of the stimuli (the stimuli were presented to the participants through headphones). The speaker volume was set at a low level to avoid interference with voice key activation during the experiment.

Design

The study was a within-subject repeated measure design. All participants were measured in all conditions. The major part of the study tested two independent variables, two levels each, a 2 x 2 design. The independent variables were neighborhood density (high and low densities) and neighbor type (nasal-final

mismatch and vowel-final mismatch). The dependent variables of the study were reaction time and response accuracy rate.

Two conditions (vowel-final mismatch and nasal-final mismatch conditions) of word sets contrasting in density were used to test whether neighborhood density affects spoken word recognition in TMC, and whether CVC and CV syllables are equivalent neighbors of a CV target. The high and low density word sets in both conditions were matched in number of items, initial consonant, word frequency, stimulus durations, and neighborhood frequency. In the vowel-final mismatch condition, all items in the high and low density sets had only vowel-final neighbors, and the two sets differed in the number of vowel-final neighbors, with a density difference of 1.81. In the nasal-final mismatch condition, the high and low density word sets were matched in the number of vowel-final neighbors, but items in the high density set all had nasal-final neighbors in addition, while those in the low density set did not; thus, the high and low density word sets differed in the number of nasal-final neighbors. If nasal-final syllables were not considered as neighbors, the two word sets were similar in neighborhood density; yet, if the nasal-final syllables were considered as neighbors, the high and low density word sets had a neighborhood density difference of 2. This allows an examination of whether items ending with a nasal are neighbors of items which do not.

Another set of high and low neighborhood density word sets with a large density difference was constructed to test the effects of neighborhood density, because the density difference was very small in the main part of the study. If the effects of neighborhood density exist in a language, a large density difference would

be more likely to reflect the effects than a small density difference would. In case of a null effect of neighborhood density in the main study, this set of comparisons would test if the effects of neighborhood density exist in the language at all. In this large difference condition, high and low density sets had a large density difference, and were matched in number of items, initial consonant, word frequency, stimulus durations and neighborhood frequency.

A total of 91 test items were presented to all participants. Test items were completely randomized by the computer for presentation for each participant to control for potential order effects.

Procedures

The study took place in a quiet room with one participant at a time. All participants were first introduced to the general purpose of the study and relevant information provided along with the consent form. Participants then completed a questionnaire about their language background (age of acquisition, self rating of language proficiency, estimated daily use of each language in proportion, languages used in family, the longest place of residency, and so forth). Participants were then seated in a chair facing a monitor with the experimenter sitting behind the participants.

The experimenter explained the general process and assisted with adjusting sitting posture and adjusting equipment. Participants were instructed to sit in a comfortable posture to reduce movements during the experiment. The headset was provided and volume was adjusted from the headphone amplifier to a comfortable level as reported by the participants while listening to practice items. The Button Box

Voice Key was tested by first having participants repeat items they heard through the headphones, using practice items; participants were reminded to respond with a comfortable vocal loudness level and keep it as steady as possible throughout the trials. The volume of the voice key was adjusted to the lowest level at which the voice key would be triggered by a verbal response and nothing else. Upon obtaining a steady state of response and activation with the lowest volume level of the voice key, the experimenter then instructed participants to stop responding, and checked for any oversensitive detection of signals by the voice key (such as the voice key triggered by breathing). Occasionally, participants were requested to respond with increased vocal loudness as a result of over-detection of irrelevant signals.

The experimenter provided task instructions after adjusting all equipment. Participants were informed that they would hear a list of single syllable TMC words (all real words of TMC ("guo2yu3") and "no Taiwanese, Hakka or English words" was emphasized), one at a time through the headset, and that they were to quickly repeat each word they heard into the microphone as quickly and accurately as possible. Reaction time was measured and recorded by the computer.

Test Trials

In the experiment, the first trial began 500 ms after the participant pressed a key to start the study, and the initiation of following trials was triggered by participants' voice responses. Stimuli were presented from the beginning of each trial. The reaction time was measured from the stimulus onset to the activation of the Button Box Voice Key. Upon activation of the Voice Key, there was a 1500 ms intertrial interval before the initiation of the next trial. Throughout each trial, a cross

was presented in the center of the screen for participants to focus on, preventing distraction; however, they were not required to look at the cross.

Ten practice items not included in the 91 selected stimuli were provided under the same setting as the experimental trials. Participants were allowed to repeat practice trials as they felt necessary, and no response was recorded during practice. At the beginning of the test trials, the participants saw the written instructions on the screen to "press any key to start". The test items were presented after they press a key on the keyboard in front of them. When test experiment was over, a statement appeared on the monitor screen informing the participants that all trials were completed. The duration of the experiment was approximately 20 minutes.

