Constructing the Mandarin Phonological Network: Novel Syllable Inventory Used to Identify Schematic Segmentation

Karl David Neergaard, Chu-Ren Huang

Hindawi
Complexity
Volume 2019, Article ID 6979830, 21 pages
https://doi.org/10.1155/2019/6979830

Research Article

Constructing the Mandarin Phonological Network: Novel Syllable Inventory Used to Identify Schematic Segmentation

Karl D. Neergaard [1,2] **and Chu-Ren Huang** [2]

[1] *Laboratoire Parole et Langage, Aix/Marseille University, Aix-en-Provence 13604, France*
[2] *Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong*

Correspondence should be addressed to Karl D. Neergaard; karlneergaard@gmail.com

Received 29 June 2018; Revised 15 October 2018; Accepted 5 November 2018; Published 23 April 2019

Guest Editor: Cynthia Siew

The purpose of this study was to construct, measure, and identify a schematic representation of phonological processing in the tonal language Mandarin Chinese through the combination of network science and psycholinguistic tasks. Two phonological association tasks were performed with native Mandarin speakers to identify an optimal phonological annotation system. The first task served to compare two existing syllable inventories and to construct a novel system where either performed poorly. The second task validated the novel syllable inventory. In both tasks, participants were found to manipulate lexical items at each possible syllable location, but preferring to maintain whole syllables while manipulating lexical tone in their search through the mental lexicon. The optimal syllable inventory was then used as the basis of a Mandarin phonological network. Phonological edit distance was used to construct sixteen versions of the same network, which we titled phonological segmentation neighborhoods (PSNs). The sixteen PSNs were representative of every proposal to date of syllable segmentation. Syllable segmentation and whether or not lexical tone was treated as a unit both affected the PSNs' topologies. Finally, reaction times from the second task were analyzed through a model selection procedure with the goal of identifying which of the sixteen PSNs best accounted for the mental target during the task. The identification of the tonal complex-vowel segmented PSN (C_V_C_T) was indicative of the stimuli characteristics and the choices participants made while searching through the mental lexicon. The analysis revealed that participants were inhibited by greater clustering coefficient (interconnectedness of words according to phonological similarity) and facilitated by lexical frequency. This study illustrates how network science methods add to those of psycholinguistics to give insight into language processing that was not previously attainable.

1. Introduction

The meeting of network science and the study of phonological processing has allowed for the examination of the mental lexicon according to mathematical principles that have both theoretical and methodological import. Researchers have used what are known as phonological networks, in which words (nodes) are connected to other words (edges) based on phonological similarity. Phonological networks have been used in basic research to examine speech processing during word recognition [1, 2], word production [3], word learning [4], and working memory [5]. Most recently, they have been applied to the study of speech pathologies in the examination of both aphasic speech [6] and stuttering [7, 8].

Common among the phonological networks that have been examined thus far is that phonological similarity is measured at the level of the phoneme. This is due to the relational parameter between nodes being defined by phonological edit distance, wherein two words are "neighbors" if they differ through the addition, deletion, or substitution of a single phoneme [9]. A given node's degree is thus the total number of words that are immediate neighbors, most commonly referred to as phonological neighborhood density [10]. One possible problem with using the phoneme as the basic unit between words is its generalizability to non-European languages. While English has gained attention in modeling network topologies of phonology [9–13], little has been done outside English. The two studies to date [14, 15]

implemented the single-phoneme edit metric. Yet, is a one size fits all approach appropriate cross-linguistically? This is an especially pertinent question in light of Mandarin Chinese. Mandarin not only has unique lexical features that set it apart from the languages studied to date but has long enjoyed a debate as to both its phonological annotation and syllable segmentation, i.e., two aspects that would likely affect network dynamics.

Mandarin has become a recent focus in the psycholinguistic literature due to a set of linguistic features that test the limits of models of speech processing previously developed for European languages. Perhaps the most unique is the status of the syllable, which is tonal, of equivalent size to the primary orthographic unit, and highly homophonic. Unlike English or Dutch, which both have over 10,000+ syllables, the Mandarin syllable inventory is small, featuring ~1,300 syllables plus tone and ~400 without tone. Excluding a select number of high-frequency lexical items that do not regularly carry tone, each syllable carries one of four tones: tone 1 (high level pitch, 55), tone 2 (low rising pitch, 35), tone 3 (low dipping pitch, 214), or tone 4 (high falling pitch, 51). Aside from the dialectal phenomena known as erhua [16] in which the character 儿 (*er2*) is added to another character yet pronounced as a single syllable (玩儿, *wan2 er2* = *war2*), each syllable in the inventory matches one or more Chinese characters. Mandarin has been shown to be largely disyllabic in nature; in fact it has been calculated that two-thirds of all Mandarin words consist of two characters [17, 18]. Yet, characters that do not exist as monosyllabic words, meaning they only exist in multisyllabic words, are still lexical items that contribute to the count of homophone neighbors. In context, the same roughly 1,300 tonal syllables service all lexical combinations from monosyllabic to multisyllabic words. This leads to a homophone density (i.e., the number of homophone neighbors a given word has) of up to 48 when tone is considered [19]. To put this in context, 11.6% of Mandarin words have homonyms, compared to 3.15% in English [20]. High homophony has been shown to lead to lexical competition in spoken word recognition, as seen by slower reaction times, and lower accuracy [19, 21]. This is uniquely important to Mandarin given the relation of orthography to the syllable.

Researchers have used two methods to describe how segmental units comprise a syllable in Mandarin. One method recognizes a maximum of 4 segments, CGVX, such that C represents initial consonants, G medial glides, the V monophthongs, and the final X the second part of a diphthong, or a final consonant. Early accounts proposed segmentation schemas based on the constituents of the rime or whether the medial glide constituted a unique phonological role within the syllable: C_GVX [22]; C_G_VX [23, 24]; CG_VX [25]; and CG_V_X [26, 27]. Note that here an underscore denotes a separation between phonological units. The second method of describing the Mandarin syllable collapses all glide and vowel information, leaving a maximum of 3 units: C_V_C [28]. The methods used to arrive at the various schemas, and whose evidence has informed the creation of syllable inventories, come through production tasks that have participants read sentences so as to measure syllable durations [29–31], or produce phonological neighbors in rhyming games [25, 32, 33]. More recent approaches depart from these methods to instead investigate segmentation as a product of either perception or production.

O'Séaghdha and colleagues [34, 35] hypothesized that the first phonological units available for selection below the level of the word or morpheme, titled proximate units, correspond to nontonal syllables in Mandarin. Their thesis was that unit sizes would vary across languages, granting phonemes and clusters of phonemes in Indo-European languages such as English, while larger units such as morae in Japanese. Speech error analyses have supported this trend, such that in English the dominant unit size is segmental [36, 37] and in Mandarin syllabic [38–40]. For speakers from alphabetic languages, like English and Dutch, sensitivity to syllable onsets between two lexical items has been documented in numerous studies and across multiple paradigms [35, 41–46]. These studies show that prior preparation to segmental units shared between lexical primes and target lexical items speeds production of the target word, implying that temporary storage occurs for segmental information. A corresponding series of priming studies have shown syllabic priming results yet no significant onset priming with Chinese orthography in the implicit priming [35, 47] and masked priming tasks [48–50]. To counter the syllable bias of Chinese characters, similar studies were conducted with picture [49, 51] and auditory stimuli [51]. Supporting evidence for the proximate unit has also been advanced in priming studies with both speakers of Cantonese [52–54] and Japanese [55–58]. While no syllable schema was explicitly proposed by the authors related to the proximate unit proposal, their statement that the primary unit to be selected is nontonal suggests that either a nontonal unsegmented schema is the target (CGVX), or its tonal counterpart (CGVX_T) seeing as the syllable is combined with tone prior to production.

To stand in contrast to speech production studies is a growing body of evidence to support the claim that Mandarin speakers, during speech recognition, process segmental information incrementally and in parallel with tonal information. Differential processing between lexical units was analyzed within a picture-word matching paradigm with both ERP [59–61] and eye-tracking [62]. Malins and colleagues found that whole-syllable mismatches did not produce effects greater than those found with individual components, mirroring results previously found in English [63]. This has motivated their claim that processing was segmental, an assertion not entirely supported in [61], which found greater evidence for syllable-level processing. One important difference in the latter study however is the fact that they also used Chinese characters during the presentation of their picture stimuli. Their results have likely an effect due to the activation of syllable-sized orthography. One limit to the claims put forward by Malins and colleagues, which implies words reside within a tonal fully segmented schema (C_G_V_X_T), is that they did not feature mismatch pairs according to glides, or the X unit [59, 60, 62]. Thus, to date these studies provide evidence for the tonal complex-onset/rime schema (CG_VX_T).

The current study began from the ground up through the creation of a novel phonological annotation system, also concurrently referred to as a syllable inventory. The creation of a novel inventory was necessary because (1) differences between existing inventories [23, 64–66] can be quite substantial, and (2) none of the existing inventories were made specifically with phonological similarity in mind. The novel inventory was constructed through participant-elicited phonological neighbors in two phonological association tasks. Phonological association tasks have been used with both nonword [67–71] and word stimuli [2, 72, 73] and provide information pertinent to syllable segmentation and the units being manipulated in that participants are asked to create minimal pairs. Minimal pairs have long played an important role in the identification of phonological units [74, 75] because of their ability to distinguish allophones from phonemes. In both tasks, we identified the respective salience of syllabic and tonal units according to edit distance (the difference between two words in number of segmental units), edit type (whether the manipulation between one word and the next is made through the addition, deletion, or substitution of a segmental unit), and edit location (i.e., the structural unit that a manipulated segment corresponds to in a fully segmented syllable: C_G_V_X_T).

In Experiment 1 we used our participants' productions to evaluate 2 existing annotation systems. We then constructed a new annotation system based on the gaps where either inventory disagreed with our participants' productions. In Experiment 2 we then validated which system optimally represented Mandarin phonology in light of phonological similarity.

The optimal annotation system was then used to construct sixteen phonological networks (8 with tone and 8 without tone), each built from an existing proposal, or suggested permutation, of Mandarin syllable segmentation. To avoid confusion between terms such as network, or schema, we introduce the term phonological segmentation neighborhood (PSN). Each of the sixteen PSNs is a representation of the same lexicon built upon a different schematic representation. While they all share the same lexical items, they differ in what constitutes neighbors. For example, given the phonological word *xiang4*, (as written in Mandarin Romanization, a.k.a. pinyin) the neighbors for the tonal fully segmented PSN (C_G_V_X_T) differ from its nontonal equivalent (C_G_V_X) by three items (*xiang1*, *xiang2*, and *xiang3*). Differences in degree and other topological network statistics accordingly arise between each PSN due to combinations of segmental units and whether or not tone is included in the calculation of similarity between words. The topological network characteristics of each PSN were analyzed similar to the analysis of [14].

Finally, an analysis was performed to identify which of the possible schematic representations of the Mandarin phonological mental lexicon best represented the task demands. We proceeded under the modeling assumption that a given target word within the metrical frame of the Levelt model [76], or the phonological representation frame of the Dell model [77], would share the same segmentation properties as the words they are connected to in long-term memory. We exploited the differences in local network features (i.e., word level) between the sixteen PSNs in a model selection procedure. Reaction times from the second association task were fitted to multiple lexical statistics per PSN with the goal of identifying which best represented the underlying mental procedure of retrieving phonological neighbors. Due to previous findings in Mandarin speech production studies, our hypothesis was that an unsegmented syllable, either tonal (CGVX_T) or nontonal (CGVX), would be identified in our modeling procedure and that like [70], greater density, as calculated on the network's degree, would facilitate mental search.

2. Experiment 1

2.1. Methods

2.1.1. Participants. Thirty-four native-Mandarin speaking participants (Female: 21; Age: M, 24.74; SD, 5.29) took part in this experiment. None of the participants reported a history of speech or hearing disorders. Prior to the experiment, participants were asked to complete a short biographical survey. Contents of the survey included, besides age and sex, the name of their home province, self-rated spoken fluency on a scale of 1 (beginner) to 10 (native speaker) in English (M: 6.26; SD: 1.11), and other Chinese languages/dialects and/or other non-Chinese languages. From their home province, we classified the speakers into two groups based on whether the region was traditionally a Mandarin (Guanhua) speaking region (Guanhua: 21; non-Guanhua: 13). To represent increased competition between similar Chinese languages/dialects, we summed the number of Chinese languages/dialects for all self-rated values from 3-10. This gave us a value that roughly reflects the number of Chinese languages/dialects (M: 2.14; SD: 0.56) that would have words similar to our target Mandarin stimuli. All participants reported native-level proficiency in Mandarin.

The Hong Kong Polytechnic University's Human Subjects Ethics Subcommittee (reference number: HSEARS20140908002) reviewed and approved the details pertinent to all experiments conducted in this study prior to beginning recruitment. The participants gave their informed consent and were compensated with 50HKD for their participation.

2.1.2. Stimuli. The material consisted of 155 Mandarin monosyllabic words, which can be seen in Table 1. The stimuli belonged to three groups according to the phase in which they were given to the participants: Example minimal pairs, 32; Practice, 10; Test, 113. A female speaker from the Beijing area produced all of the stimuli by speaking target monosyllables at a normal speaking rate 5 times into a high-quality microphone. Clearly produced items that were closest to the group mean in length were chosen. The pronunciation of each monosyllabic word was verified through transcriptions done by native-Mandarin speaking volunteers. Stimuli that did not have full agreement between transcribers were rerecorded and rerated until all stimuli were verified by at least 10 volunteers. All stimuli were edited using Audacity 2.1.2 and were 415ms in length.

Table 1: Experiment 1 stimuli.

