Exploring Orthographic Neighborhood Size Effects in a Computational Model of Chinese Character Naming

Ya-Ning Chang[ac], Stephen Welbourne[b], Chia-Ying Lee[c]

[a]Department of Psychology, Lancaster University, UK
[b]Neuroscience and Aphasia Research Unit (NARU), University of Manchester, UK
[c]Institute of Linguistics, Academia Sinica, Taiwan

Correspondence to:
Dr Ya-Ning Chang
yaningchang@gmail.com
Department of Psychology
Lancaster University
LA1 4YF
Tel: +44 (0) 1524 593698

Abstract

Orthographic neighborhood (N) size effects have been extensively studied in English consistently producing a facilitatory effect in word naming tasks. In contrast, several recent studies on Chinese character naming have demonstrated an inhibitory effect of neighborhood size. Response latencies tend to be inhibited by inconsistent characters with large neighborhoods relative to small neighborhoods. These differences in neighborhood effects between languages may depend on the characteristics (depth) of the mapping between orthography and phonology. To explore this, we first conducted a behavioral experiment to investigate the relationship between neighborhood size, consistency and reading response. The results showed an inhibitory effect of neighborhood size for inconsistent characters but a faciliatory effect for consistent characters. We then developed two computational models based on parallel distributed processing principles to try and capture the nature of the processing that leads to these results in Chinese character naming. Simulations using models based on the triangle model of reading indicated that consistency and neighborhood size interact with the division of labor between semantics and phonology to produce these effects.

Keywords: orthographic neighborhood size; phonetic radical neighbors; consistency; Chinese character naming; semantics; computational modeling;

Exploring orthographic neighborhood size effects
in a computational model of Chinese character naming

1. Introduction

Orthographic neighborhood size is one of the key lexical variables that affect word response latencies during visual word recognition. The most widely used measure of neighborhood size (denoted by the statistic of *N*) is defined as the number of words that could be created by changing one letter in a target word (Coltheart, Davelaar, Jonasson, & Besner, 1977). For example, *dog* has a number of orthographic neighbors such as *jog*, *dot*, *dig*, *log*, and *doe*. Many studies have explored the effects of orthographic neighborhood across a range of tasks including naming and lexical decision (Andrews, 1989, 1992; Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Carreiras, Perea, & Grainger, 1997; Coltheart et al., 1977; Forster & Shen, 1996; Grainger, 1990; Sears, Hino, & Lupker, 1995). While the focus of this paper is naming, it is useful to consider both lexical decision and naming data together as these can constrain the theoretical explanations of the neighborhood effect. In lexical decision, the findings concerning the neighborhood size effect appear to be somewhat mixed (Balota et al., 2004). Andrews (1989, 1992) reported a facilitatory effect of neighborhood size in the lexical decision task, in particular for low frequency words. However, other studies have found that response latencies for words having high frequency neighbors tend to be prolonged in comparison with words having low frequency neighbors (Carreiras et al., 1997; Grainger, 1990; Grainger, Oregan, Jacobs, & Segui, 1989). This has been referred to as the *neighborhood frequency effect*. Despite this, facilitation in lexical decision has been reported when both neighborhood size and neighborhood frequency are considered in the same experiment (Forster & Shen, 1996; Sears et al., 1995). Balota et al. (2004) examined the neighborhood size effect by conducting multiple regression analyses on a group of younger readers and another group of older readers in both naming and lexical decision tasks. They showed that younger readers' lexical decision performance was facilitated by neighborhood size particularly for low frequency words, which is consistent with Andrews (1989, 1992). However, in older and slower readers the lexical decision performance was inhibited by neighborhood size. These results suggest that the neighborhood size effect in lexical decision may depend on decision strategies and the processing speed of the subjects.

The effect of neighborhood size in naming tasks is much more consistent, showing a robust facilitatory effect across many studies (Andrews, 1989, 1992; Balota et al., 2004; Carreiras et al., 1997; Grainger, 1990; Peereman & Content, 1995, 1997; Sears et al., 1995), particularly when words are low in frequency. This has been supported by studies using either a factorial design (Andrews 1989, 1992; Sears et al., 1995) or a regression technique (Balota et al., 2004); and the effect also has been found in different alphabetic languages such as Dutch (Grainger, 1990), French (Peereman & Content, 1995) and Spanish (Carreiras et al., 1997).