Reliability

Reliability was conducted on accuracy ratings and data coding. In accuracy rating, two raters listened to the audiotape recording of the responses independently, and transcribed what was heard into MPS-I in trial sequence for each participant; in addition, trials with incidents such as failure of the voice key to activate and coughing were marked. The transcription was then compared with a list of actual trial stimuli and coded as "1" (correct trials) and "0" (incorrect trials). Overall agreement on accuracy rating (both correct and incorrect trials) was calculated by dividing the total number of agreed correct and incorrect trials by the total number of trials. An agreement on "correct" rating was calculated by dividing the number of agreed correct trials by the total number of agreed and disagreed correct trials, and the agreement on "incorrect" rating is calculated by dividing the number of agreed incorrect trials by the total number of agreed and disagreed incorrect trials. A

reliability check on accuracy ratings was conducted on 8 randomly selected participants (30% of the data). A native speaker of TMC, blind to the purpose of the study, was invited as a second rater. Although the recording of responses was poor in quality, overall agreement on accuracy rating was high, 99.2%. The agreement on "incorrect" rating was 71.4%, and the agreement on "correct" rating was 99.2%.

Reliability of reaction time (RT) measures made by the computer was checked by comparing computer measures with hand measures. The audio recording of responses were transferred into digital files, separated by participant number, and analyzed on computer using Cool Edit 2000. Each pair of stimulus- response signals was adjusted in amplitude with normalized peak level to facilitate determination of signal onset and offset. RT was defined as the beginning of the preceding stimulus offset to the beginning of the response onset based on waveform patterns. The experimenter measured each trial and recorded the RTs in trial sequence. Data from two participants were randomly selected for hand measurement. Correlations between computer and hand measures for the two participants were poor ($r = .14$) and fair ($r = .81$) with mean difference of 105 ms and 103 ms. A second coder, experienced in analyzing digital signals, independently measured one of the two participants' data by hand and recorded RTs in trial sequence. Good correlation was obtained between the two coders ($r = .93$), with a mean difference of 25 ms.

By examining the differences between computer and hand measured RT data in the two participants that were hand measured by the experimenter, certain patterns were found. Among the trails that were 2 standard deviations below the mean measurement difference (that is, the computer measured RTs were much slower than

the hand measured RTs), 75% were s/sh-initial syllables. It is possible that the weak onset of s/sh-initial syllables was too low in intensity to trigger the voice key, and thus resulting in longer reaction time measured by the computer. Among the trials that were 2 standard deviations above the mean difference (that is, the computer measured RTs were much faster than the hand measured RTs), 80% were measured as responses below 150 ms by the computer. It is likely that during these trials, the voice key was triggered by sounds prior to verbal responses, such as noise caused by movements. Measurement differences were then examined by items. Any item with a measurement difference of 1 standard deviation beyond the mean found in both participants' data was also selected for hand measure; two items were found, "pa4" and "ma1". The computer measured RTs of these two items were not consistently faster or slower for the two participants (that is, the computer measured RTs for each item were much slower in one participant and much faster in the other participant than the hand measured RTs), and the reason for this discrepancy is unclear. The two outliers with the most discrepancy (beyond 3 and 4 standard deviations from the mean difference) were trials in which the voice key failed to activate upon a typical response by the participant, for some unclear reason. In summary, it was decided that trials would be hand measured if meeting one of the following criteria: 1) RTs less than 150 ms measured by computer, 2) trials in which the voice key failed to activate (as marked during accuracy rating), 3) trials with s-/sh-initial syllables, or 4) items of "pa4" and "ma1", which had a large discrepancy in both participants' data.

After hand correction of RTs on the selected trials based on the three criteria, correlation was measured between the partial corrected data and the fully hand

measured data for the two participants. Fair correlations were obtained (.85 and .83), yet with still large mean difference of 65ms and 94ms. No further modification to the criteria was made since correlations were fair. Trials meeting the criteria were then selected and corrected for all participants. A total of 205 trials out of the total 2366 trials (9%) required hand correction of RT.

Data Analyses

Response accuracy data were calculated by dividing the number of correct responses by total number of responses for each participant and for each item. Analysis of response accuracy was conducted by participant and by item.

Reaction time measured by the computer was set to start at the stimulus onset, rather than the offset, to ensure measurement and recording of any fast response prior to stimulus offset. Reaction time data for analyses were obtained by subtracting stimulus duration from the RTs measured by the computer. Reaction time data were then partially corrected manually as described previously. Trials with a reaction time beyond 2 standard deviations from each participant's mean were excluded from the rest of reaction time analyses. Average reaction time was obtained across items and across participants. A separate average reaction would be calculated by discarding the items and participants that showed an accuracy rate below 80%, yet no item or participant had an accuracy rate below 80%. Analysis of the reaction time was conducted by subject and also by item.

Statistical data analyses by participant were conducted using two-way repeated measures ANOVA for comparison among neighborhood density and neighbor types, and paired t-tests were used for the comparison between high and low density with

41

large density difference. Analyses by item were conducted using a two-way fixed

factor ANOVA for comparison among neighborhood density and neighbor type, and

an independent t-test was used for the comparison between high and low density with

large density difference.

Chapter 3: Results

The average reaction times by participants in this study showed a wide range, from 194 ms to 898 ms with a mean of 440 ms and a standard deviation of 152 ms, but responses were overall high in accuracy rate with a mean of 98%, ranging from 93% to 100%. The average reaction time by items showed a mean of 419 ms and a standard deviation of 62 ms, ranging from 210 ms and 561 ms; average accuracy rate by item showed a mean of 98%, ranging from 81% to 100%. Trials with reaction times beyond 2 standard deviations from the participant's mean were excluded, and the remaining trials all had reaction times between 0 and 2 seconds. In total, 178 out of 2366 (8%) trials were excluded from reaction time analyses. Table 3 lists the group means of reaction time and accuracy rate analyzed by participants.