Example pairs							
bi3~bian3		*chi3~zi3*		*diu1~di1*		*fo2~fei2*	
huang2~hua2		*ka3~kua3*		*lie4~luo4*		*mian2~miao2*	
miu4~you4		*nie4~nue4*		*ou1~sou1*		*piao4~pao4*	
ran2~rang2		*shan1~shan4*		*tian2~tuan2*		*zhe4~zhen4*	
Practice							
	bing1	*cai2*	*chu1*	*fa3*	*guai4*		
	mei2	*reng2*	*shuo1*	*song4*	*zui4*		
Test							
Base Rime	+C	+C		Base Rime	+C	+C	
a1	*ba3*	*ma1*		*wang4*	*kuang2*	*zhuang1*	
ai4	*gai1*	*zai4*		*wei2*	*chui1*	*tui3*	
an4	*nan2*	*san1*		*wen4*	*hun4*	*zun1*	
ang2	*shang4*	*tang3*		*weng1*			
ao4	*kao4*	*lao3*		*wo3*	*cuo4*	*huo2*	
e4	*che1*	*de2*		*wu3*	*fu4*	*ru2*	
ei4	*hei1*	*pei2*		*ya4*	*dia3*	*xia4*	
en1	*fen4*	*gen1*		*yan3*	*tian1*	*qian2*	*pian4*
	deng3	*zheng4*		*yang3*	*niang2*	*xiang3*	
	ren2	*sen1*		*yao4*	*tiao2*	*xiao3*	
er2				*ye2*	*bie2*	*jie1*	
	di4	*li3*	*ni3*	*yi1*	*ji1*	*qi3*	
	lin2	*pin1*		*yin1*	*jin4*	*xin1*	
	ming2	*ting1*		*ying2*	*jing3*	*qing3*	
	ri4	*si3*		*yong4*	*qiong2*	*xiong2*	
	bo1	*mo2*		*you3*	*niu2*	*qiu2*	
	hong2	*cong2*		*yu3*	*lv4*	*nv3*	
ou4	*hou4*	*rou4*		*yuan2*	*juan4*	*quan2*	
wa1	*gua4*	*zhua1*		*yue4*	*jue2*	*xue2*	
wai4	*kuai4*	*shuai4*		*yun4*	*lun2*	*xun2*	
wan2	*guan3*	*suan4*					

The example minimal pairs were exposed to the participants prior to the practice phase of the experiment. The idea of providing auditory examples of sound similarity came about during piloting. Upon given instructions to create minimal pairs or similar sounding syllables, our pilot participants were by and large unsure of what constituted similarity. Luce and Large [71] avoided this possible pitfall by providing their participants the one-phoneme difference rule, while Wiener and Turnbull [70] made it explicit which segment was to be manipulated in three of their four experimental blocks. We chose to provide example pairs because we did not want to bias our participants towards a tonal fully segmented syllable (C_G_V_X_T); however, it was not possible to provide a perfectly even example per each segmentation schema specifically because syllables can be interpreted in multiple ways depending on the number of units in the syllable, or the interpretation of the segments within the syllable. For example, the syllable pairs *bi3~bian3* can be interpreted as both C_GVX_T and CG_VX_T with the Z&L inventory (/pi^{214}/, /pian^{214}/), while only representing C_GVX_T with the Lin inventory (Lin: /pi^{214}/, /pjɛn^{214}/). Of the 17 example pairs presented to our participants, 7 consisted of a single-segment manipulation according to both inventories (*chi3~zi3*; *ou1~sou1*; *miu4~you4*; *nie4~nue4*; *ka3~kua3*; *piao4~pao4*; *shan1~shan4*), while 9 consisted of multiple segment manipulations according to either Lin or Z&L (*fo2~fei2*; *ran2~rang2*; *zhe4~zhen4*; *huang2~hua2*; *tian2~tuan2*; *lie4~luo4*; *mian2~miao2*; *bi3~bian3*; *diu1~di1*).

The test stimuli were created with the goal of representing all syllable structures in the Mandarin language. This was done by adding two lexical items per each base rime syllable from the syllable inventory through the addition of a consonant, regardless of lexical tone. For example, the addition of the consonants, /k/ and /ts/, to the nontonal base syllable /ai/ produced *gai1*, *zai4*, and *ai4*, respectively. Noteworthy about the choice of stimuli, which can be seen in Table 1, are some peculiarities due to the nature of the syllable inventory. First, the base rime syllables, *er2*, and *weng1*, do not have corresponding initial consonant phonological neighbors. Conversely, the pinyin syllables ending in "*eng*" (Lin and Z&L: /əŋ/), such as *deng3*, and *zheng4*; "*i*" as found in *ri4, si3* (Lin: /ɻ̩/ and Z&L: /ɿ/); and "*ong*" (Lin and Z&L:

/uŋ/), located in *hong2* and *cong2*, do not have corresponding base rime syllables. Next, based on phonotactic concerns and a high incidence of certain onsets, we included extra entries. The pinyin onset consonants "*j*" /tɕ/, "*q*" /tɕʰ/, and "*x*" /ɕ/ only cooccur with a small range of rimes, most notably accompanied by the three glides (Lin: /j, ɥ, w/; Z&L: /i, y, u/). Their greater occurrence with glide rimes meant that choosing them was unavoidable and would consequently lead to their overrepresentation in the stimuli set ("*j*" /tɕ/: 6 stimuli; "*q*" /tɕʰ/: 6 stimuli; "*x*" /ɕ/: 7 stimuli). We thus added onsets with lower occurrence for the pinyin rimes ending in "*i*" (*di4, li3, ni3*), "*in*" (*lin2, pin1*), "*ing*" (*ming2, ting1*), and "*en*" (*ren2*, *sen1*).

2.1.3. Procedure. Seated in a quiet room in front of a computer running E-Prime 2.0 [78] and wearing headphones equipped with an adjustable microphone, each participant was exposed to three phases: pretraining, practice, and test. Prior to beginning the experiment participants were instructed to not produce nonitems, which included syllables that do not correspond to an existing Chinese character. For the pretraining phase, participants were told to listen and not respond as they were exposed to 17 word pairs as examples of similar sounding syllables. Each pair was presented according to the same procedure: an auditory stimulus was presented during a blank screen that lasted the word's duration followed by a slide that read "听起来像" (sounds like) for 500ms, that was then immediately followed by its minimal pair during a blank screen that lasted the duration of the stimulus. Between each pair, a dark grey slide that featured, "…", in the center of the screen remained for 2000ms.

For the practice phase participants were told to produce a similar sounding syllable for each of the 10 items. Each stimulus was presented on a blank screen with no time limit. Participants were told that their spoken responses would advance the next trial by activating the PST Serial Response Box. A pause of 500ms followed each participant's response, followed by a slide that read "下一个词" (next word) for 500ms before the next trial. The test phase followed the same procedure for all 113 randomized test items. The entire task took an average of 15 minutes to complete. The audio was recorded on a second computer using Audacity 2.0.6 for offline analysis.

2.2. Results and Discussion. Two native-Mandarin speaking volunteers transcribed into pinyin the participants' spoken productions, with an agreement rate of 93%. A third transcriber resolved disagreements or classified unresolved items as nonitems. The pinyin responses were then translated into a sampa (ascii phonological transcription) that accommodated both the Lin and Z&L syllable inventories.

Our participants responded with large numbers of legal syllables that corresponded to existing Chinese characters, but were not monosyllabic words. We did not discount these items due to their qualification as lexical items. The online dictionary www.zdic.net [79] was used to classify nonitem status by identifying whether a given syllable corresponded to an existing Chinese character. Zdic.net, which includes definitions and pronunciations for 75,983 characters, has been used as a resource in the disambiguation of out-of-vocabulary words in several studies [80–83].

Missing (67), identical responses (138), and nonitems (260) were excluded, accounting for 12.16% of the total 3,842 observations.

2.2.1. Syllable Inventory Creation. In the current section, we detail the creation of a novel syllable inventory through the use of our participants' productions. It is first important to note that neither the Lin nor Z&L inventory, which will be used in the process described below, was constructed specifically according to phonological similarity. While Lin's inventory was informed through phonetic analysis, the Z&L inventory was created to be used in computational models of lexical processing. They are valuable for the current purpose because they have critical differences. The two inventories differ according to glides, as can be seen in syllables such as *ying2*, and *qing3* (Lin: /jəŋ^{35}/, /tɕʰjəŋ^{214}/; Z&L: /iŋ^{35}/, /tɕʰiŋ^{214}/), *yu3* and *yue4* (Lin: /ɥy^{214}/, /ɥe^{51}/; Z&L: /y^{214}/, /yɛ^{51}/), and *hun2* and *kuai4* (Lin: /xwən^{35}/, /kwai^{51}/; Z&L: /xuən^{35}/, /kuai^{51}/). They also differ according to certain vowels such as those found in the syllables *ou4*, and *hou4* (Lin: /ou^{51}/, /xou^{51}/; Z&L: /əu^{51}/, /xəu^{51}/), and *ye1*, *yan2*, and *juan3* (Lin: /je^{55}/, /jɛn^{35}/, /tɕɥɛn^{214}/; Z&L: /iɛ^{55}/, /ian^{35}/, /tɕyan^{214}/).

The fact that the two inventories differ should reminde us of Chao's nonuniqueness theory [84]. The uniqueness theory held that due to there being more than one way to represent a phonological system, there was no absolute better inventory per a given language, but rather an inventory more appropriate for a given purpose. Our purpose in creating a novel inventory was to ensure that a network built upon phonological similarity depended on a syllable inventory equally constructed on phonological similarity.

In the creation of the inventory, we did not seek to redefine Mandarin phonology through the classification of novel phonemes, but instead compare and contrast existing inventories with our participants' minimal pair creations so as to choose which phonemes best accounted for their productions. Thus, we first sought to identify where our participants' minimal pairs disagreed with the annotations of either the Lin and/or Z&L inventories. Agreement between the annotation systems and our participants' productions was assessed through the calculation of mean edit distance per stimuli. High agreement meant that a given stimuli's mean edit was near 1. Prior to calculating mean edit distance per stimuli, tonal neighbors were removed due to their segmental units being identical.

Stimuli of both high and low agreement were informative as to identifying changes in transcriptions that would follow our participants' minimal pair productions. For instance, the stimuli *an4*, which garnered a mean edit of 1.42 for both Z&L and Lin, garnered a lower mean edit of 1.26 for the newly formed Neergaard and Huang inventory (N&H). This was due to modifying the rime, /ɑŋ/, to /aŋ/ (N&H: /an^{51} ~ aŋ^{51}/, edit = 1; Lin and Z&L: /an^{51} ~ ɑŋ^{51}/, edit = 2). The mean edit for the stimuli, *qing3*, (Lin: 2.86; Z&L: 1.71; N&H: 1.71) illustrated that the addition of the glide, /j/, in

the Lin inventory created lower agreement (Lin: /tɕʰjəŋ^{214} ~ tɕʰin^{214}/, edit = 3; Z&L and N&H: /tɕʰiŋ^{214} ~tɕʰin^{214}/, edit = 1).

Another means to identify low agreement, and thus a means to improve the N&H inventory, was through targeting specific annotation choices. Lin's glide annotations /j,w,ɥ/ were shown to have lower agreement across multiple minimal pairs, such as *xin1~xian1* (Lin: /ɕin^{55} ~ ɕjɛn^{55}/, edit = 2; Z&L: /ɕin^{55} ~ ɕian^{55}/, edit = 1; N&H /ɕin^{55} ~ ɕiɛn^{55}/, edit = 1), *mo2~mu2* (Lin: /mwo^{35} ~mu^{35}/, edit = 2; Z&L: /mo^{35} ~ mu^{35}/, edit = 1; N&H /muo^{35} ~mu^{35}/, edit = 1), *quan2~qun2* (Lin: /tɕʰɥɛn^{35} ~tɕʰyn^{35}/, edit = 2; Z&L: /tɕʰyan^{35} ~tɕʰyn^{35}/, edit = 1; N&H /tɕʰyɛn^{35} ~tɕʰyn^{35}/, edit = 1). Similarly, both Lin and Z&L showed low agreement according to pinyin syllables that have the "ong" rime, annotated as /uŋ/. Participants preferred to produce phonological neighbors that contained /o/ rather than /u/, as can be seen in the example, *yong4~you4* (Lin: /juŋ^{51} ~jou^{51}/, edit = 2; Z&L: /iuŋ^{51} ~iəu^{51}/, edit = 2; N&H /ioŋ^{51} ~ioʊ^{35}/, edit = 1).

There were two cases in which the N&H inventory collapsed existing categories. Participants made neighbors ignoring the difference between the Lin and Z&L phonemes /ɤ/ and /ɘ/. By collapsing them into the single phoneme, /ɘ/, N&H reduced the mean edit compared to both Lin and Z&L for syllables such as *er2* (Lin: 2.5: Z&L: 2.5; N&H: 2.13), as is illustrated in the pair *er2~e2* (Lin: /ɤr^{35} ~ ɘ^{35}/, edit = 2; Z&L: /ɤɻ^{35}~ ɘ^{35}/, edit = 2; N&H /ɘr^{35} ~ ɘ^{35}/, edit = 1). N&H collapsed the Lin distinction of /ɔ/ and /ou/, and the Z&L distinction of /o/ and /əu/, into the single diphthong /oʊ/. This decision was based on garnering lower mean edits for the N&H inventory for syllables such as *bo1* (Lin: 1.7: Z&L: 2; N&H: 1.65), *mo2* (Lin: 1.94: Z&L: 1.94; N&H: 1.76), and *huo2* (Lin: 1.71: Z&L: 1.71; N&H: 1.59). It was also based on edit distances for minimal pairs such as *ou4~o1* (Lin: /ou^{51} ~ ɔ^{55}/, edit = 3; Z&L: /əu^{51} ~o^{55}/, edit = 3; N&H /oʊ^{51} ~oʊ^{35}/, edit = 1).

Examples of 10 syllables across the Lin, Z&L and N&H syllable inventories can be seen in Table 2. See Table 3 for the N&H phoneme inventory.