One interpretation of the neighborhood size effect proposed by Andrews (1989) is based on the interactive activation theory of word recognition (McClelland & Rumelhart, 1981). Since a target word and its neighbors only differ in one letter, when the target word is presented, the word nodes for the neighboring words would be activated early in processing along with the target word. The activations in turn feedback to facilitate the activations of the constituent letter nodes. The feedback activations are particularly helpful for naming low frequency words. On this view, the facilitation results from lexical contribution to the orthographic activation, although as argued by Peereman and Content (1995), lexical activation of neighbors could contribute to phonological computation rather than orthographic processing. One problem with this account is that it would predict the same facilitation for lexical decision as naming, but as we have indicated the data for lexical decision is much more complex.

An alternative hypothesis is that the neighborhood size effect is related to phonological computation (Peereman & Content, 1995; 1997). According to this view, the effect is not limited to orthographic processing; rather it can be attributed to the variability of phonological properties among orthographic neighbors. Evidence for this view comes from a study by Peereman and Content (1997), in which they examined the influence of different types of orthographic and phonological neighbors on naming. The results showed that when the orthographic neighbors were also phonological neighbors (i.e., they are phonographic neighbors), the facilitation in naming was the strongest compared with other types of neighbors. Thus, they argued that the extent to which orthographic neighborhood size could accelerate phonological computation is dependent on the similarity between the phonological codes of neighboring words and the target word. This finding is corroborated by the results reported in a multiple regression study on four large English naming datasets (Adelman & Brown, 2007), where the number of phonographic neighbors was a stronger predictor than the conventional neighborhood size in accounting for naming data.

The phonological computation account of the neighborhood size effect is supported by most current theories of reading (Adelman & Brown, 2007). According to the dual-route cascade (DRC) models (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Perry, Ziegler, & Zorzi, 2007), the facilitatory effect in naming is expected because the models allow the processing to activate orthographic neighbors of word stimuli in the orthographic lexicon, which in turn activates phonological entries and phonemes. Phonetic activation generated from the lexical route along with that generated from the non-lexical route would speed the naming latencies. Within the parallel distributed processing (PDP) models (Chang, Furber, & Welbourne, 2012, Harm & Seidenberg, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) the effect emerges as the system adjusts its connection weights following exposure to the shared orthographic structure of the neighboring words.

While the facilitatory effect of neighborhood size in naming is robust in alphabetic languages, recent studies in Chinese have showed a contradictory pattern of neighborhood size effects (Li, Bi, Wei, & Chen, 2011; Zhao, Li, & Bi, 2012), where the orthographic neighbors of phonetic radicals tend to increase naming latencies. To our knowledge, there are no studies in alphabetic languages reporting an inhibitory effect of neighborhood size in naming; this effect seems to be reported only in studies based on Chinese characters. In Chinese, over 80% of characters are phonograms, which consist of a semantic radical (usually on the left) and a phonetic radical (usually on the right) (Zhou, 1978). In general the semantic radical provides some information relating to meaning, while the phonetic radical provides some information about pronunciation. The *neighborhood size of phonetic radicals* is defined as the number of characters that share the same phonetic radical. Two relevant measures are the *orthography-to-phonology consistency* of a character, indicating whether the pronunciation of a character agrees with other characters containing the same phonetic radical, and *regularity,* which is defined as whether a character is pronounced the same as its phonetic radical under the constraint that the phonetic radical is pronounceable (Fang, Horng, & Tzeng, 1986). These definitions are based on similar concepts to those used in English (Coltheart, 1978; Glushko, 1979). Despite the fact that Chinese has a very different orthographic system from alphabetic languages, most typical reading effects such as frequency effects (Balota et al., 2004; Forster & Chambers, 1973; Hue, 1992; Lee, Tsai, Su, Tzeng, & Hung, 2005) and regularity or consistency effects (Glushko, 1979; Lee et al., 2005; Taraban & McClelland, 1987) tend to have a similar pattern across English and Chinese, and those effects also have been simulated by computational models based on the same general learning principles in Chinese (Hsiao & Shillcock, 2004; Yang, McCandliss, Shu, & Zevin, 2009). It remains unclear how a more language-specific effect (i.e., the inhibitory effect of neighborhood size) seen in Chinese emerges in the reading system.

1.1. The effect of orthographic neighborhood size in Chinese character naming

On the basis of the orthographic structures of Chinese phonograms, two different types of orthographic neighbors can be defined: semantic radical neighbors and phonetic radical neighbors (Feldman & Siok, 1999). Of particular interest here is the phonetic radical neighborhood size, also known as *phonetic combinability,* because it is directly linked to phonology. Throughout this paper we will use orthographic neighborhood size in Chinese to refer to the neighborhood size of phonetic radicals, unless stated otherwise.