Table 3. Group means (by participants) of each word set under the three conditions

	Nasal-final neighbor mismatch	Vowel-final neighbor mismatch	Large density difference
Mean reaction times (msec)			
High density	431.41	441.34	451.84
Low density	434.24	429.76	448.98
Mean accuracy rates (%)			
High density	97.51	96.63	97.07
Low density	97.96	98.32	97.80

The main part of the study examined the effects of neighborhood density (high vs. low) and neighbor mismatch type (nasal-final neighbor mismatch vs. vowel-final neighbor mismatch) on reaction time and response accuracy. There were four groups of words being tested, the vowel-high, the vowel-low, the nasal-high and the nasal-low groups. In addition, the nasal-high and the nasal-low items contained similar numbers of vowel-final neighbors, but the two sets differed in that the nasal-high group allowed nasal-final neighbors yet the nasal-low group did not. It will be recalled that in terms of density, words with high neighborhood density were predicted to show slower and/or less accurate responses than words with low neighborhood density. In terms of neighbor type, it was predicted that there would be no difference in reaction time or response accuracy between the nasal-final neighbor mismatch and the vowel-final neighbor mismatch conditions; that is, both types of neighbors would contribute equally to the effects of neighborhood density.

In the main part of the study, results of reaction time analyses by participant were consistent with the prediction that words with high neighborhood density would be responded to more slowly than words with low neighborhood density, but results were inconsistent with the prediction that responses in the nasal-final and the vowel-final neighbor mismatch conditions would show similar patterns.

Two-way repeated measure ANOVA analyses of reaction time data by participant showed a statistically significant interaction between neighborhood density and neighbor type, $F(1,25)=8.29$, $p<.01$ (Figure 1, left), and no main effect of neighborhood density, $F(1,25)=1.39$, $p>.2$, or neighbor type, $F(1,25)=1.17$, $p>.25$, was found. The significant interaction suggests that there were differential effects of

neighborhood density across neighbor types. Analyses of response accuracy by participant showed no difference among the four groups (Figure 1, right).
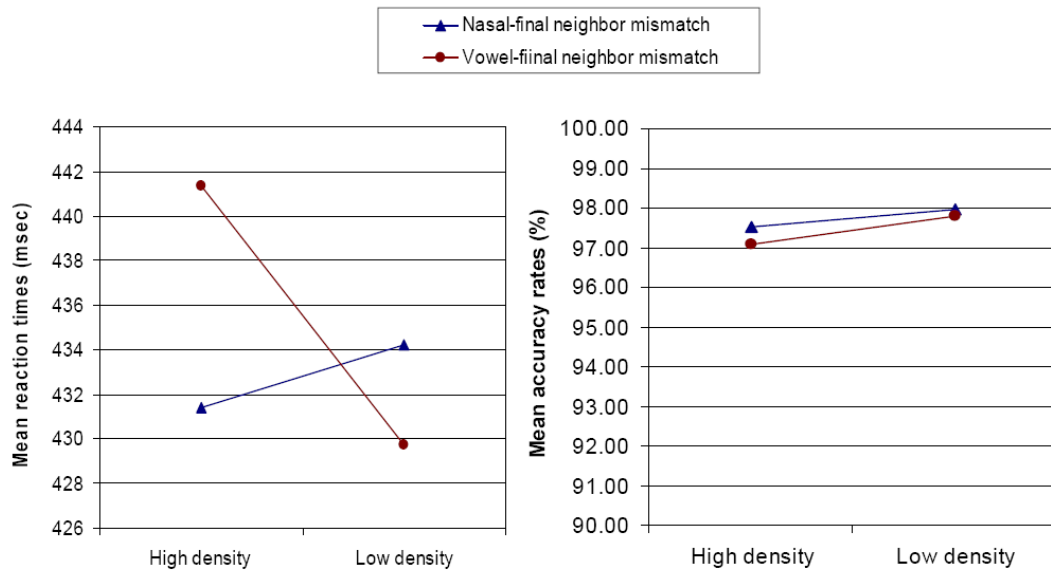

Figure 1. Mean reaction times and mean accuracy rates by neighborhood density level and neighbor type (analyses by participant)

Paired *t*-tests were used for follow-up comparisons. Two comparisons were conducted between the two density levels within each neighbor type (i.e. the nasal-high vs. the nasal-low in the nasal-final neighbor mismatch condition and the vowel-high vs. the vowel-low in the vowel-final neighbor mismatch condition). Analyses showed a simple effect of neighborhood density in the vowel-final mismatch condition; that is, participant's responses were significantly slower to the vowel-high items than the vowel-low items, $t(25)=3$, $p<.01$ (mean difference: 11ms). No simple effect of neighborhood density was found in the nasal-final mismatch condition, that is, there was no difference between the nasal-high and the nasal-low groups, $t(25)=-.56$, $p>.5$. The predictions and results of this main part of the study are summarized in Table 4.