A final step in evaluating the three syllable inventories consisted of an ANOVA between edit distance values (excluding tonal neighbors): Lin (M: 1.90; SD: 0.92); Z&L (M:1.72; SD: 0.81); N&H (M: 1.67; SD: 0.79). The main effect was significant ($F=43.46$; $p < 0.001$). Pair-wise comparisons showed that both the Z&L ($p < 0.001$) and N&H ($p < 0.001$) inventories outperformed the Lin inventory. No significant difference was found between the edit distance values of Z&L and N&H.

2.2.2. Edit Information. Edit distance (including tonal neighbors) was used to calculate similarity according to the Lin, Z&L and N&H syllable inventories. Single-segment edits made up between 61 and 67% of correct responses (Lin: 61%; Z&L: 65%; N&H: 67%) while two-segment edits comprised over 20% (Lin: 23%; Z&L: 24%; N&H: 24%), three-segment edits accounted for around 10% (Lin: 12%; Z&L: 9%; N&H: 8%), and four- and five-segment edits combined were roughly 3% of correct responses (Lin: 4%; Z&L: 2%; N&H: 2%).

Table 2: Comparisons between syllable inventories.

Pinyin	Lin	Z&L	N&H
e	/ɤ/	/ɤ/	/ə/
ai	/ai/	/ai/	/aɪ/
ei	/ei/	/ei/	/eɪ/
o	/ɔ/	/o/	/oʊ/
ou	/ou/	/əu/	/oʊ/
ao	/ɑu/	/ɑu/	/aʊ/
ang	/ɑŋ/	/ɑŋ/	/aŋ/
yu	/ɥy/	/y/	/y/
yue	/ɥe/	/yɛ/	/yɛ/
yuan	/ɥɛn/	/yan/	/yɛn/

The single-segment edits can be further described by addressing which segments within the fully segmental schema (C_G_V_X_T) were altered to make a minimal pair (edit location) and the edit type (addition, deletion, or substitution) that was made per manipulation. The predominant segment to be changed within single-edit manipulations was that of lexical tone, which accounted for 34% of all correct responses across the three inventories. The second most often manipulated segment was the initial consonant, accounting for around 18% (Lin: 18%; Z&L: 18%; N&H: 19%). The remaining segments featured in less than 8% of all correct responses. The medial glide was manipulated roughly 2% across all inventories. The monophthong was around 5% (Lin: 4%; Z&L: 5%; N&H: 5%) and the final X between 3 and 8% (Lin: 3%; Z&L: 6%; N&H: 8%). As for edit type, the majority of manipulations made for correct responses were made through substitution (Lin: 55%; Z&L: 55%; N&H: 56%). Edits made from the addition of a segment accounted for between 5 and 8% (Lin: 5%; Z&L: 7%; N&H: 8%), while deletion type edits accounted for roughly 2% of correct responses (Lin: 1%; Z&L: 3%; N&H: 3%).

2.3. Discussion. In this experiment, participant-elicited minimal pairs served in the creation of a novel syllable inventory as well as provide insight into awareness of the units within the Mandarin syllable. As it stands currently, the Lin inventory was outperformed by both Z&L and the newly created N&H inventories with no statistical difference between the latter two. In terms of segmentation, while results show that all units are subject to manipulation, there was a strong prevalence towards two principle units: the unsegmented syllable and lexical tone. These results perhaps do not in themselves lessen the status of each segment but emphasize a tonal route for mental search of minimal pairs. Of another note on this experiment's findings is the fact that our Mandarin-speaking participants produced a lower percentage of single-edit responses (Lin: 61%; Z&L: 64%; N&H: 67%) than did the English speaking participants of [71] at 71%, [2] at 74.5%, and [73] at 84.21%. It is likely safe to assume that the lower values for the Luce and Large [71] study were the result of their participants having given spoken responses, whereas in both studies by Vitevitch and colleagues [2, 73] the recorded responses were written. Another reason for a lower percent

Table 3: N&H phoneme inventory.

	IPA	Sampa	Pinyin word	Sampa word	Ortho word		IPA	Sampa	Pinyin word	Sampa word	Ortho word
Vowels	a	a	b**a**3	p**a**3	把	Plosives	p	p	**b**u4	**p**u4	不
	ə	@	sh**e**4	S**@**4	蛇		p^h	P	**p**ao3	**P**aU3	跑
	e	e	g**e**i3	k**e**I3	给		k	k	**g**e0	**k**@0	个
	ɛ	E	y**e**3	i**E**3	也		k^h	K	**k**e4	**K**@4	课
	ɨ	1	zh**i**1	Z**1**1	之		t	t	**d**ou1	**t**oU1	都
	i	i	d**i**4	t**i**4	第		t^h	T	**t**a1	**T**a1	他
	ɪ	I	su**i**4	sue**I**4	岁	Fricatives	s	s	**s**uo3	**s**uo3	所
	o	o	ru**o**4	ru**o**4	若		f	f	**f**ang4	**f**aN4	放
	ʊ	U	cho**u**3	Co**U**3	丑		x	x	**h**ui4	**x**ueI4	会
	u	u	**w**o3	**u**o3	我	Affricates	ʂ	S	**sh**i4	**S**14	是
	y	y	**yu**an2	**y**En2	元		ɕ	X	**x**ia4	**X**ia4	下
Nasals	m	m	m**a**1	m**a**1	妈		tɕ	J	**j**iu4	**J**ioU4	就
	n	n	**n**eng2	**n**@N2	能		$tɕ^h$	Q	**q**ing3	**Q**iN3	请
	ŋ	N	xia**ng**3	Xia**N**3	想		ts^h	c	**c**ong2	**c**oN2	从
Liquids	l	l	**l**ie4	**l**iE4	列		$tʂ^h$	C	**ch**u1	**C**u1	出
	r	r	**r**ang4	**r**aN4	让		ts	z	**z**i4	**z**14	字
							tʂ	Z	**zh**e	**Z**@4	这

of single-edit manipulations might be due to the nature of our example pairs. Given our participants did produce examples of manipulations at all units, we decided in a second phonological association task to validate the performance of the three annotation systems using an explicit single-edit example, as was done in [71]. We expected this to increase the number of single-edit manipulations and in turn aid in discriminating which of the three inventories best aligns with our participants' manipulations.

3. Experiment 2

3.1. Methods

3.1.1. Participants. Of the thirty-four newly recruited native Mandarin speakers, one participant was removed from further consideration because they rated Mandarin as being their nondominant language. The thirty-three remaining participants reported native-level fluency in Mandarin (Female: 22; Age: M, 23; SD, 4). None of the participants reported speech, hearing, or visual disorders. Participant characteristics did not differ from those from the first experiment in self-rated spoken English proficiency (M: 6.55; SD: 1.23), traditionally Mandarin-speaking region (Guanhua: 23; non-Guanhua: 10), or number of Chinese languages/dialects spoken (M: 2.39; SD: 0.74).

As with Experiment 1, all participants gave their informed consent and were compensated with 50HKD for their participation as stipulated by the local ethics committee.

3.1.2. Stimuli. The stimuli for Experiment 2 consisted of 200 test items and 10 practice items. Two items, however, were discounted for not existing in the www.zdic.net dictionary, reducing our total to 198 stimuli test items. Ninety-seven stimuli were used from the Experiment 1 stimuli set. The 101 new stimuli were created with the same voice and procedure. A determining factor in the selection of new stimuli and rejection of stimuli from Experiment 1 was whether or not lexical frequency could be accounted for using the word list from Subtlex-CH [85]. The Subtlex-CH wordlist, created through aggregated movie subtitles, was chosen because the subtitle genre has been shown to greater predict language processing tasks when compared to frequencies calculated from written sources [85, 86]. Further information on the transcription of the wordlist's 99,125 Chinese characters to pinyin and subsequent sampa can be found in the Database of Mandarin Neighborhood Statistics [87].

As can be seen in Table 4, we included each of the base rime syllables accompanied by between three to six consonant neighbors. As with the stimuli in Experiment 1, certain stimuli lacked base rimes, while others did not have consonant neighbors. Those stimuli with only three consonant neighbors were tied to the base rime syllables *yuan2* and *yong3*. They were limited to three consonant neighbors because there are only the three possible onsets, "*j*, *q*, *x*" /tɕ, $tɕ^h$, ɕ/, available for these base rimes and we did not want to repeat a nontonal syllable.

3.1.3. Procedure. The procedure differed from Experiment 1 in that no pretraining phase was given. In place of this, participants were given oral instructions as to what consisted of a similar sounding monosyllable through the use of the target syllable ling2 (e.g., 零), which they were told had the neighbors: ling4 (e.g., 另), ning4 (e.g., 宁), lang2 (e.g., 狼), and lin2 (e.g., 磷). All other procedural aspects were the same as in Experiment 1.

3.2. Results. As with Experiment 1, the same transcription procedure was followed. Missing (53), identical (210), nonitem (505), and semantically related responses (3) were excluded from the analysis.

Table 4: Experiment 2 test stimuli.

Base Rime	+C	+C	+C	+C	+C	+C
a1	*ba3*	*fa3*	*ka3*	*la1*	*ma1*	
ai4	*cai2*	*dai4*	*gai1*	*mai3*	*zai3*	
an4	*ban1*	*nan2*	*ran2*	*san3*	*shan1*	*zhan4*
ang2	*gang1*	*rang4*	*sang1*	*shang4*	*tang3*	
ao1	*gao3*	*lao3*	*mao1*	*pao4*	*zao3*	
e4	*che1*	*de2*	*ge1*	*ke3*	*zhe4*	
ei4	*bei3*	*hei1*	*fei2*	*mei2*	*pei2*	
en1	*fen4*	*hen3*	*men2*	*ren2*	*zhen4*	
	feng1	*reng1*	*sheng3*	*zeng4*	*zheng4*	
er3						
	ci4	*chi1*	*ri4*	*shi2*	*si3*	*zi3*
	hong2	*cong2*	*nong4*	*song4*	*zhong1*	*zong3*
ou4	*bo1*	*fo2*	*hou4*	*mo2*	*rou4*	
wa1	*gua4*	*hua2*	*kua3*	*shua1*	*zhua1*	
wai4	*guai4*	*huai4*	*kuai4*	*shuai4*		
wan2	*guan3*	*huan4*	*ruan3*	*suan4*	*tuan2*	
wang4	*guang1*	*huang2*	*shuang3*	*kuang2*	*zhuang1*	
wei2	*chui1*	*sui4*	*shui3*	*tui3*	*zui4*	
wen4	*hun4*	*lun2*	*gun3*	*zun1*		
weng4						
wo4	*cuo4*	*duo1*	*huo2*	*ruo4*	*shuo1*	
wu3	*chu1*	*du4*	*fu4*	*ru2*	*zhu4*	
ya4	*dia3*	*jia1*	*lia3*	*qia1*	*xia4*	
yan3	*bian3*	*mian2*	*pian4*	*qian2*	*tian2*	
yang3	*jiang1*	*liang2*	*niang2*	*qiang2*	*xiang3*	
yao4	*diao4*	*miao2*	*piao4*	*tiao1*	*xiao3*	
ye2	*die1*	*jie1*	*bie2*	*lie4*	*xie2*	
yi1	*ji1*	*li3*	*ni3*	*qi3*	*di1*	
yin1	*jin4*	*qin2*	*xin1*	*pin1*	*min2*	
ying2	*bing1*	*jing3*	*qing3*	*ting1*	*ming2*	
yong3	*jiong3*	*qiong2*	*xiong2*			
you3	*diu1*	*jiu3*	*liu1*	*niu2*	*qiu2*	
yu3	*ju2*	*lv4*	*nv3*	*qu4*	*xu1*	
yuan2	*juan4*	*quan2*	*xuan3*			
yue4	*jue2*	*nue4*	*que1*	*xue2*		
yun4	*jun1*	*kun4*	*qun2*	*xun1*		

We again evaluated which of the three syllable inventories optimally accounted for phonological similarity according to our participants' minimal pair productions. Repeating the same procedure, we excluded tonal neighbors prior to conducting an ANOVA on the edit distances of the three inventories: Lin (M: 1.86; SD: 0.87); Z&L (M: 1.71; SD: 0.78); N&H (M: 1.64; SD: 0.75). The main effect was significant ($F=65.3$; $p < 0.001$). Pair-wise comparisons showed that both the Z&L ($p < 0.001$) and N&H ($p < 0.001$) inventories outperformed the Lin inventory. Meanwhile the N&H inventory outperformed the Z&L inventory ($p = 0.002$).

Edit information, including edit distance, location, and type, was then calculated for the three inventories. All calculations were derived from the correct responses, including tonal neighbors.

Single-segment edits accounted for between 68 and 73% (Lin: 68%; Z&L: 71%; N&H: 73%). Two-segment edits accounted for around 21% (Lin: 20%; Z&L: 22%; N&H: 20%). Three-segment edits accounted for 5 to 10% (Lin: 10%; Z&L: 7%; N&H: 5%), and four- and five-segment edits combined were between 1 and 2% (Lin: 2%; Z&L: 1%; N&H: 1%).

Edit location for the single-segment edits again was dominantly at the lexical tone position, accounting for 46% of correct responses. The second most common manipulation, at 15%, again occurred at the initial consonant. The remaining syllable position saw a combined 5 to 16% instance of manipulation (Final X: Lin: 3%; Z&L: 5%; N&H: 7%, monophthong: Lin: 3%; Z&L: 4%; N&H: 4%, and medial glide: 1%).

Edit type again was dominantly substitution, occurring between 64 and 66% (Lin: 64%; Z&L: 65%; N&H: 66%). Edits

made from the addition of a segment accounted for between 3 and 5% (Lin: 3%; Z&L: 4%; N&H: 5%), while deletion type edits accounted for roughly 2% of correct responses (Lin: 1%; Z&L: 2%; N&H: 2%).