Several studies in Chinese character reading have examined the effects of orthographic neighborhood size and consistency simultaneously because they are closely related to phonetic radicals (Hsu, Lee, & Tzeng, 2014; Hsu, Tsai, Lee, & Tzeng, 2009; Li et al., 2011; Zhao et al., 2012). Li et al. (2011) found an inhibitory effect of neighborhood size for inconsistent characters while a null effect was observed for consistent characters. However, when the high frequency neighboring characters in the inconsistent condition were removed, the effect became facilitatory. They suggested that the neighboring characters of a target character might accelerate activation at the orthographic level; however, any high frequency neighbors of the target would cause interference at the phonological processing stage, resulting in an inhibitory effect. This interpretation is partly consistent with the phonological computation account (Adelman & Brown, 2007; Peereman & Content, 1997), suggesting that the effect of orthographic neighborhood size is not limited to the orthographic level but it is also dependent on the stage of phonological processing. Further evidence for this comes from a study of event-related potentials (ERPs) conducted by Hsu et al. (2009). They examined the effects of orthographic neighborhood size and consistency in Chinese character reading. They demonstrated that characters with large neighborhoods facilitated the earlier stages of orthographic (N170) and phonological processing (P200) relative to characters with small neighborhoods. They also elicited larger negativity at the later stage of semantic processing (N400), suggesting an increase of semantic competition for high neighborhood characters. Their results suggest the effect of orthographic neighborhood size is widespread throughout the reading system.

According to the phonological computation account (Peereman & Content, 1997), one might expect a similar effect of neighborhood size in both English and Chinese, if considering the orthographic activation of phonology alone. However, as agreed by most current theories of reading, there are two different pathways active during reading (Coltheart et al. 2001; Plaut et al. 1996). Within the PDP models of reading (Harm & Seidenberg, 2004; Plaut et al., 1996) there is a phonological pathway from orthography to phonology and a semantic pathway from orthography to phonology via semantics. The division of labor between pathways is greatly shaped by the nature of the orthographic systems (Yang, Shu, McCandliss, & Zevin, 2013). In English, the mappings between orthography and phonology are mostly consistent, which contrasts with the arbitrary mappings between orthography and semantics. So learning the mappings in the phonological pathway is much faster and the connection weights are optimized for the consistent spelling-to-sound mappings. Thus, consistent words can utilize the phonological pathway very efficiently for their pronunciations. While inconsistent words can utilize the phonological pathway, at the same time they may also partly rely on the semantic pathway for their pronunciations (Plaut et al. 1996).

Chinese has less transparent mappings between orthography and phonology but more regular mappings between orthography and semantics compared with those in English. Moreover, in English, inconsistent words still have many subcomponents that are shared among words. For instance, the pronunciation of an inconsistent word, *pint,* can benefit from the pronunciation of *pant* because they share the same onset and coda and it is only the vowel section that is inconsistent. In Chinese most inconsistent characters do not share any phonetic components with their phonetic radicals (e.g., 灑 /sa3/ - 麗 /li4/). However, some characters may share either onset or rime. For example, the inconsistent character 結 /jie2/ shares the same onset (i.e. /j/) with its phonetic radical 吉 /ji2/ while the inconsistent character 妒 /du4/ shares the same rime (i.e. /u/) with its phonetic radical 戶 /hu4/. However, in Chinese this partial information derived from the phonetic radical is not very helpful in determining the pronunciation of inconsistent characters (Chen, Shu, Wu, & Anderson, 2003). Overall, these properties might suggest that the semantic pathway would play a more important role in Chinese character naming than it does in English. This might explain the inhibitory effect of neighborhood size seen in inconsistent Chinese characters (Li et al. 2011). Specifically, a performance cost observed for inconsistent words with large neighborhood may be due to the conflict between orthographic activation of phonology from the phonological pathway and semantic activation of phonology from the semantic pathway.

1.2. Computational models of Chinese character naming

Although a number of theoretical models of Chinese reading have been proposed in the literature (Perfetti, Liu, & Tan, 2005; Perfetti & Tan, 1999; Taft & Zhu, 1997), only recently have large-scale computational models been developed (Hsiao & Shillcock, 2004; Yang et al., 2009; Yang et al., 2013). These provide explicit details about the connections between the core processing layers within the system and allow for effective evaluation of reading effects. In particular, a recent parallel distributed processing model of Chinese character naming by Yang et al. (2009) has demonstrated that the same statistical learning principles can be applied to both English and Chinese. They developed a computational model of Chinese character naming on the basis of the previous models in English (Harm &

Seidenberg, 1999), with revised representations to represent Chinese orthography and phonology. The model was able to capture the pattern of interaction between frequency and consistency seen in skilled Chinese readers (Hue, 1992; Lee et al., 2005). Analyses of internal representations revealed that phonetic radicals emerged as critical processing units over learning. These findings show that the processing of orthography-to-phonology conversion for Chinese has strong similarities to English. However, the relationship between neighborhood size of phonetic radicals and consistency has not been investigated in their model.