Table 4. Summary of predictions and results in the nasal-final and vowel-final mismatch conditions

	Pair 1 Nasal-final neighbor mismatch		Pair 2 Vowel-final neighbor mismatch
Neighborhood density	N+V	V-only	V-only
High density	29.94 (slow)	26.29 (no difference)	28.00 (slow)
Low density	25.94 (fast)	25.94 (no difference)	26.19 (fast)
Density difference	2	0.35	1.81
Result		✓	✓

"N+V" indicates neighborhood density including both nasal- and vowel-final neighbors; "V-only" indicates neighborhood density including vowel-final neighbors only. In the parentheses are the predictions (the predicted patterns of response speed in the two conditions), and the highlighted are the results.

Two-way fixed factor ANOVA analyses of reaction time data by item showed no main effect of neighborhood density or neighbor-mismatch type and no statistically significant interaction between the two factors. Though the reaction time analyses by item did not reach significance, the data appeared to show similar patterns as the data by participant. Reaction time data by item showed a pattern of slower responses in the vowel-high group than in the vowel-low group (average reaction times: 430 ms and 407 ms respectively), the same between the nasal-high and the nasal-low groups (average: 403 ms in both groups), slower responses in the vowel-high group than in the nasal-high group (430 ms and 403 ms respectively), and similar reaction time in the nasal-low and the vowel-low groups (403 ms and 407 ms). Response accuracy showed no difference among the four groups, either by participant or by item (all group means were within 1% difference).

The separate comparison set up to examine the effects of neighborhood density with a large density difference (average difference: 7.87) predicted that words with high neighborhood density would be responded to more slowly and/or less accurately than words with low neighborhood density in this large density difference condition. However, contrary to this prediction, no group difference was found in any of the analyses between this pair of word sets, as shown in Figure 4. Analyses included the following: reaction time by participant using a paired sample t-test, $t(25)=.69$, $p>.5$ (high and low density group means: 452 ms and 449 ms), reaction time by item using an independent t-test, $t(30)=.44$, $p>.5$ (high and low density group means: 439 and 431 ms), response accuracy rate by participant, $t(25)=-1.77$, $p>.05$, and response accuracy by item [$t(30)=-1.14$, $p>.5$]; all group means were within 2% difference.

Chapter 4: Discussion

In general, the results of this study provided support for the hypothesis that words with high phonological neighborhood density would inhibit spoken word recognition in TMC. In this study, words with more neighbors were responded to more slowly than words with fewer neighbors. However, the results did not support the hypothesis that nasal-final CVC syllables are neighbors of vowel-final CV syllables; the study actually suggests that nasal-final syllables are not neighbors of a vowel-final target syllable.

Neighborhood Density

The inhibitory effect of neighborhood density found in the main part of the study is consistent with what has been found in English spoken word recognition. The pair of high and low density word sets with a small density difference (vowel-high and vowel-low word sets) showed a significant difference in reaction time by participants; yet, it should be noted that the difference did not reach the significance level by item. In the large density difference condition, however, results did not show a difference in the responses between the high and the low density word sets. If the effects of neighborhood density exist in a language, a large density difference between two word sets is more likely to reflect the difference in lexical access than a small density difference. Yet, contrary to this prediction, the study showed an inhibitory effect of neighborhood density only in the small density difference condition but not in the large density difference condition. The contradicting results

led to a closer examination of the two sets of words (high and low densities) in the

two conditions (small and large density differences).

In both conditions, the high and low density word sets were matched in initial

consonant, syllable duration, syllable frequency, word frequency and neighborhood

frequency, with vowels left uncontrolled. As mentioned earlier, the vowels in

Mandarin Chinese show complimentary distribution, and certain phonotactic rules

have been suggested in the language (Wan & Jaeger, 2003). This uncontrolled factor

might lead to an issue of differences in phonotactic probability between each pair of

contrasting word sets, if phonotactic probability plays a role in this study. The

prelexical effects of phonotactic probability have been found on spoken (non)word

processing in English (Pÿlkkanen, Stringfellow, & Marantz, 2002; Vitevitch & Luce,

1998, 1999, 2005), and a similar effect was suggested in Japanese in speeded auditory

naming tasks (Yoneyama, 2002), both suggesting that this factor have an influence on

word recognition at the sublexical (prelexical) level. A measure relevant to

phonotactic rules, biphone probability, indicates the relative frequencies of a position-

specific phoneme sequence in a language. The calculation of biphone probability in

English is to sum up the log-adjusted frequency values of all words that contain the

position-specific phoneme sequence, and divide it by the sum of log-adjusted

frequency values of all words that contain phonemes in the two specified adjacent

positions (Vitevitch & Luce, 2004). Yet, with the simple CV(C) syllable structure in

Mandarin, the biphone probability of a CV sequence in Mandarin is basically the

probability of the CV syllable itself (no other syllables would share the same CV

sequence with the target syllable, except for the syllables that allow an additional

final nasal consonant, CVC). In this study, the CV syllables in the high and low density word sets in both the small and large density difference conditions did not allow any nasal final consonant, and both pairs of word sets were matched in syllable frequency. This means that in the present study, biphone probability would not differ between word sets in either small or large density difference condition (see Table 5).