3.3. Discussion. The second phonological association task identified an optimal annotation system while providing repeated evidence of segmentation biases, specifically towards the manipulation of lexical tone while maintaining a whole syllable. Changes made in comparison to Experiment 1 included (1) changing instructions so as to provide a single-edit example and (2) increasing the number of stimuli, which respectively increased the percentage of single-edit productions (Experiment 1: 61-65%; Experiment 2: 67-73%) and gave greater discriminative power in identifying the newly formed N&H inventory as the optimal syllable inventory. In applying the principle of the nonuniqueness theory [84], we can surmise that the N&H inventory, built on phonological similarity, is the optimal choice to model Mandarin vocabulary in a phonological network that is as well constructed on phonological similarity.

4. Phonological Segmentation Neighborhoods

The goal of previous investigations into phonological networks has been to infer aspects of the nature of language processing and/or the development of the lexicon from constructed, random, and real language graphs. A number of topological measures have been used. Those that we will be reporting on come from the same six studies [9, 11–15]. The igraph package in R [88] was used for the construction and measurement of all the following graphs.

The first value to consider is degree. When expressed at the word level, annotated as k, it is the number of single-edit neighbors a given word has. At the topological level, annotated as $\overline{k}$, it is the mean of neighbors per node across the entire network, or from the network's largest fully connected subgraph, also referred to as the network's giant component. The giant components of phonological networks studied thus far have been shown to take between 32-66% of available nodes, which is lower than phonological networks built from artificial corpora [9]. All topological measures featured below will be reported from each network's giant component.

Interconnectedness between neighbors is expressed through the measure known as clustering coefficient. At the word level, annotated as CC, it is the proportion of neighbors who are also neighbors of each other. The mean value taken at the macrolevel is annotated as $\overline{CC}$. Phonological networks have shown $\overline{CC}$ values of between 0.191-0.383 for giant components.

Another measure of the relationship to the density of interconnectedness is the correlation between the density of a given node and the density of its neighbors, known as mixing by degree (M) [89, 90]. When positive, referred to as assortative mixing by degree, the value indicates that the network's nodes tend to have dense nodes connected with other dense nodes. Thus far, phonological networks, whether from real vocabulary lists or artificially constructed vocabularies, have all been assortative. Networks constructed from real vocabulary lists have shown M values between 0.556 and 0.762.

A final measure of network density, which we annotate as $\overline{L}$, is that of a networks' mean shortest path length. It is the average distance between a given node and the rest of the nodes within the giant component and thus a measure of spreading through the network. Phonological networks have been shown to have $\overline{L}$ values between 6.08 and 10.40 for their giant components.

We categorized the PSNs as to whether they had small world characteristics. A network that shows small world characteristics ($\overline{CC} > \overline{CC}\text{-}RN$; $\overline{L} > \overline{L}\text{-}RN$) has values of $\overline{CC}$ and $\overline{L}$ greater than those generated from random networks ($\overline{CC}\text{-}RN$, $\overline{L}\text{-}RN$). We report on the mean and standard deviations of 10 iterations of Erdos-Renyi random networks constructed from the same number of nodes and edges as the networks they were compared to. The small world structure is believed to aid speed during search [91] and thus generalize to the spreading of lexical activation [13, 15]. All language networks thus far have shown small world characteristics in their giant components.

Finally, we also report on whether the PSNs' degree distributions can be described as having a power-law degree distribution. Vitevitch [13] drew attention to the distributions of phonological networks as a possible cue to vocabulary formation due to the association between power-law degree distributions with self-organization [92] and the two principles underlying scale-free networks: growth and preferential attachment [93, 94]. Preferential attachment describes a process whereby new nodes establish connections to already densely connected nodes. A limitation to this association of scale-free characteristics and power-law degree distributions is that scale-free networks can come about from other growth methods [95]. Phonological networks have not shown clear cut power-law distributions, but instead a power-law with cut-off [3, 5, 9, 14, 15]. The term cut-off refers to the process of choosing a starting point from which the distribution is fit [96], meaning that only a portion of a given distribution is being described. The present degree distributions were fitted using [97].

While the initial studies were optimistic about which of the many variables were indicative of cognitive processes [13, 15], few seem to be likely candidates. The construction of pseudo lexicons and their subsequent comparisons to real phonological networks has shown that small world characteristics are not intrinsic to the nature of vocabulary [11, 98] and that preferential attachment is not a likely account of vocabulary growth due to portions of power-laws also occurring from distributions made through random sampling [9]. Turnbull and Peperkamp [12] placed their hope in assortative mixing by degree due to it being the only value to distinguish an English phonological network from 5 types of random graphs. Stella and Brede [9] similarly had higher assortativity for their English network when compared to their constructed networks. Yet, it is not clear whether a difference of either 0.103 [12] or 0.117 [9] between their real networks and their second highest constructed networks is meaningful. Other studies have suggested that word length

Table 5: Mandarin segmentation schemas according to the example monosyllables, *lian2* /liɛn^{35}/ and *liao3* /liaʊ^{214}/.

Without Tone			With Tone		
C_V_C	/l_iɛ_n/	/l_iaʊ/	C_V_C_T	/l_iɛ_n_35/	/l_iaʊ_214/
C_G_V_C	/l_i_ɛ_n/	/l_i_aʊ/	C_G_V_C_T	/l_i_ɛ_n_35/	/l_i_aʊ_214/
C_G_V_X	/l_i_ɛ_n/	/l_i_a_ʊ/	C_G_V_X_T	/l_i_ɛ_n_35/	/l_i_a_ʊ_214/
C_G_VX	/l_i_ɛn/	/l_i_aʊ/	C_G_VX_T	/l_i_ɛn_35/	/l_i_aʊ_214/
C_GVX	/l_iɛn/	/l_iaʊ/	C_GVX_T	/l_iɛn_35/	/l_iaʊ_214/
CG_V_X	/li_ɛ_n/	/li_aʊ/	CG_V_X_T	/li_ɛ_n_35/	/li_aʊ_214/
CG_VX	/li_ɛn/	/li_aʊ/	CG_VX_T	/li_ɛn_35/	/li_aʊ_214/
CGVX	/liɛn/	/liaʊ/	CGVX_T	/liɛn_35/	/liaʊ_214/

plays a unique role in the networks. Network statistics are influenced by word length [9, 12, 98], because of the negative correlation found between length and phonological similarity according to the single-edit metric. Languages with greater morphological richness have shown sparser distributions [14, 98], hinting at cross-linguistic differences based on graph measures.

4.1. Constructing the PSNs. With the N&H inventory validated, it was then used to create a database of neighborhood statistics from all schematic representations proposed or suggested. In order to provide all possible permutations we add the segmented diphthongal schema, previously proposed for Taiwanese speakers [99] in its nontonal (C_G_V_C) and tonal form (C_G_V_C_T). The possibility of diphthongs was proposed for Mandarin by [100]. Table 5 presents sixteen segmentation schemas, each with two example syllables.

Lexical frequencies and subsequent neighborhood frequency counts (the average frequency of a words' neighbors) were again adapted from Subtlex-CH [85] as detailed in the Database of Mandarin Neighborhood Statistics [87]. Prior to calculating phonological neighbors, all homophones were collapsed into single items, and their frequencies summed, i.e., the definition of a phonological word. Each PSN was then created from the top 30,000 most frequent phonological words, roughly the same size as the Mandarin network analyzed by Arbesman et al. [14]. This led to slight differences in degree (PND) from the existing resource of similar structure and content [87] that calculated similarity from the top 17,000 most frequent phonological words. Monosyllables that were featured in the stimuli but that were not present in the Subtlex-CH word list were added for the sake of calculating their degree, but were given a frequency count of 1 and thus were not part of the top 30,000 phonological words from which degree calculations were made. Lastly, we removed edges between monosyllables in the CGVX PSN, consisting of 397 monosyllabic neighbors per target monosyllable word, due to there being no meaningful relationship between them.

4.2. Topology. As can be seen in Table 6, the PSNs exhibit network characteristics both within and outside expected ranges compared to past phonological networks. Unlike previous networks, $\overline{L}$ (2.79-17.72), M (0.454-0.918), and the proportion of the network covered by the giant component (Size: 30.53-88.15%) all showed a large range of values, some of which were double those previously reported. $\overline{CC}$ (0.247-0.628) was perhaps the only measure with relative stability across PSNs. In line with past networks, all PSNs exhibited small world characteristics ($\overline{CC} > \overline{CC}$-$RN$; $\overline{L} > \overline{L}$-$RN$).

In Table 6, we characterize the sixteen Mandarin PSNs according the number of units within each PSNs' maximal syllable (Units). In Figure 1, we see that both Units and lexical tone determine how each PSN patterns according to their network characteristics. What first stands out is the distance the unsegmented PSNs (CGVX, CGVX_T) take from their segmented counterparts. While CGVX_T groups according to the nontonal PSNs in Size (a) and $\overline{L}$ (b), it stands apart in M (c), yet is similar to CGVX in its high $\overline{CC}$ (d). CGVX, meanwhile, has a uniquely high $\overline{k}$ (139.97) and Size, similar to the collaboration networks reported by Newman [101]. $\overline{L}$ in contrast is very low, illustrating that high $\overline{k}$ and Size equate short distances between any given neighbors. Only in M does CGVX pattern according to the nontonal PSNs. The segmented PSNs on the other hand show some gradient distributions. Size shows a negative trend for greater segmentation, particularly for tonal PSNs, which is opposite to the positive trend found in $\overline{L}$. There is no linear effect of segmentation for either M or $\overline{CC}$. M exhibits the only split distribution among the network statistics. Unfortunately, there is no immediate indication why the low M group (C_GVX_T, C_V_C_T, CG_V_X_T) would have roughly half of the values of the high M group (CG_VX_T, C_G_VX_T, C_G_V_C_T, C_G_V_X_T).

In Table 6 we see that not all PSNs contain portions of power-law distributions. Those that did contain power-law portions were both nontonal and tonal and of varying unit lengths (nontonal: CGVX, C_GVX, C_V_C, CG_V_X; tonal: CGVX_T, CG_VX_T, CG_V_X_T, C_G_VX_T, C_G_V_C_T, C_G_V_X_T), which similarly can be said for those that did not (nontonal: CG_VX, C_G_VX, C_G_V_C, C_G_V_X; tonal: C_GVX_T, C_V_C_T). Segmental units were also not to blame seeing as all individual units and their collapsed combinations occur in both distribution groups.

4.3. Syllable Length. Of the top 30,000 phonological words, monosyllables account for 3.80% (n=1,141), disyllables 72.17% (n=21,652), trisyllables 14.84% (n=4453), quadrasyllables 8.73% (n=2618), and less than 1% for the remaining 5-, 6-, and 7-syllable phonological words (n=136). In Figure 2 we

Table 6: Topological network measures of the PSNs' giant components.

	CGVX	C_GVX	CG_VX	C_V_C	CG_V_X	C_G_VX	C_G_V_C	C_G_V_X
Units	1	2	2	3	3	3	4	4
Size	88.15	71.91	70.41	70.52	69.1	70.33	69.13	68.79
$\overline{k}$	139.97	14.53	13.2	10.19	10.54	11.73	9.79	8.93
M	0.577	0.602	0.628	0.613	0.689	0.593	0.622	0.633
$\overline{L}$	2.79	5.31	5.58	6.49	7.47	5.77	6.85	7.65
$\overline{L}$-RN	2.47 (1.93^{e-5})	4.01 (2.75^{e-4})	4.14 (4.09^{e-4})	4.56 (3.16^{e-3})	4.50 (3.16^{e-3})	4.33 (6.59^{e-4})	4.62 (1.13^{e-3})	4.79 (1.14^{e-3})
$\overline{CC}$	0.578	0.319	0.336	0.303	0.435	0.278	0.310	0.340
$\overline{CC}$-RN	5.29^{e-3} (6.25^{e-6})	6.53^{e-4} (4.01^{e-5})	6.10^{e-4} (3.30^{e-5})	4.81^{e-4} (4.19^{e-5})	4.97^{e-4} (3.48^{e-5})	5.36^{e-4} (3.19^{e-5})	4.54^{e-4} (4.86^{e-5})	4.25^{e-4} (3.24^{e-5})
Cut-off/p	226/∗∗	27/∗	24/*NS*	19/∗∗	20/∗∗	37/*NS*	31/*NS*	27/*NS*
	+T	+T	+T	+T	+T	+T	+T	+T
Units	2	3	3	4	4	4	5	5
Size	71.89	49.11	48.88	34.39	31.02	45.1	35.29	30.53
$\overline{k}$	25.64	3.61	4.98	3.10	3.14	4.61	4.47	4.51
M	0.733	0.538	0.918	0.454	0.470	0.894	0.891	0.900
$\overline{L}$	5.40	12.12	12.68	15.15	17.47	12.44	14.74	17.72
$\overline{L}$-RN	3.42 (1.81^{e-4})	7.57 (1.47^{e-2})	6.16 (3.20^{e-3})	8.17 (2.48^{e-2})	8 (2.23^{e-2})	6.40 (7.84^{e-3})	6.37 (9.66^{e-3})	6.24 (8.22^{e-3})
$\overline{CC}$	0.628	0.303	0.460	0.247	0.335	0.358	0.388	0.416
$\overline{CC}$-RN	1.21^{e-3} (1.90^{e-5})	2.16^{e-4} (9.25^{e-5})	3.27^{e-4} (5.98^{e-5})	2.99^{e-4} (1.47^{e-4})	3.85^{e-4} (9.07^{e-5})	3.43^{e-4} (8.30^{e-5})	3.96^{e-4} (9.51^{e-5})	4.91^{e-4} (1.19^{e-4})
Cut-off/p	71/∗∗	10/*NS*	3/∗∗	10/*NS*	6/∗∗	4/∗∗	3/∗∗	3/∗∗

Note: Units = number of units within each PSN's maximal syllable; Size = percent of nodes covered by the giant component; $\overline{k}$ = mean degree; M = mixing by degree; $\overline{L}$ = mean shortest path length; $\overline{L}$-RN = $\overline{L}$ of 10 iterations of random networks (mean standard deviation of $\overline{L}$-RN); $\overline{CC}$ = mean clustering coefficient; $\overline{CC}$-RN = $\overline{CC}$ of 10 iterations of random networks (mean standard deviation of $\overline{CC}$-RN); Cut-off/p = properties of a power-law degree distribution, such that Cut-off denotes where in the distribution the calculation begins, and p, the probability of said distribution accounting for a power-law distribution, expressed as either NS (nonsignificant), ∗ ($p < 0.05$), or ∗∗ ($p < 0.01$)

illustrate the distributions for $\overline{k}$ and $\overline{CC}$, for monosyllables and disyllables, according to the number of maximal syllable units (Units) in each PSN. Because of the difference between segmented and unsegmented PSNs, we will consider them separately.