The aim of the present study was to develop computational models based on the parallel distributed processing framework and to test whether the division of labor between phonological and semantic pathways is the key to accounting for the inhibitory effect of neighborhood size observed in Chinese character naming. We investigated this by conducting a behavioral naming task in order to replicate previous results of Li et al. (2011). We then developed two computational models of Chinese character naming based on previous models (Chang, Furber, & Welbourne, 2012; Plaut et al., 1996; Yang et al., 2013): one contained only the phonological pathway from orthography to phonology and the other one contained both the phonological and semantic pathways by providing both orthographic and semantic inputs to phonology. We also incorporated a visual processing stage into models, which allowed orthographic representations to be developed over the course of learning (Chang et al., 2012). We expected that both models would account for the typical interaction between frequency and consistency shown in the previous Chinese reading models (Hsiao & Shillcock, 2004; Yang et al., 2009). However, if the semantic pathway contributes to the emergence of the inhibitory effect of neighborhood size, we would expect that only the model including a semantic pathway would show the inhibitory effect of phonetic neighbors as seen in skilled Chinese readers.

2. Behavioral Experiment

The aim of this experiment was to replicate previous findings of neighborhood size effects in Chinese character naming (Li et al., 2011) by manipulating character consistency and neighborhood size of phonetic radicals. Specifically, we investigated whether there is an inhibitory effect for inconsistent characters with many neighbors and how the processing of consistent characters is affected by neighborhood size. Although Li et al., (2011) reported a null effect of neighborhood size for consistent characters, it might be worthy of note that the average consistency score in their high consistent condition was about 0.6, which was closer to a medium level of consistency. The present experiment thus attempted to provide a better control of the different levels of consistency.

2.1. Method

2.1.1. Participants

Twenty-seven native Chinese speakers from National Chengchi University in Taiwan participated in the naming task (male=10; average age=22). The official and most commonly used spoken language in Taiwan is Mandarin Chinese, though there is a widespread dialect known as Taiwanese. In addition, the writing system used in Taiwan is based on traditional characters. All participants had normal or corrected-to-normal vision, and none were reported to have any cognitive problems. They were paid for their participation. This study was approved by human subjects research ethics, Academia Sinica, Taiwan.

2.1.2. Stimuli

Character consistency was quantified in terms of the ratio of the number of characters sharing a phonetic radical that have the same pronunciation relative to the total number of characters sharing that phonetic radical; tonal differences were ignored in line with prevous studies (Fang et al., 1986, Lee et al. 2005; Hsu et al. 2009). For example, the phonetic radical

少 (/shao3/, "little") is shared by six characters; three of which (沙, 砂 and 紗) are pronounced in the same way (/shao/) so they have a consistency value of 0.5 (i.e., 3/6). Orthographic neighborhood size, also known as phonetic combinability (Feldman & Siok, 1999) or phonetic radical frequency by type, was defined as the number of characters that share the same phonetic radical. For example, the character 鯨 (/jing1/, "whale") has neighbors including 涼 (/liang2/), 掠 (/lue4/), 晾 (/liang4/), 諒 (/liang4/), and 黥 (/qing2/).

Orthographic neighborhood size had two conditions: large and small. Consistency also had two conditions: consistent and inconsistent. Each condition comprised 30 items. As can be seen in Table 1, all selected stimuli were of low frequency (average 31 per million)[1]. As intended the consistent group had significantly higher consistency scores than the inconsistent group, $t(118) = 38.28$, $p < .001$, and the large neighborhood size group had significantly more neighbors than the small neighborhood size group, $t(118) = 17.74$, $p < .001$. To control for a potential confounding effect in naming, characters across conditions were matched for number of strokes, character frequency, number of semantic radical neighbors and number of phonological neighbors. The number of semantic neighbors measures the neighborhood size of the semantic radical. In addition, the number of phonological neighbors measures the number of characters that can be generated by substituting a single phoneme. Note that in Chinese, the syllable structure is relatively simple and most Chinese syllables can be mapped onto more than one character so the number of phonological neighbors for a given character is generally rather large. Between-items one-way ANOVA analyses confirmed that the condition groups did not differ significantly on any