Another way of looking at phonotactic probability is to examine the phonotactic rules. The 22 initials (consonant or zero consonant) and 37 finals (a vowel, vowel combination, or vowel/s with a nasal consonant) in Mandarin Chinese phonology would create 814 possible combinations of an initial and a final to form a syllable, but there are only approximately 400 legal syllables in the language (Li & Thompson, 2003). The complimentary distribution of phonemes in Mandarin Chinese shows that certain types of initials are more likely to take certain types of finals. The phonotactic rules generally reflect the relations between the place of articulation of the initial consonant and the quality of its following vowel (the leading vowel of a final). Traditionally, there are four types of finals defined by the quality of the leading vowel: *Kaikou* includes finals led by either mid- or low-vowels (e.g. /e, o, a, ei, au, ou/), *Qichi* includes finals led by the high-front vowel /i/ (e.g. /iau, iou, ia/), *Hekou* includes finals led by the high-back vowel /u/ (e.g. /u, ua, uo, uei/), and *Cuokou* includes finals led by the high-front rounded vowel /y/ (e.g. /yn/). Using a corpus of approximately one million words, Zhang (cited by Zhang, 1996) reported the relative frequency of the four types of finals combined with each of the nine types of initials. For example, most of the labial-dental initial consonants occurred with *Kaikou* finals (85%) (none combined with *Qichi* or *Cuokou* finals), and most syllables with the zero

consonant consist of *Qichi* finals (55%). There is a difference between the two pairs

of word sets examined in the current study (see Table 5). Over 50% of the items in

the word sets with a small density difference contained initial consonants combined

with their most probable types of finals. However, in the large difference condition,

over 75% of the items in the large-high word set contained initials combined with

their most probable types of finals, but the large-low word set had only 31% of such

items. This lower probability of initial-final combinations in the large-low word set

might have a negative impact on the speed of processing. Thus, it is likely that the

effects of neighborhood density and the effects of phonotactic probability could

cancel each other, resulting in a null effect on the speed of spoken word recognition.

It has been suggested that Mandarin-listeners utilize such phonotactic rules (the

probability of types of initial-final combinations) and syllable structure to facilitate

speech perception, especially in recognizing the place of articulation of a consonant

(Zhang, 1996). However, the relation between phonological neighborhood density

and this joint probability of initial-final combination and their effects along the time

course of Mandarin Chinese spoken word recognition require further research.

Another source of discrepancy between the small and large density difference

conditions is the number of vowels in a vowel unit. Though syllable durations were

matched between word sets, the amount of phonological information conveyed in a

syllable appeared to be different among word sets. The two word sets with a small

density difference contained mostly a two-vowel combination as a vowel unit (62%

of the items in the vowel-high and 76% of the items in the vowel-low), but in the

large density difference condition, most items in the large-high word set (88%)

contained two-vowel combinations while most of the items in the large-low word set (50%) contained three-vowel combinations as a vowel unit. Three-vowel combination is relatively infrequent in the present study (0% in the nasal-high, 6% in the nasal-low, 5% in the vowel-high, 5% in the vowel-low, 6% in the large-high and 50% in the large-low word sets). However, this discrepancy in the number of vowels, with the low density word set having more three-vowel combinations than the high density word set, would confound the effect of neighborhood density because with controlled duration, words containing more phonological information have been found to be associated with stronger lexical activation than words containing less phonological information, possibly as a result of fewer competing neighbors (Pitt & Samuel, 2006). It would still be predicted that the large-low word set (with many three-vowel combinations) would be responded to more quickly than the high density set, but this was not found in the study. Thus, the difference between the two conditions in terms of varying amount of phonemic information could not explain the contradicting findings between the two conditions.

To summarize, the vowel-high and vowel-low word sets were best matched in terms of vowel unit for examining the effect of neighborhood density. The two word sets, though different in vowel units, were similar in phonotactic probability (biphone probability), both had mainly high-probability initial-final combinations given their matching initial consonants, and both contained mainly diphthongs. An inhibitory effect was clearly shown from the comparison between this pair of word sets. Although reaction time analyses by item did not show any significant difference, those results were probably not as valid because items in each group were not

randomly selected, and several items overlapped between conditions. Different from the well matched vowel-high and vowel-low word sets with a small density difference, the word sets in the large density difference condition were not well matched in vowels. The null effect of neighborhood density in the large density difference condition suggests that the predicted inhibitory effect of neighborhood density might be attenuated by other factors with opposite effects on spoken word recognition, such as phonotactic probability. Based on the hypothesis that neighborhood density mediates lexical access with an inhibitory effect, and phonotactic probability mediates the process with a facilitative effect when lexical access is weak, the null effect found in the large density difference condition might be related to both factors. Yet, the contradicting results from the small and large difference conditions require further research to determine the factors involved in lexical access besides neighborhood density.