In Figures 2(a) and 2(b) we see that greater segmentation and the addition of lexical tone led to fewer neighbors for both monosyllables and disyllables. The PSNs with the lowest $\overline{k}$ were built from five units and are both tonal (C_G_V_C_T, C_G_V_X_T). Conversely, the segmented syllables that have only two units, CG_VX and C_GVX, are both nontonal and have the highest $\overline{k}$ of the segmented PSNs. Compared to $\overline{k}$, the story of $\overline{CC}$ for monosyllables and disyllables is less clear. There is no trend between Units and $\overline{CC}$ for monosyllables (Figure 2(c)) from segmented PSNs. Conversely, $\overline{CC}$ among disyllables of segmented PSNs are affected by lexical tone. Figure 2(d) shows that nontonal PSNs all have higher $\overline{CC}$ than tonal PSNs.

To account for the outlier behavior of unsegmented PSNs, we first address the switch in $\overline{k}$ between monosyllables ($\overline{k}$ = 286) and disyllables ($\overline{k}$ = 26). For monosyllables in CGVX_T, every phonological word of a given tone assignment is a neighbor of every other monosyllable with that same tone, leading to 5 distinct subgraphs (tones 0-4). The complete interconnectedness of neighbors for monosyllables means that these words have CC values nearing 1, as seen in Figure 3(c). Disyllabic words of the CGVX_T PSN, on the other hand, have 3 of 4 units that must match with another word to classify as a neighbor and as such do not diverge greatly from the linear relationship between $\overline{k}$ and Units of the segmented PSNs. The increase in Units reduces not just $\overline{k}$, but also $\overline{CC}$.

In contrast to the CGVX_T PSN, the nontonal unsegmented PSN (CGVX) has an opposite switch in $\overline{k}$ for monosyllables ($\overline{k}$ = 117) and disyllables ($\overline{k}$ = 186). The number of neighbors increases for disyllabic words due to the ability of nontonal disyllabic words to link to monosyllables, other disyllables, and trisyllables. Unlike with monosyllables, in which $\overline{k}$ and $\overline{CC}$ distributions differ according to Units, disyllables show a linear relation between Units and both $\overline{k}$ (b) and $\overline{CC}$ (d).

To aid in comprehending the role of segmentation and lexical tone on network features at the word level, in Figure 3 we illustrate the tonal monosyllabic word *niao3* /niaʊ^{214}/ and its nontonal counterpart *niao* /niaʊ/. The nontonal *niao* of CGVX (Figure 3(a)) is the sole monosyllable in

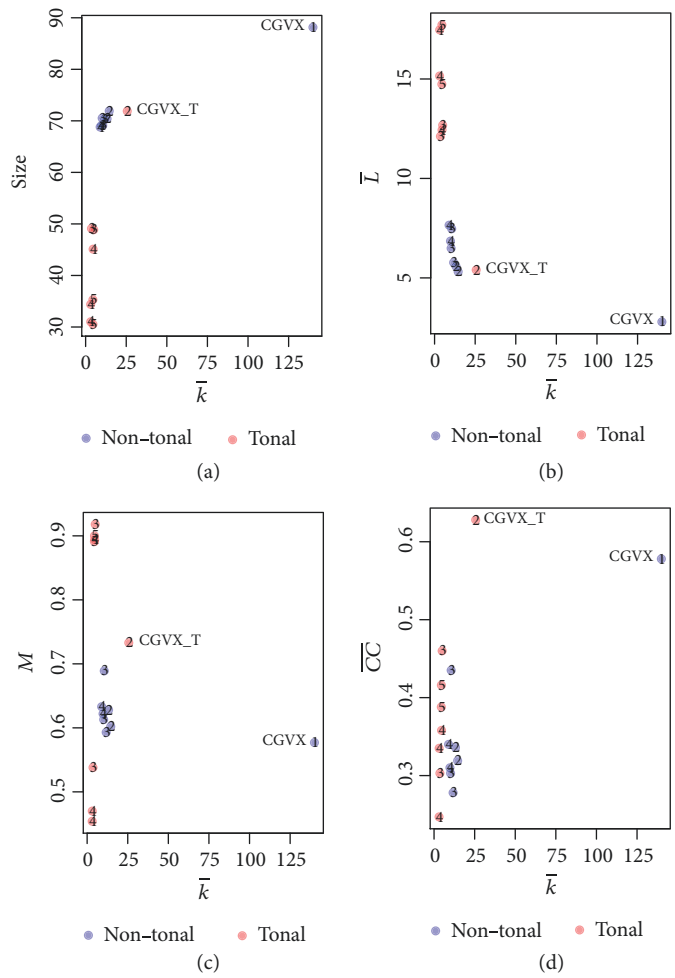

Figure 1: Topological features of each PSN according to $\overline{k}$ (mean degree) and (a) Size (percent of giant component), (b) $\overline{L}$ (mean shortest path length), (c) M (mixing by degree), and (d) $\overline{CC}$ (mean clustering coefficient). Units (the number of units within each PSNs' maximal segmentation schema) is featured within each data point.

a network of disyllables. Through the addition of lexical tone (Figure 3(b)), all disyllables are excluded, and neighbor classification is based on whether the monosyllables share the same tone. Meanwhile, a segmented *niao* (Figure 3(c)) has both monosyllabic neighbors that differ by a single segment, and disyllabic neighbors, such as *ni hao* /ni xao/. This is not the case for *niao3* (Figure 3(d)), which has only other monosyllables as neighbors.

4.4. Discussion. Sixteen Mandarin PSNs were constructed that differed according to both syllable segmentation and lexical tone. Network statistics revealed that both characteristics determined what constituted similarity between phonological words. Greater segmentation and the presence of tone mean less density for segmented neighbors. Plots of the PSNs' $\overline{k}$ was informative as to each segmented network's Size (larger for nontonal PSNs), $\overline{L}$ (larger for tonal PSNs), M (assortative for all PSNs, and split for tonal PSNs), and $\overline{CC}$ (no clear trend). Unsegmented PSNs, in contrast, behaved differently from segmented PSNs for each network measure at both the scale of the giant component and when isolating monosyllables and disyllables. Inspection of word-level graphs illustrated that for monosyllables, the presence of tone limited the choice of available neighbors to other monosyllables, while monosyllables of nontonal PSNs had both monosyllabic and disyllabic neighbors. For unsegmented PSNs, this was exacerbated, such that monosyllabic words from the nontonal unsegmented PSN (CGVX) had only disyllabic neighbors, and monosyllabic words from the tonal unsegmented PSN (CGVX_T) had only monosyllabic tonal neighbors.

We now turn to the principle goal of inspecting topological features of language networks, and phonological networks in particular. Under the conceit that properties of language processing and vocabulary formation can be inferred from phonological networks built from vocabulary lists, we ask whether we can predict which of the sixteen PSNs is the most likely candidate for Mandarin based on previous network measures.

Previous phonological networks showed between 32 and 66% in Size. This range includes all segmented and tonal PSNs, while excluding the nontonal segmented group and

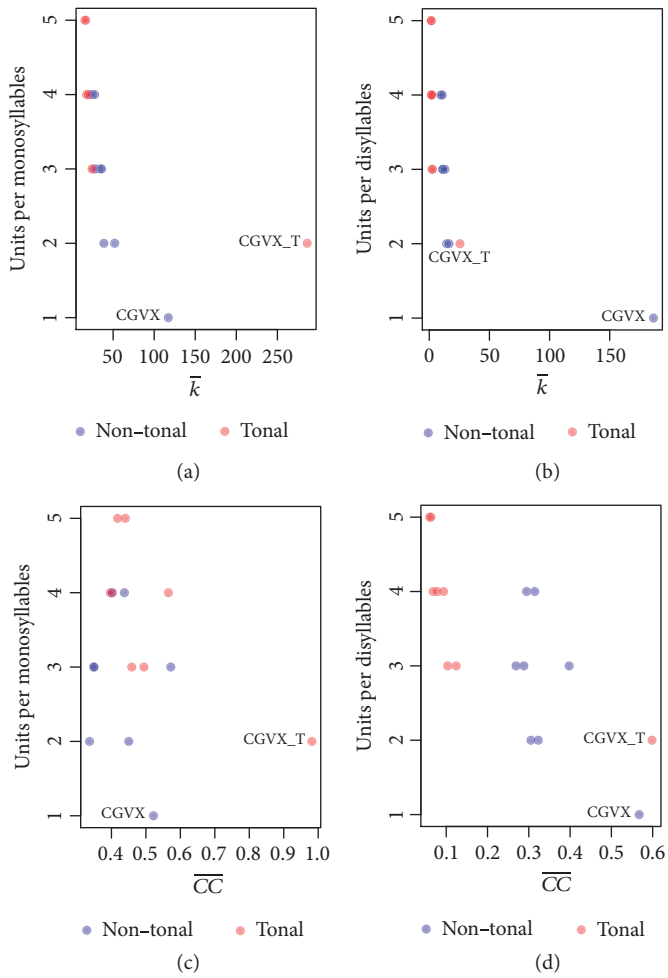

Figure 2: Units (the number of units within each PSNs' maximal segmentation schema) plotted against $\overline{k}$ (mean degree) for monosyllables (a) and disyllables (b), and $\overline{CC}$ (mean clustering coefficient) for monosyllables (c) and disyllables (d).

both unsegmented PSNs, which fell within a range of 69-88%. Thus, using Size alone would predict the likely candidate as both tonal and segmented.

Phonological networks have shown between 0.191 and 0.383 in mean clustering coefficient ($\overline{CC}$). The tonal and nontonal segmented PSNs comprise one group falling in between 0.247 and 0.460. Using $\overline{CC}$ as an indicator would exclude the unsegmented PSNs that have higher values (CGVX: 0.578; CGVX_T: 0.628).

Despite the possibility of a phonological network falling within the negative range (disassortative) of M, phonological networks have been positive (assortative), falling between 0.556-0.762. Our PSNs were also assortative, but did not follow a specific trend. Nontonal PSNs were tightly grouped between 0.577-0.689, which patterned similarly with previous phonological networks. The split in distributions for tonal PSNs meant that the low group fell below (C_V_C_T: 0.454; CG_VX_T: 0.470) the expected range, while the high group far above (CG_VX_T: 0.918; C_G_V_X_T: 0.900; C_G_VX_T: 0.894; C_G_V_C_T: 0.891). Only two tonal networks were near or within the expected range (C_GVX_T: 0.538; CGVX_T: 0.733).

Phonological networks have shown values in $\overline{L}$ between 6.08 and 10.40. This range excludes the nontonal unsegmented PSN (CGVX: 2.79), and the tonal segmented PSNs, which had higher values falling between 12.12 and 17.72. The past networks however are near or within the range of the nontonal segmented PSNs (5.31-7.65) and the tonal unsegmented PSN (CGVX_T: 5.40).

Finally, while all PSNs met the conditions for small world networks, not all were suggestive of being scale-free networks. Neither segmentation nor tone accounted for why four nontonal networks (CG_VX, C_G_VX, C_G_V_C, C_G_V_X) and two tonal networks (C_GVX_T, C_V_C_T) did not have power-law degree distributions.

No single PSN patterned according to past phonological networks. However, discounting Size, three nontonal segmented PSNs, C_GVX, C_V_C,and CG_V_X, meet the remaining criteria. In the next section, we evaluate the reaction times from Experiment 2 with the goal of identifying which of the sixteen PSNs was the likely candidate.