Table 5. Phonotactic probabilities in the small and large difference conditions

	Small density difference (Vowel-final neighbor mismatch)	Large density difference
Biphone probability		
High density	0.002784	0.002452
Low density	0.002776	0.002408
Vowel probability		
High density	0.048335	0.041411
Low density	0.048537	0.037537
Iinitial-final (I-F) type probability		
High density	0.829041	0.821284
Low density	0.822664	0.783327
% of most vowel combinations		
High density	2-vowel (62%)	2-vowel (88%)
Low density	2-vowel (72%)	3-vowel (50%)
% of most probable I-F combination		
High density	57%	75%
Low density	67%	31%

The other main question of the study is whether nasal-final CVC syllables and vowel-final CV syllables are equivalent neighbors of a target CV syllable, all differing from the target in one phoneme including substitution, addition and deletion. The study showed that nasal-final syllables are not neighbors of a CV vowel-final syllable, even though they differ from the target in one phoneme (the additional final nasal consonant). One feature of nasal-final syllables is the effect of anticipatory coarticulation, that is, when a nasal consonant follows a vowel, the nasal feature would spread forward to the preceding vowel (vowel nasalization). A facilitative effect of coarticulation (both carryover and anticipatory coarticulation) on word recognition has been found in a lexical decision task with controlled frequency-weighted neighborhood density (Scarborough, 2004). In Mandarin Chinese, nasal consonant is the only legal final consonant or coda; thus, a vowel without nasalization would indicate that the word does not have a nasal-final component. It is likely that the system treats nasal-final syllables differently from vowel-final syllables; that is, nasal-final syllables are not considered as neighbors to a vowel-final target syllable. In the present study, the finding that adding a nasal-final consonant to a CV target does not make the syllable a neighbor of the CV target could suggest that perhaps some phonetic features should be considered in defining a neighbor.

The nasal-final neighbors and vowel-final neighbors defined in this study differed not only in the phonetic features (vowel vs. nasal) but also in the way they were defined as neighbors (one-phoneme difference of substitution vs. addition). In the study, the vowel-final neighbors differed from the target CV syllable mostly by

substituting one phoneme, but all the nasal-final neighbors differed from the target syllable by adding one phoneme. The finding that adding a phoneme to the target syllable did not increase neighborhood density could suggest that the one-phoneme difference, including addition, deletion and substitution, might not be sufficient in defining phonological neighbors. However, it would be challenging to test this hypothesis in Mandarin Chinese because an additional phoneme of a CV syllable (the most common syllable structure in the language) is always a nasal final consonant. It would be difficult to separate the two factors (phonetic features of nasalization vs. types of one-phoneme difference).

As mentioned earlier, a difference in biphone probability is expected between the nasal-high and the nasal-low word sets because all CV syllables in the nasal-high word set had nasal-final CVC "neighbors" while items in the nasal-low word set did not. The nasal-final CVC "neighbors" shared the same CV sequence as the target syllable, and this increased the probability of the CV sequence and the overall biphone probability of the nasal-high word set. However, according to the reported probability of initial-final combinations (4 types of finals combined with 9 types of initials) (cited by Zhang, 1996), both the nasal-high and the nasal-low sets used for the present study contained mainly the highest probable type of finals given their matched initial consonants. Therefore, it is unclear at this point what role phonotactic probability plays in the process and its relation to neighborhood density in Mandarin Chinese spoken word recognition. The phonotactic probabilities and vowel differences in the nasal- and vowel-final mismatch conditions are summarized in Table 6 below.

Table 6. Phonotactic probabilities in the nasal- and vowel-final mismatch conditions

	Nasal-final neighbor mismatch	Vowel-final neighbor mismatch
Biphone probability		
High density	0.007049	0.002784
Low density	0.002664	0.002776
Vowel probability		
High density	0.118029	0.048335
Low density	0.046033	0.048537
Iinital-final (I-F) type probability		
High density	0.825578	0.829041
Low density	0.808972	0.822664
% of most vowel combinations		
High density	1-vowel (88%)	2-vowel (62%)
Low density	2-vowel (82%)	2-vowel (72%)
% of most probable I-F combinations		
High density	77%	57%
Low density	59%	67%

In brief, the study showed that nasal-final CVC syllables are not equivalent to vowel-final CV syllables when they differ from a CV target by one phoneme. Several possible explanations were provided, including redundant phonetic information provided through coarticulation (vowel nasalization), effects from different types of one-phoneme difference (substitution vs. addition) and phonotactic probability. Determining which factors are involved in the difference between vowel-final neighbors and nasal-final neighbors requires further research.

Limitation of the Study

The study is limited by the phonotactic constraints of the target language, in which vowels were difficult to match between word sets for the purpose of the present study, which was to examine neighborhood density. The unbalanced vowels, as mentioned earlier, were not expected to influence the results in terms of the

number of vowels or the overall duration; rather, the difference in vowels would result in potential difference in phonotactic probability. It should be noted that this limitation rests on the assumption that phonotactic probability might play a role in spoken word recognition in this study. It would be ideal to control for relevant factors, including word frequency, syllable frequency, neighborhood frequency, initial consonant, vowel, and stimulus duration. However, with the phonotactic constraints of the language, it would be difficult to construct word sets with different density levels, and match both initial consonant and vowel between word sets.