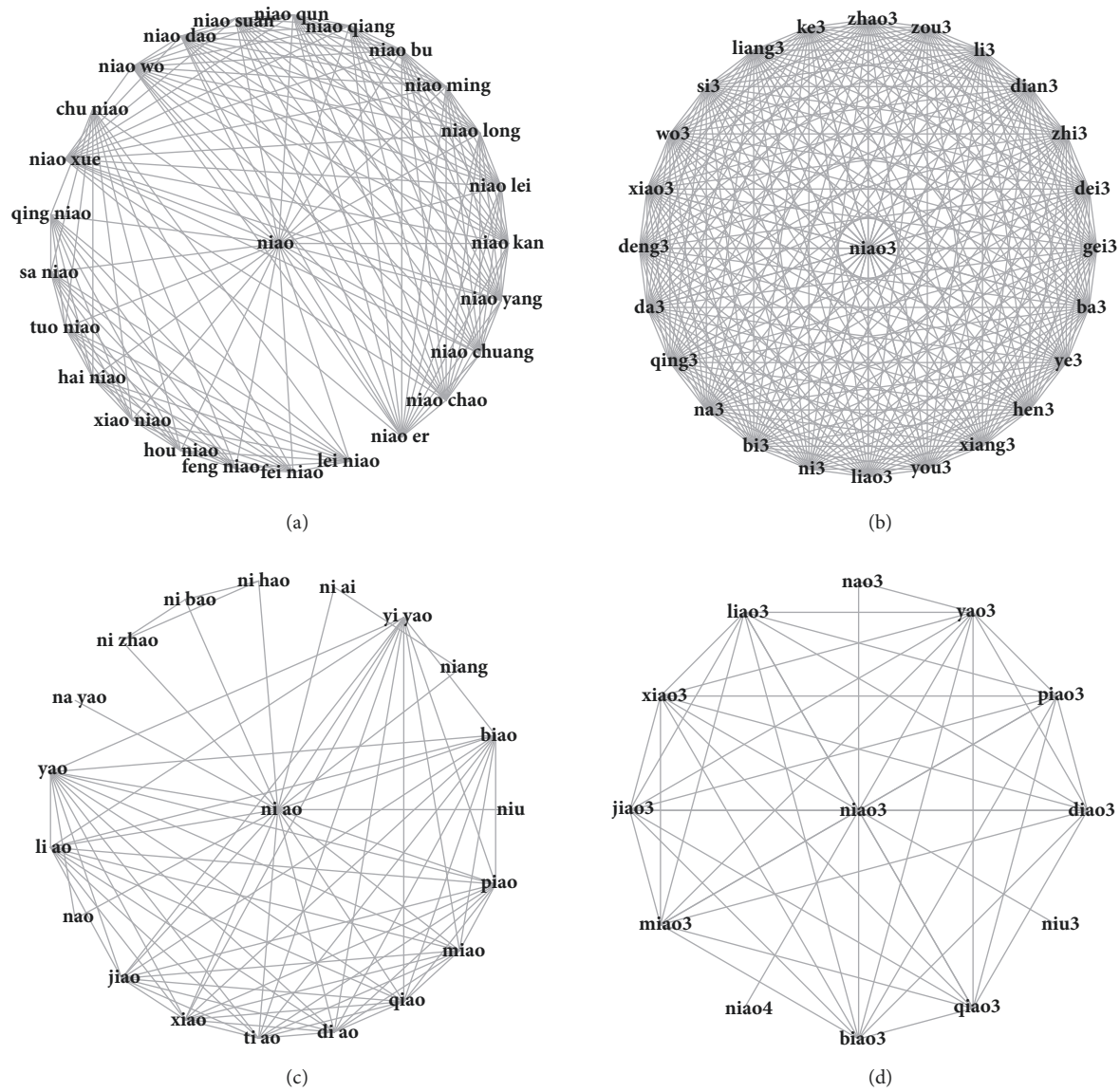

FIGURE 3: Word-level phonological networks for the monosyllabic word *niao3* /niau^{214}/, according to (a) the nontonal unsegmented PSN (CGVX), (b) the tonal unsegmented PSN (CGVX_T), (c) the nontonal fully segmented PSN (C_G_V_X), and (d) the tonal fully segmented PSN (C_G_V_X_T). For visualization purposes, (b) was restricted to just 20 visible neighbors due to all CGVX_T values being well over 200.

5. Model Selection Procedure

The goal of the current methods was to identify an optimal PSN through the lexical statistics that were tied to them. From the outset, this implied the identification of an optimal model in comparison to many other models. We began with backwards selection to identify which of the participant-related and stimuli-related predictors merited inclusion in the random effects structure. The purpose of having a complex random effects structure was not to increase generalizability of a confirmatory analysis, as proposed by Barr et al. [102], but instead to both restrict the current exploratory analysis from overestimating the effects of our network predictors and to guide future confirmatory analyses in dealing with participant- and stimuli-related characteristics. As is a current norm in the psycholinguistic literature, Subject and Item were included as random intercepts in all models featured. Upon identifying a random effects structure, sixteen full models were assessed according to R^2 using the Kenward-Roger approximation [103]. We used the "r2glmm" package in R [104] to (1) measure both marginal R^2 for full models, and semipartial R^2 for each fixed effect, and (2) to perform an R^2 difference test between our top ranked models.

Reaction times were measured offline using SayWhen [105]. One participant was excluded due to mean reaction times greater than 2.5 standard deviations above the group mean. Outliers with reaction times greater than 3000ms and lower than 415ms were then excluded, followed by three stimuli (*dia3*, *fo2*, *gun3*) with error rates greater than a third of the number of participants. From the remaining 6,240 trials, 31 false starts, 391 nonitems, 187 identical items, 24 missing,

Table 7: Experiment 2 lexical statistics.

	HD M(SD)	*k* M(SD)	*CC* M(SD)	Freq M(SD)	NF M(SD)
CGVX	18.91	119.86	0.53	4.15	2.47
	(14.76)	(85.67)	(0.11)	(0.89)	(0.47)
C_GVX	18.91	51.18	0.34	4.15	4.91
	(14.76)	(13.66)	(0.1)	(0.89)	(0.43)
CG_VX	18.93	37.13	0.46	4.15	4.8
	(14.76)	(11.23)	(0.16)	(0.89)	(0.34)
C_V_C	18.91	33.48	0.34	4.15	4.91
	(14.76)	(9.67)	(0.1)	(0.89)	(0.5)
CG_V_X	18.93	28.13	0.57	4.15	4.85
	(14.76)	(9.58)	(0.19)	(0.89)	(0.4)
C_G_VX	19.02	34.67	0.33	4.15	4.85
	(14.77)	(9.34)	(0.08)	(0.89)	(0.41)
C_G_V_C	19.02	26.41	0.37	4.15	4.89
	(14.77)	(7.24)	(0.12)	(0.89)	(0.45)
C_G_V_X	19.02	23.71	0.4	4.15	4.9
	(14.77)	(6.82)	(0.14)	(0.89)	(0.45)
CGVX_T	5.58	285.87	0.99	3.71	4.18
	(4.59)	(34.04)	(0.01)	(0.99)	(0.18)
C_GVX_T	5.58	24.98	0.49	3.71	4.22
	(4.59)	(8.19)	(0.12)	(0.99)	(0.47)
CG_VX_T	5.58	22.86	0.51	3.71	4.17
	(4.59)	(8.13)	(0.14)	(0.99)	(0.49)
C_V_C_T	5.58	16.45	0.39	3.71	4.24
	(4.59)	(6.35)	(0.12)	(0.99)	(0.53)
CG_V_X_T	5.58	18.82	0.57	3.71	4.22
	(4.59)	(7.74)	(0.19)	(0.99)	(0.56)
C_G_VX_T	5.58	18.16	0.39	3.71	4.22
	(4.59)	(7.5)	(0.11)	(0.99)	(0.5)
C_G_V_C_T	5.58	15.26	0.39	3.71	4.27
	(4.59)	(6.06)	(0.14)	(0.99)	(0.54)
C_G_V_X_T	5.58	14.19	0.41	3.71	4.28
	(4.59)	(5.81)	(0.16)	(0.99)	(0.55)

and 1 semantically related item were excluded, giving us a mean of 1530ms (SD: 557ms).

After exclusion, participants' responses consisted of edit distances between 1-5: Edit 1, 3532 observations (M: 1496ms; SD: 556ms); Edit 2, 934 observations (M: 1617ms; SD: 550ms); Edit 3, 245 observations (M: 1642ms; SD: 553ms); Edit 4, 49 observations (M: 1766ms; SD; 552ms); Edit 5, 6 observations (M: 1756ms; SD: 677).

Age, sex, self-rated spoken English, and whether a speaker was from a traditionally Guanhua speaking region were all nonsignificant, as were segment length (SegLen 1: 5, SegLen 2: 47; SegLen 3: 98; SegLen 4: 45) and lexical tone (tone 1: 49; tone 2: 47; tone 3: 43; tone 4: 56). The number of Chinese languages/dialects spoken by our participants (Num_Chinese) did significantly account for a portion of the variance. Our preliminary models revealed that higher values of Num_Chinese led to slower reaction times. Due to this variable representing variation at the participant level, it was added to the random effects structure as a random slope of Subject.

The fixed effects under consideration include Edit and five variables that vary due to PSN construction: homophone density (HD), lexical frequency (Freq), neighborhood frequency (NF), word-level degree (*k*), and word-level clustering coefficient (*CC*). All mean and standard deviations for the 80 network predictors (16 PSNs $*$ 5 network predictors) can be found in Table 7. Edit was not centered due to it being an interval measurement of only 5 levels, while the variables representing the PSNs (HD, *k*, *CC*, Freq, NF) were centered. Model selection output can be seen in Table 8.

The results identify the tonal complex-vowel segmented PSN (C_V_C_T) as the optimal model with a marginal R^2 of 0.162. The second highest ranking models belonged to two nontonal PSNs (C_V_C, C_G_VX) with marginal R^2 values of 0.154. An R^2 difference test showed that the C_V_C_T PSN was significantly higher than both competitors ($p < 0.001$).

Table 8: Model selection output for Experiment 2 according to marginal R^2 values for each model, and semipartial R^2 values for each fixed effect.

	CGVX	C_GVX	CG_VX	C_V_C	CG_V_X	C_G_VX	C_G_V_C	C_G_V_X
Model	0.133	0.135	0.114	0.154	0.108	0.154	0.124	0.121
Edit	0.004	0.005	0.005	0.004	0.004	0.005	0.005	0.004
Freq	0.002	0.018	0.028	0.021	0.026	0.038	0.033	0.032
HD	0.023	0.002	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
NF	0.001	0.005	< 0.001	0.004	< 0.001	0.008	0.001	< 0.001
k	0.043	0.004	0.010	0.003	0.002	0.017	0.003	0.001
CC	< 0.001	0.030	0.001	0.055	< 0.001	0.054	0.022	0.020
	+T	+T	+T	+T	+T	+T	+T	+T
Model	0.140	0.126	0.128	0.162	0.117	0.133	0.118	0.134
Edit	0.004	0.005	0.005	0.005	0.004	0.005	0.005	0.005
Freq	0.038	0.036	0.031	0.047	0.031	0.040	0.041	0.041
HD	0.008	< 0.001	0.006	0.005	0.005	0.005	0.005	0.007
NF	0.001	0.006	0.002	0.003	0.001	0.002	< 0.001	0.003
k	0.002	0.008	0.001	< 0.001	0.001	0.002	0.001	< 0.001
CC	0.021	0.003	0.022	0.060	0.012	0.025	0.009	0.011

Table 8 revealed that Freq according to tonal PSNs accounted for a greater portion of the variance than those of nontonal PSNs. NF played a limited role across all of the PSNs, while HD accounted for a portion of the variance in unsegmented and tonal PSNs (excluding C_GVX). Finally, despite *k* accounting for a portion of the variance for four of the PSNs (CGVX, CG_VX, C_G_VX, C_GVX_T), *CC* outranked *k* in semipartial R^2 for twelve PSNs.

The model estimates for the C_V_C_T PSN model, shown in Table 9, reveal that monosyllabic words greater in *CC* inhibited mental search and the production of phonological neighbors. Both high Freq and low Edit sped the search for neighbors. Tensor product smooths within a contour graph [106], as seen in Figure 4, were used to visualize a significant interaction between *CC*.C_V_C_T and Edit (adjusted R-sq. = 0.002; F = 11.85; $p < 0.001$). The graph reveals that contrary to the facilitative effect of low Edit, when the stimuli were low in *CC*, low Edit responses tended to be produced slower than high Edit responses.

6. Discussion

The model selection procedure used the lexical statistics tied to each of the sixteen PSNs to identify the likely structure used during mental search of phonological neighbors. Based on previous findings from the phonological association task of Wiener and Turnbull [70], we predicted a facilitative effect to high *k*. We also predicted the identification of an unsegmented PSN (CGVX, CGVX_T) based on the findings of production studies that hold that syllables are the first units for retrieval in Mandarin, i.e., "proximate units". Contrary to our predictions, model selection identified the tonal complex-vowel segmented PSN (C_V_C_T) without the expected facilitative effect of *k*. Interestingly, C_V_C_T was the same segmentation schema used to define phonological similarity in the Wiener and Turnbull study [70]. Meanwhile, the principle predictor within the C_V_C_T model, with a semipartial R^2 of 0.060, belonged to *CC* and was inhibitory in its effect on mental search.

The literature related to *CC* in the English mental lexicon entails inhibited retrieval and lower accuracy to high *CC* words. High *CC* has been tied to greater speech errors (Chan & Vitevitch 2010), lower accuracy in a perceptual identification task (Chan & Vitevitch, 2009), and the retention of newly learned nonwords (Goldstein & Vitevitch, 2014). Directly relevant to the current evidence is that high *CC* has also been shown to slow the retrieval of picture names (Chan & Vitevitch 2010), the judgment of lexical status of auditory words (Chan & Vitevitch, 2009; Goldstein & Vitevitch 2017) and visually presented orthographic words (Siew, 2018). The previous inhibitory *CC* findings are suggestive that our reaction times represent the selection of the target lexical item prior to production.

There are several indications as to why our results point to the C_V_C_T PSN. The first indication is the use of vowel information during the task. For example, glides, which are collapsed in this schema, were the least manipulated units in both experiments (Experiment 1: 2%; Experiment 2: 1%). A second indication is the length of our stimuli. Of the 198 stimuli in Experiment 2, nearly half consisted of three segments (SegLen 3 = 98). Through the disregard of the medial glides, which are obligatory in four-segment items, the 45 four-segment items were likely treated in the same manner as their three-segment counterparts. Given that 66% of all manipulations were of the substitution edit type, three-segment stimuli were primarily manipulated into three-segment responses. The final indication can be found in our participants' bias in producing tone neighbors (Experiment 1: 34%; Experiment 2: 46%). The influence of lexical tone was especially noted in the significant Freq effect across all tonal PSNs.

Of final concern is the significant effect of Edit. Participant-produced phonological neighbors that shared greater phonological similarity with the stimuli (i.e., lower

Table 9: Model estimates for Experiment 2.