Another factor not controlled in the study was the morpheme density. In Mandarin Chinese, about 85% of monosyllabic words differentiated by tonal contrasts are ambiguous in meaning, and 55% have five or more homophones (Yip, 2000). Spoken word recognition presumably would have to identify the target item from a large number of homophones. The number of morphemes (morpheme density based on simplified Chinese characters) sharing the same syllable (phonological form) has been found to mediate the recognition process similarly as the phonological neighborhood density does; that is, homophonic morphemes compete and inhibit the process of lexical access (Zeng & Mattys, 2004). However, it would be difficult to determine which lexical item was actually identified, and it is unclear whether the average frequency of the morpheme neighborhood has an influence on the process. The present study assumed that the processing system would bias towards the most frequent morpheme in the neighborhood because there was no contextual information in the naming task for the selection of a specific lexical item, and thus ignored the properties of the morpheme neighborhood altogether. Yet, if morpheme

neighborhood does have an influence on spoken word processing of monosyllabic words in isolation (e.g. relative frequency of the target in a neighborhood), it would weaken the findings in the present study. Morpheme density was not calculated in the current study as this study focused on the syllable frequency. Major lexicons in traditional Chinese organize morphemes (characters) based on both phonology and orthography. Though the number of homophonic morphemes can be easily calculated from any traditional Chinese lexicon, the morpheme frequency is not listed. Thus, it is difficult to determine the number of homophonic morphemes that actually occur in a database, and conducting a search for the frequencies of approximately 10,000 to 50,000 morphemes is beyond the scope of the present study

Future Work

This study provides support of the effects of neighborhood density in Mandarin Chinese spoken word recognition, which is consistent with what has been found in English. At the same time, it found that there might be potential effects of phonotactic probability in the spoken word recognition process. It would be important to examine the effects of the two factors (neighborhood density and phonotactic probability) in Mandarin Chinese lexical access to further specify the primary level of processing or the processing strategies used by typical Mandarin-speaking adults in processing monosyllabic spoken word presented in isolation.

Another question of interest would be to follow up on examining the three types of differences in the common definition of neighbors (one-phoneme difference of addition, deletion and substitution). The present study showed that adding a nasal final consonant to a target syllable would not create an equivalent neighbor to

substituting a phoneme from a target syllable. Future research could further examine the potential factors for this finding across languages to determine whether the difference is related to the nature of the definition, the phonetic features of nasal final consonants, or whether it is language-specific.

Since the present study showed an inhibitory effect of phonological neighborhood density, future research could further explore the effect of homophonic morpheme density (neighborhoods composed of tone-specific and tone-general homophonic morphemes) with controlled phonological neighborhood density. The assumption for the effects of phonological neighborhood is that a lexical item would be chosen from a group of similar sounding words (lexical access). In Mandarin Chinese, identifying a phonological word (syllable) is nowhere close to selecting a lexical item because of the large number of homophones (with or without tonal contrasts). Though it was assumed that morpheme frequency would bias the system in the present study, which targeted monosyllabic word processing isolated from context, morpheme density is still a potential factor in processing, especially the relative frequency of the target morpheme in a morpheme neighborhood.

Conclusion

The effects of phonological neighborhood on spoken word recognition have been found in English, French, Spanish and Japanese. The present study targeted Mandarin Chinese to examine the effects of neighborhood density and also the definition of neighbor. The study tested words with different neighborhood density levels (high vs. low) and words with different types of neighbors according to the one-phoneme definition (nasal-final neighbors defined by an additional nasal final

consonant vs. vowel-final neighbors defined otherwise). The present study found an inhibitory effect of phonological neighborhood density on spoken word recognition in Mandarin Chinese, which is consistent with what has been found in general. The study also found that for a target syllable, neighbors with an additional nasal final consonant were not equivalent to neighbors defined otherwise (primarily by substituting a phoneme in the study). The results, however, suggest a potential effect of phonotactic probability when this factor is unbalanced. Phonotactic probability and neighborhood density have been associated with different levels of processing. It would require further research to determine the effects of phonotactic probability in spoken word recognition and its relation with neighborhood density in Mandarin Chinese. This would facilitate our understanding of the primary level of processing and the processing strategies used by its speakers.

References

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

Amano, S., & Kondo, T. (2000). Neighborhood and cohort in lexical processing of Japanese spoken words. In A. Cutler, J. McQueen, and R. Zondervan (Eds.), Proceedings from SWAP-2000: *ISCA Tutorial and Research Workshop on Spoken Word Access Processes* (pp. 91-94). Nijmegen: MPI for Psycholinguistics.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 802-814.

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language, 52,* 339-362.

Bashford, J. A., Warren, R. M., & Lenz, P. W. (2006). Polling the effective neighborhoods of spoken words with the verbal transformation effect. *Journal of the Acoustical Society of America, 119,* EL55-EL59.

Bi, H., Hu, W., & Weng, X. (2006). Orthographic neighborhood effects in the pronunciation of Chinese words. *Acta Psychologica Sinica, 38,* 791-797.

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46,* 751-781.

Chen, H. C. & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin and Review, 8,* 511-518.

Cheng, C. C., et al. (2005). Digital Resources Center for Global Chinese Teaching and Learnging. http://elearning.ling.sinica.edu.tw/index.html. Taipei: Academia Sinica, Institute of Linguistics.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25,* 257-271.

Cutler, A. & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics, 59,* 165-179.

De Cara, B., & Goswami, U. (2002). Similarity relations among spoken words: The special status of rimes in English. *Behavioral Research Methods, Instruments, and Computers, 34,* 416-423.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing system.* Honolulu: Universtiy of Hawaii press.

Dupoux, E., & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language, 29,* 316-335.

Ellis, A. W. & Morrison, C. M. (1998). Real age of acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 515-523.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effect of segment probability and length of the processing of nonwords. *Journal of Memory and Language, 42,* 481-496.