Random effects		Variance	SD	Corr		
Item		0.001	0.035			
Subject		0.075	0.274			
Num_Chinese:Subject		0.003	0.055	0.590		
Residual		0.181	0.425			
Fixed effects	Estimate	SE	df	t value	p value	R^2
Intercept	1.536	0.066	32.15	23.22	< 0.001	
Edit	0.049	0.011	4713.00	4.67	< 0.001	0.005
Freq.C_V_C_T	-0.021	0.007	189.50	-3.08	0.002	0.047
HD.C_V_C_T	0.007	0.007	177.60	0.95	0.345	0.005
NE.C_V_C_T	0.006	0.007	183.30	0.78	0.438	0.003
k.C_V_C_T	0.001	0.008	191.30	0.14	0.891	< 0.001
CC.C_V_C_T	0.027	0.007	212.60	3.71	< 0.001	0.060

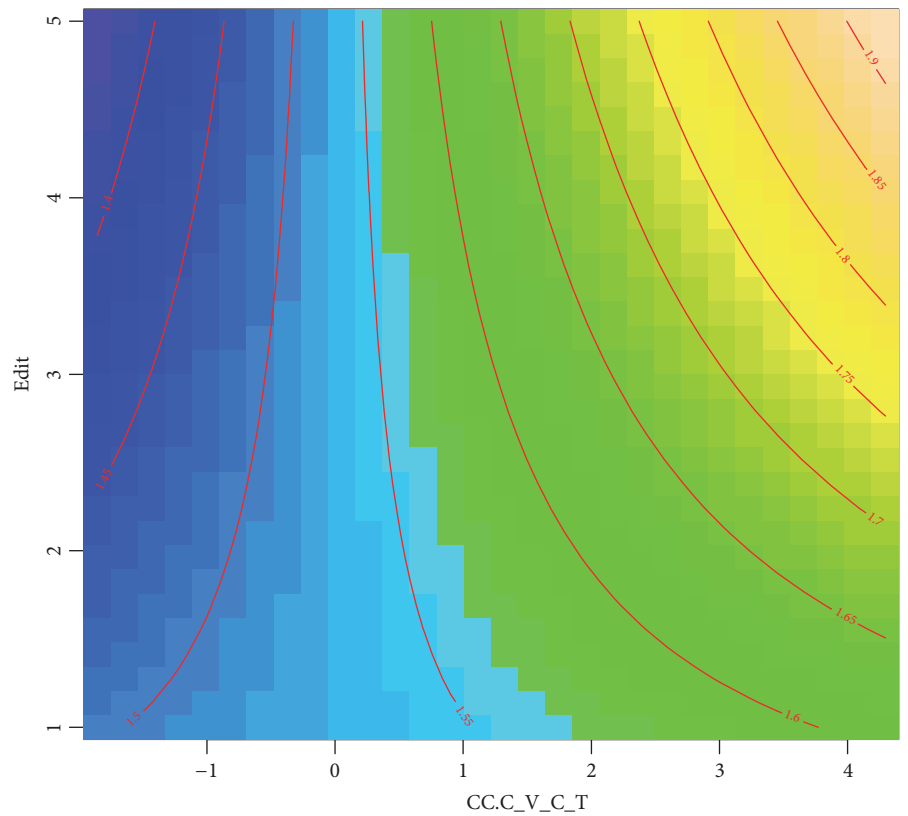

Figure 4: Tensor product smooth for the interaction between word-level clustering coefficient according to the tonal complex-vowel segmented PSN (*CC*.C_V_C_T) and the number of units (segments and/or tone) that differed between auditory stimuli and participant-produced phonological neighbors (Edit). Reaction times, as noted in the red contour lines, blend from cool (shorter latencies) to warm colors (longer latencies).

Edit) were produced faster than low similarity responses. These findings address a question posed by Vitevitch and colleagues [73] as to whether the edit distance between the stimuli and participant-produced phonological neighbors affects the time it takes to generate a phonological neighbor. In their study, they used neighbor generation as a means to investigate the types of neighbors that would occur to a given target if the target were incorrectly perceived. According to their hypotheses, our current result is suggestive that less time is needed to recover from the misperception of a spoken word when the misperceived item shares greater phonological similarity with its intended target.

7. Conclusion

In this study we constructed, measured, and then identified a possible schematic representation of phonological processing in the tonal language Mandarin Chinese. We began with

the identification of an optimal syllable inventory through 2 phonological association tasks. In Experiment 1, we used the edit distance between our participants' spoken responses and the annotation of the stimuli according to each syllable inventory to build and validate, (in Experiment 2), a novel syllable inventory that outperformed both prior inventories. On the premise of the nonuniqueness theory [84], the N&H inventory, built on phonological similarity, is the optimal choice to model Mandarin vocabulary in a phonological network in which relations between lexical items depend on phonological similarity.

The phonological association tasks aided in the identification of segmentation biases through spoken productions of phonological neighbors. Both tasks showed a strong tendency to use replacement as the method of manipulating units. The most commonly manipulated units were the items' lexical tones. In contrast, the most often ignored segments were medial glides.

The novel syllable inventory was used to build networks that we titled phonological segmentation neighborhoods (PSNs), in which schematic representations of segmentation determined phonological similarity. Each PSN was defined by it being built from one of sixteen phonological segmentation schemas. In using the same lexicon and number of nodes (30,000) within each PSN we were able to analyze the effects of segmentation and lexical tone on network statistics, both at the topological level and among monosyllables and disyllables. Segmented PSNs showed gradient differences according to the number of units within the syllable or whether or not they featured tone as a unit. For example, PSNs of less segmentation had greater $\overline{k}$ for both monosyllables and disyllables. $\overline{CC}$ did not show this pattern for monosyllables (except for the nontonal unsegmented PSN (CGVX_T), but did for disyllables, which is contrary to prior network findings [3, 5]. For nontonal segmented PSNs, the lack of tone also led to a greater $\overline{k}$ due to the mixing of syllable length.

The similarities between the sixteen PSNs and previous phonological networks were found in the presence of assortative mixing by degree and small world characteristics. The sixteen PSNs varied in Size, $\overline{k}$, $\overline{CC}$, $\overline{L}$, and whether or not they had power-law degree distributions. Discounting Size, three nontonal segmented PSNs, C_GVX, C_V_C, and CG_V_X, met all of the characteristics of the previously analyzed phonological networks. Contrary to our initial predictions, and those informed by the network analysis, our reaction time analysis revealed the tonal complex-vowel segmented PSN (C_V_C_T), with a significant inhibitory *CC* effect and a facilitative effect of low edit distance and high lexical frequency.

The current study began under the premise that the one-size fits all approach taken to phonological networks might not be sufficient. Yet, given the results of Experiment 2, do we have evidence to support this contention? The identification of the C_V_C_T PSN is likely the result of the stimuli we presented to the participants (majority 3 segments in length), and our participants navigation through the task demands, in that (1) the collapsing of vowel information in this PSN mirrors the lack of medial glide manipulations found in our participants' responses, (2) the primary method of substitution in order to produce a phonological neighbor meant that most productions were 3-segment neighbors of 3-segment stimuli, and (3) the presence of lexical tone in the featured PSN is likely the result of our participants' bias to use lexical tone as a guide through the mental lexicon.

The results are suggestive of complex adaptation wherein through manipulating the content and demand during a given task we identified the objects of those mental transformations. Thus far, network science methods have assisted both in the formation of the questions and how the results have been interpreted, despite the fact that the influence of topological features is still unclear. Future work will need to explore whether changes in the stimuli and task demands lead to the identification of different PSNs and whether those changes are meaningful representations of lexical processing.

Data Availability

The experiment data used to support the findings of this study have been deposited in github.com (https://github.com/karlneergaard/Constructing_the_Mandarin_phonological_network). The lexical database from which the network statistics were calculated has been deposited in github.com (https://github.com/karlneergaard/Database_of_word-level_statistics).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank the creators of the SUBTLEX-CH database, who provided the corpora used in the present study. We would like to thank Hongzhi Xu for his part in the creation of the lexical database and both Stephen Politzer-Ahles and Michael Tyler for their advice on the manuscript. The funding for this study was made available to the first author through a Hong Kong Polytechnic University (PolyU) International Postgraduate Scholarship and through the PolyU Faculty of Humanities International Collaboration project: 1-ZVKX Conversational Brains: A Multidisciplinary Approach.

References

[1] K. Y. Chan and M. S. Vitevitch, "The influence of the phonological neighborhood clustering coefficient on spoken word recognition," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 35, no. 6, pp. 1934–1949, 2009.

[2] M. S. Vitevitch, K. Y. Chan, and R. Goldstein, "Insights into failed lexical retrieval from network science," *Cognitive Psychology*, vol. 68, no. 1, pp. 1–32, 2014.

[3] K. Y. Chan and M. S. Vitevitch, "Network structure influences speech production," *Cognitive Science*, vol. 34, no. 4, pp. 685–697, 2010.

[4] R. Goldstein and M. S. Vitevitch, "The influence of clustering coefficient on word-learning: how groups of similar sounding words facilitate acquisition," *Frontiers in Psychology*, vol. 5, 2014.

[5] M. S. Vitevitch, K. Y. Chan, and S. Roodenrys, "Complex network structure influences processing in long-term and short-term memory," *Journal of Memory and Language*, vol. 67, no. 1, pp. 30–44, 2012.

[6] M. S. Vitevitch and N. Castro, "Using network science in the language sciences and clinic," *International Journal of Speech-Language Pathology*, vol. 17, no. 1, pp. 13–25, 2015.

[7] N. Castro, K. M. Pelczarski, and M. S. Vitevitch, "Using network science measures to predict the lexical decision performance of adults who stutter," *Journal of Speech, Language, and Hearing Research*, vol. 60, no. 7, pp. 1–8, 2017.

[8] C. S. Q. Siew, K. M. Pelczarski, J. S. Yaruss, and M. S. Vitevitch, "Using the OASES-A to illustrate how network analysis can be applied to understand the experience of stuttering," *Journal of Communication Disorders*, vol. 65, pp. 1–9, 2017.

[9] M. Stella and M. Brede, "Patterns in the English language: phonological networks, percolation and assembly models," *Journal of Statistical Mechanics: Theory and Experiment*, no. 5, 2015.

[10] M. Stella, N. M. Beckage, M. Brede, and M. De Domenico, "Multiplex model of mental lexicon reveals explosive learning in humans," *Scientific Reports*, vol. 8, no. 1, 2018.

[11] T. M. Gruenenfelder and D. B. Pisoni, "The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons," *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 3, pp. 596–609, 2009.

[12] R. Turnbull and S. Peperkamp, "What governs a language's lexicon? Determining the organizing principles of phonological neighbourhood networks," in *Proceedings of the 5th International Workshop on Complex Networks and their Applications (Complex Networks 2016)*, pp. 83–94, 2016.

[13] M. S. Vitevitch, "What can graph theory tell us about word learning and lexical retrieval?" *Journal of Speech, Language, and Hearing Research*, vol. 51, no. 2, pp. 408–422, 2008.

[14] S. Arbesman, S. H. Strogatz, and M. S. Vitevitch, "Comparative analysis of networks of phonologically similar words in english and spanish," *Entropy*, vol. 12, no. 3, pp. 327–337, 2010.

[15] S. Arbesman, S. H. Strogatz, and M. S. Vitevitch, "The structure of phonological networks across multiple languages," *International Journal of Bifurcation and Chaos*, vol. 20, no. 3, pp. 679–685, 2010.

[16] W.-S. Lee, "A phonetic study of the "er-hua" rimes in Beijing Mandarin," in *Proceedings of the 9th European Conference on Speech Communication and Technology*, pp. 1093–1096, Portugal, September 2005.

[17] M.-F. Li, W.-C. Lin, T.-L. Chou, F.-L. Yang, and J.-T. Wu, "The role of orthographic neighborhood size effects in chinese word recognition," *Journal of Psycholinguistic Research*, vol. 44, no. 3, article no. 2, pp. 219–236, 2015.

[18] J.-T. Wu, F.-L. Yang, and W.-C. Jin, "Beyond phonology matters in character recognition," *Chinese Journal of Psychology*, vol. 55, no. 3, pp. 289–318, 2013.

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

[20] W. Wen, "A study of Chinese homophones from the view of English homographs," *Modernizing our Language*, vol. 2, pp. 120–124, 1980.

[21] W.-F. Chen, P.-C. Chao, Y.-N. Chang, C.-H. Hsu, and C.-Y. Lee, "Effects of orthographic consistency and homophone density on Chinese spoken word recognition," *Brain and Language*, vol. 157-158, pp. 51–62, 2016.

[22] S. Xu, *Phonology of Standard Chinese*, Wenzi Gaige Chubans, Beijing, China, 1980.

[23] R. L. Cheng, "Mandarin phonological structure," *Journal of Linguistics*, vol. 2, no. 2, pp. 135–158, 1966.

[24] R. You, N. Qian, and Z. Gao, "On the phonemic system of standard chinese," *Zhongguo Yuwen*, vol. 5, no. 158, pp. 328–334, 1980.

[25] Z. M. Bao, "Fan-Qie languages and reduplication," *Linguistic Inquiry*, vol. 21, 1990.

[26] B. X. Ao, "The non-uniqueness condition and the segmentation of the Chinese syllable," *Working Papers in Linguistics*, vol. 42, pp. 1–25, 1992.

[27] S. Duanmu, *The Phonology of Standard Chinese*, Oxford University Press, Oxford, UK, 2nd edition, 2007.

[28] W. Lee and E. Zee, "Standard Chinese (Beijing)," *Journal of the International Phonetic Association*, vol. 33, no. 1, pp. 109–112, 2003.

[29] C. Shih and B. Ao, "Duration study for the bell laboratories mandarin test-to-speech system," in *Progress in Speech Synthesis*, J. van Santen, R. Sproat, J. Olive, and J. Hirschberg, Eds., pp. 383–399, Springer-Verlag, New York, NY, USA, 1997.

[30] D. Wu, *Cross-Regional Word Duration Patterns in Mandarin*, University of Illinois at Urbana-Champaign, 2017.