Gerhand, S. & Barry, C. (1999). Age of acquisition and frequency effects in speeded word naming. *Cognition, 73,* B27-B36.

Hahn, U. & Bailey, T. M. (2005). What makes words sound similar? *Cognition, 97,* 227-267.

Hsiau, A. C. (1997). Language ideology and ethnic politics in Taiwan. *Journal of Multilingual and Multicultrual Development, 18,* 302-315.

Huang, H. W., Lee, C. Y., Tsai, J. L., Lee, C. L., Hung, L., & Tzeng, J. L. (2006). Orthographic neighborhood effects in reading Chinese two-character words. *Neuroreport, 17,* 1061-1065.

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.

Li, P., Liu, Y., & Shu, H. (2006, November). Word naming and psycholinguistic norms: Data from Chinese. Poster presented at Society for Computers in Psychology, Houston.

Li, C. N. & Thompson, S. A. (2003). *Mandarin Chinese: A functional reference grammar.* Taipei: Crane Publishing Co.

Lipinski, J. & Gupta, P. (2005). Does neighborhood density influence repetition latency for nonwords? Separating the effects of density and duration. *Journal of Memory and Language, 52,* 171-192.

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

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29-63.

McClelland, J. & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

Morrison, C. M. & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 116-133.

Mulatti, C., Reynolds, M. G., & Besner, D. (2006). Neighborhood effects in reading aloud: New findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 799-810.

Newman, R. S., Sawusch, J. R. & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 873-889.

Newman, R. S., Sawusch, J. R. & Luce, P. A. (2005). Do postonset segments define a lexical neighborhood? *Memory and Cognition, 33,* 941-960.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences,* 23, 299-370.

Oldfield, R. C. & Wingfield, A. (1965). Response latencies in naming objects. *The Quarterly Journal of Experimental Psychology, 17,* 273-281.

Pÿlkkanen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language, 81,* 666-678.

Peereman, R. & Content, A. (1995). Neighborhood size effect in naming: Lexical activation or sublexical correspondences? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 409-421.

Pitt, M. A. & Samuel, A. G. (2006). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human, Perception and Performance, 32,* 1120-1135.

Scarborough, R. A. (2004). Degree of coarticulation and lexical confusability. In P. Nowak, C. Yoquelet, and D. Mortensen (Eds.), *Annual Meeting of the Berkeley Linguistics Society, Vol. 29. Phonetic Sources of Phonological Patterns* (pp. 367-378). Berkeley: Berkeley Linguistics Society.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1140-1154.

Tsai, J. L., Lee, C. Y., Lin, Y. C., Tzeng, J. L., & Hung, L. (2006). Neighborhood size effects of Chinese words in lexical decision and reading. *Language and Linguistics, 7,* 659-675.

van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory and Cognition, 24,* 573-583.

Vitevitch, M. S. (2007). The spread of the phonological neighborhood influences spoken word recognition. *Memory and Cognition, 35,* 166-175.

Vitevitch, M. S. & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science, 9,* 325-329.

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

Vitevitch, M. S. & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers, 36,* 481-487.

Vitevitch, M. S. & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken word nonword repetition. *Journal of Memory and Language, 52,* 193-204.

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

Wan, I. P. & Jaeger, J. (2003). The phonological representation of Taiwan Mandarin vowels: A psycholinguistic study. *Journal of East Asian Linguistics, 12,* 205-257.

Wu, J. T. & Chen, H. C. (2000). Evaluating semantic priming and homophonic priming in recognition and naming of Chinese characters. *Chinese Journal of Psychology, 42,* 65-86.

Wu, J. T. & Chou, T. L. (2000). The comparison of relative effects of semantic, homophonic, and graphic priming on Chinese character recognition and naming. *Acta Psychologica Sinica, 32,* 34-41.

Ye, Y. & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14,* 609-630.

Yip, M. C. W. (2000). Spoken word recognition of Chinese homophones: The role of context and tone neighbors. *Psychologia, 43,* 135-143.

Yoneyama, K. (2002). Phonological neighborhoods and phonetic similarity in Japanese word recognition. *Dissertation Abstracts International, A: The Humanities and Social Sciences, 63,* 170A-171A.

Zeng, B. & Mattys, S. (2004, May). Neighborhood effect in auditory Chinese word recognition. Paper presented at the 16th North American Conference on Chinese Linguistics, Iowa.

Zhang, J. L. (1996). On the syllable structures of Chinese relating to speech recognition. In Proceedings from ICSLP-1996: *The 4th International Conference on Spoken Language Processing, Vol. 4* (pp. 2450-2453). New York: Institute of Electrical and Electronics Engineers.

Zhang, Q. F. & Yang, Y. F. (2004). The time course of semantic, orthographic and phonological activation in Chinese word production. *Acta Psychologica Sinica, 36,* 1-8.

Zhang, Q. F. & Yang, Y. F. (2005). The phonological planning unit in Chinese monosyllabic word production. *Psychological Science, 28,* 374-378.

Zhou, X. L. & Zhuang, J. (2000). Lexical tone in the speech production of Chinese words. In Proceedings from ICSLP-2000: *The 6th International Conference on Spoken Language Processing, Vol. 2* (pp. 51-54). Beijing, China: Chinese Military Friendship Publishers.

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,* 779-793.