[31] F. Wu and M. Kenstowicz, "Duration reflexes of syllable structure in Mandarin," *Lingua*, vol. 164, pp. 87–99, 2015.

[32] Y. R. Chao, *Eight Varieties of Secret Language Based on the Principle of Fan-qie*, 1931.

[33] M. Yip, "Reduplication and C-V skeleta in chinese secret language," *Linguistic Inquiry*, vol. 13, no. 4, pp. 637–660, 1982.

[34] P. G. O'Séaghdha and J.-Y. Chen, "Toward a language-general account of word production: the proximate units principle," in *Proceedings of the CogSci... Annual Conference of the Cognitive Science Society*, pp. 68–73, July-August, 2009.

[35] P. G. O'Seaghdha, J.-Y. Chen, and T.-M. Chen, "Proximate units in word production: phonological encoding begins with syllables in mandarin chinese but with segments in english," *Cognition*, vol. 115, no. 2, pp. 282–302, 2010.

[36] V. A. Fromkin, "The non-anomalous nature of anomalous utterances," *Language*, vol. 47, no. 1, p. 27, 1971.

[37] S. Shattuck-Hufnagel, "Speech errors as evidence for a serial-ordering mechanism in sentence production," in *Sentence processing: Psycholinguistic Studies Presented to Merrill Garrett*, W. E. Cooper and E. C. T. Walker, Eds., pp. 295–342, Erlbaum, Hillsdale, NJ, USA, 1979.

[38] J.-Y. Chen, "A small corpus of speech errors in mandarin chinese and their classification," *Word Chinese Language*, vol. 69, pp. 26–41, 1993.

[39] J.-Y. Chen, "The representation and processing of tone in mandarin chinese: evidence from slips of the tongue," *Applied Psycholinguistics*, vol. 20, no. 2, pp. 289–301, 1999.

[40] J.-Y. Chen, "Syllable errors from naturalistic slips of the tongue in Mandarin Chinese," *Psychologia: an International Journal of Psychology in the Orient*, 2000.

[41] A. S. Meyer, "The time course of phonological encoding in language production: phonological encoding inside a syllable," *Journal of Memory and Language*, vol. 30, no. 1, pp. 69–89, 1991.

[42] N. O. Schiller, "The effect of visually masked syllable primes on the naming latencies of words and pictures," *Journal of Memory and Language*, vol. 39, no. 3, pp. 484–507, 1998.

[43] N. O. Schiller, "Masked syllable priming of English nouns," *Brain and Language*, vol. 68, no. 1-2, pp. 300–305, 1999.

[44] N. O. Schiller, "Single word production in english: the role of subsyllabic units during phonological encoding," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 26, no. 2, pp. 512–528, 2000.

[45] J. D. Jescheniak and H. Schriefers, "Priming effects from phonologically related distractors in picture–word interference," *The Quarterly Journal of Experimental Psychology*, vol. 54, no. 2, pp. 371–382, 2001.

[46] A. S. Meyer and H. Schriefers, "Phonological facilitation in picture-word interference experiments: effects of stimulus onset asynchrony and types of interfering stimuli," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 17, no. 6, pp. 1146–1160, 1991.

[47] J.-Y. Chen, T.-M. Chen, and G. S. Dell, "Word-form encoding in mandarin chinese as assessed by the implicit priming task," *Journal of Memory and Language*, vol. 46, no. 4, pp. 751–781, 2002.

[48] R. G. Verdonschot, M. Nakayama, Q. Zhang, K. Tamaoka, and N. O. Schiller, "The proximate phonological unit of chinese-english bilinguals: proficiency matters," *Plos One*, vol. 8, no. 4, 2013.

[49] W. You, Q. Zhang, and R. G. Verdonschot, "Masked syllable priming effects in word and picture naming in chinese," *Plos One*, vol. 7, no. 10, 2012.

[50] J.-Y. Chen, W.-C. Lin, and L. Ferrand, "Masked priming of the syllable in mandarin chinese speech production," *Chinese Journal of Psychology*, vol. 45, no. 1, pp. 107–120, 2003.

[51] T.-M. Chen and J.-Y. Chen, "The syllable as the proximate unit in mandarin chinese word production: an intrinsic or accidental property of the production system?" *Psychonomic Bulletin & Review*, vol. 20, no. 1, pp. 154–162, 2013.

[52] A. W.-K. Wong and H.-C. Chen, "Processing segmental and prosodic information in Cantonese word production," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 34, no. 5, pp. 1172–1190, 2008.

[53] A. W.-K. Wong and H.-C. Chen, "What are effective phonological units in Cantonese spoken word planning?" *Psychonomic Bulletin & Review*, vol. 16, no. 5, pp. 888–892, 2009.

[54] A. W. Wong, J. Huang, H. Chen, and K. Paterson, "Phonological units in spoken word production: insights from cantonese," *Plos One*, vol. 7, no. 11, Article ID e48776, 2012.

[55] Y. Kureta, T. Fushimi, and I. F. Tatsumi, "The functional unit in phonological encoding: evidence for moraic representation in native japanese speakers," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 32, no. 5, pp. 1102–1119, 2006.

[56] Y. Kureta, T. Fushimi, N. Sakuma, and I. F. Tatsumi, "Orthographic influences on the word-onset phoneme preparation effect in native japanese speakers: evidence from the word-form preparation paradigm," *Japanese Psychological Research*, vol. 57, no. 1, pp. 50–60, 2015.

[57] K. Tamaoka and S. Makioka, "Japanese mental syllabary and effects of mora, syllable, bi-mora and word frequencies on japanese speech production," *Language and Speech*, vol. 52, no. 1, pp. 79–112, 2009.

[58] R. G. Verdonschot, S. Kiyama, K. Tamaoka, S. Kinoshita, W. La Heij, and N. O. Schiller, "The functional unit of japanese word naming: evidence from masked priming," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 37, no. 6, pp. 1458–1473, 2011.

[59] J. G. Malins and M. F. Joanisse, "Setting the tone: an ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese," *Neuropsychologia*, vol. 50, no. 8, pp. 2032–2043, 2012.

[60] J. G. Malins, D. Gao, R. Tao et al., "Developmental differences in the influence of phonological similarity on spoken word processing in Mandarin Chinese," *Brain and Language*, vol. 138, pp. 38–50, 2014.

[61] J. Zhao, J. Guo, F. Zhou, and H. Shu, "Time course of Chinese monosyllabic spoken word recognition: evidence from ERP analyses," *Neuropsychologia*, vol. 49, no. 7, pp. 1761–1770, 2011.

[62] J. G. Malins and M. F. Joanisse, "The roles of tonal and segmental information in Mandarin spoken word recognition: an eyetracking study," *Journal of Memory and Language*, vol. 62, no. 4, pp. 407–420, 2010.

[63] A. S. Desroches, R. L. Newman, and M. F. Joanisse, "Investigating the time course of spoken word recognition: electrophysiological evidence for the influences of phonological similarity," *Cognitive Neuroscience*, vol. 21, no. 10, pp. 1893–1906, 2009.

[64] S. Duanmu, "Chinese syllable structure," in *The Blackwell Companion to Phonology*, pp. 2151–2777, 2010.

[65] Y. H. Lin, *The Sounds of Chinese with Audio CD*, Cambridge University Press, 2007.

[66] X. Zhao and P. Li, "An online database of phonological representations for Mandarin Chinese," *Behavior Research Methods*, vol. 41, no. 2, pp. 575–583, 2009.

[67] A. Cutler, N. Sebastian-Galles, O. Soler-Vilageliu, and B. Van Ooijen, "Constraints of vowels and consonants on lexical selection: cross-linguistic comparisons," *Memory & Cognition*, vol. 28, no. 5, pp. 746–755, 2000.

[68] A. E. Marks, D. R. Moates, Z. S. Bond, and V. Stockmal, "Word reconstruction and consonant features in english and spanish," *Linguistics*, vol. 40, no. 2, pp. 421–438, 2002.

[69] B. Van Ooijen, "Vowel mutability and lexical selection in english: evidence from a word reconstruction task," *Memory & Cognition*, vol. 24, no. 5, pp. 573–583, 1996.

[70] S. Wiener and R. Turnbull, "Constraints of tones, vowels and consonants on lexical selection in mandarin chinese," *Language and Speech*, vol. 59, no. 1, pp. 1–24, 2015.

[71] P. A. Luce and N. R. Large, "Phonotactics, density, and entropy in spoken word recognition," *Language and Cognitive Processes*, vol. 16, no. 5-6, pp. 565–581, 2001.

[72] M. Muneaux and J. C. Ziegler, "Locus of orthographic effects in spoken word recognition: novel insights from the neighbour generation task," *Language and Cognitive Processes*, vol. 19, no. 5, pp. 641–660, 2004.

[73] M. S. Vitevitch, R. Goldstein, and E. Johnson, "Path-length and the misperception of speech: insights from network science and psycholinguistics," in *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, A. Mehler, A. Lücking, S. Banisch et al., Eds., pp. 29–45, Springer, Berlin Heidelberg, Germany, 2016.

[74] M. Swadesh, "The phonemic principle," *Language (Baltim)*, vol. 10, no. 2, pp. 117–129, 1934.

[75] K. L. Pike, "Grammatical prerequisites to phonemic analysis," *Word*, vol. 3, pp. 155–172, 1947.

[76] W. J. M. Levelt, A. Roelofs, and A. S. Meyer, "A theory of lexical access in speech production," *Behavioral and Brain Sciences*, vol. 22, no. 1, pp. 1–38, 1999.

[77] G. S. Dell, "A spreading-activation theory of retrieval in sentence production," *Psychological Review*, vol. 93, no. 3, pp. 283–321, 1986.

[78] *I. Psychological Software Tools, "E-Prime 2.0."*, Psychological Software Tools, Inc., Pitsburgh, PA, USA, 2012.

[79] "汉典 zdic.net," http://www.zdic.net/.

[80] B. Li, *Research on Chinese Word Segmentation and Proposals for Improvement*, Roskilde University, 2011.

[81] X. Li, C. Zong, and K. Su, "A unified model for solving the OOV problem of chinese word segmentation," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 14, no. 3, pp. 12–29, 2015.

[82] Y. Zhang, J. Niehues, and A. Waibel, "Integrating encyclopedic knowledge into neural language models," in *Proceedings of the 13th International Workshop on Spoken Language Translation (IWSLT)*, 2016.

[83] J. Ma, C. Kit, and D. Gerdemann, "Semi-automatic annotation of chinese word structure and linguistics," in *Proceeding of the 2nd CIPS-SIGHAN Jt. Conference Chinese Lang. Process. (CIPS-SIGHAN 2012)*, vol. 0, pp. 9–17, 2012.

[84] Y.-R. Chao, "The non-uniqueness of phonemic solutions of phonetic systems," *Bulletin of the Institute of History and Philology Academia Sinica*, vol. 4, no. 4, pp. 363–398, 1934.

[85] Q. Cai and M. Brysbaert, "SUBTLEX-CH: chinese word and character frequencies based on film subtitles," *Plos One*, vol. 5, no. 6, Article ID e10729, 2010.

[86] M. Brysbaert and B. New, "Moving beyond kučera and francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english," *Behavior Research Methods*, vol. 41, no. 4, pp. 977–990, 2009.

[87] K. Neergaard, H. Xu, and C.-R. Huang, "Database of mandarin neighborhood statistics," in *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, pp. 2270–2277, May 2016.

[88] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *International Journal of Complex Systems*, vol. 1695, no. 5, pp. 1–9, 2006.

[89] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, Article ID 208701, pp. 1–5, 2002.

[90] M. E. J. Newman and J. Park, "Why social networks are different from other types of networks," *Physical Review E*, vol. 69, no. 3, pp. 1–9, 2003.

[91] J. M. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, no. 6798, p. 845, 2000.

[92] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality," *Physical Review A*, vol. 38, no. 1, pp. 364–374, 1988.

[93] A. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[94] A. L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A: Statistical Mechanics and its Applications*, vol. 272, no. 1, pp. 173–187, 1999.

[95] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.

[96] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *Society for Industrial and Applied Mathematics*, vol. 51, no. 4, pp. 661–703, 2009.

[97] C. S. Gillespie, "Fitting heavy tailed distributions: the poweRlaw package," *Journal of Statistical Software*, vol. 64, no. 2, pp. 1–16, 2015.

[98] P. Shoemark, S. Goldwater, J. Kirby, and R. Sarkar, "Towards robust cross-linguistic comparisons of phonological networks," in *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pp. 110–120, Berlin, Germany, August 2016.

[99] Y. H. Lin, *Autosegmental Treatment of Segmental Processes in Chinese Phonology*, University of Texas at Austin, 1989.

[100] I.-P. Wan, "A psycholinguistic study of postnuclear glides and coda nasals in mandarin," *Journal of Languages and Linguistics*, vol. 5, no. 2, pp. 158–176, 2006.

[101] M. E. Newman, "The structure of scientific collaboration networks," *Proceedings of the National Acadamy of Sciences of the United States of America*, vol. 98, no. 2, pp. 404–409, 2001.

[102] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013.

[103] B. C. Jaeger, L. J. Edwards, K. Das, and P. K. Sen, "An R^2 statistic for fixed effects in the generalized linear mixed model," *Journal of Applied Statistics*, vol. 44, no. 6, pp. 1086–1105, 2017.

[104] B. Jaeger, *r2glmm: Computes R squared for Mixed (Multilevel) Models*, 2017.

[105] P. A. Jansen and S. Waiter, "Say when: an automated method for high-accuracy speech onset detection," *Behavior Research Methods*, vol. 40, no. 3, pp. 744–751, 2008.

[106] S. N. Wood, F. Scheipl, and J. J. Faraway, "Straightforward intermediate rank tensor product smoothing in mixed models," *Statistics and Computing*, vol. 23, no. 3, pp. 341–360, 2013.