[1] According to the Chinese naming database by Chang et al. (2016), the frequency scores in Chinese range from 1 to 60158 per million. To categorize the frequency scores into five levels (very low, low, medium, high and very high), we computed the percentiles of the frequency distribution in steps of 20. The 20th, 40th, 60th, and 80th percentiles of the frequency distribution are 10, 45, 182, 1005.4 per million respectively. So the average frequency score 31 per million falls between 20th and 40th percentiles of the distribution and thus is considered as low frequency.

of those measures all *ps* > .05. The measures of all the lexical factors used in stimulus selection were taken from or computed based on an online psycholinguistic naming database (Chang, Hsu, Tsai, Chen, & Lee, 2016). Table 1 shows the descriptive statistics along with a representative example of characters in each condition.

Table 1. Descriptive statistics of a range of psycholinguistic variables as a function of consistency and neighborhood size

	Consistent		Inconsistent	
	Large neighborhood	Small neighborhood	Large neighborhood	Small neighborhood
	伕 /fu1/	仟 /qian1/	橙 /cheng2/	絨 /rong2/
Consistency	0.91 (0.018)	0.91 (0.022)	0.25 (0.014)	0.29 (0.013)
Neighborhood size	9.83 (0.60)	3.47 (0.13)	11.27 (0.45)	3.6 (0.16)
Frequency	30 (4.24)	30.83 (3.37)	29.03 (4.33)	34.2 (4.16)
Stroke	13.53 (0.67)	14.7 (0.83)	13.1 (0.75)	13.73 (0.89)
Semantic radical neighbors	82.83 (10.51)	79.1 (7.65)	81.13 (9.4)	80.67 (9.9)
Phonological neighbors	1144.43 (64.74)	1046.9 (53.49)	1128.13 (60.41)	1023.3 (64.41)

Note: Means and standard errors of variables in brackets

2.1.3. Procedures

The experiment was conducted in a small test room. Participants were instructed to read aloud all the words as quickly and as accurately as possible. The presentation sequence of the stimuli was randomized for each participant. A voice key with headset connected to the computer was used to obtain participants' response latencies. Before the experiment began, there were 20 practice trials. This allowed the participants to familiarize themselves with the procedure. Participants' responses to the practice trials were also used to adjust a loudness threshold for individual participants where necessary. In the experiment, each trial started with presentation of a cross fixation point for 400 ms, followed by a target character, accompanied with a beep sound for 200 ms. The target character remained on the screen until response. There were three breaks during each session. The whole experiment was recorded by using an audio recorder.

All errors resulting from mispronunciations, inaccurate voice key activation or measurement errors were recorded at the time by the experimenter.

2.2. Results

All incorrect trials including a mispronounced error or a voice key error were excluded. In addition two characters were excluded because of error rates of over 60%. The average naming accuracy for all the items was 87.5%. To calculate the average naming latencies, any response time faster than 300 ms or slower than 2000 ms was removed. Naming latencies outside two standard deviations from the mean were also excluded. These resulted in a removal of 5.24% of responses in total. The descriptive results are presented in Table 2.

Table 2. Mean naming latencies and accuracy rates as a function of consistency and orthographic neighborhood size

	Consistent		Inconsistent		Consistency effect	
	RT	Accuracy	RT	Accuracy	RT	Accuracy
Large Neighborhood	793 (11)	0.90 (0.01)	893 (16)	0.83 (0.01)	100	0.07
Small Neighborhood	815 (10)	0.94 (0.07)	866 (15)	0.83 (0.01)	51	0.11
Neighborhood effect	22	0.04	33	0		

Naming latencies were analyzed using a linear mixed effects model (Baayen, Davidson, & Bates, 2008) and, as recommended by Jaeger (2008), the accuracy data was analyzed using a generalized linear mixed effects model with a binominal distribution. Models were fit using the lme4 package in R (version 3.2.0, 2015). For both naming latencies and accuracy data, the significance of individual and interaction factors was assessed using a likelihood ratio test to determine whether the model fit changed significantly between models with and without the factor or interaction of interest. To control for potential onset effects caused by the sensitivity of voice key to different onsets (Balota et al., 2004; Chang et al. 2016; Liu et al., 2007), the initial phonemes were also included as a fixed effect in the linear mixed effects models. Following Balota et al. (2004), the initial phoneme of each character was coded dichotomously (1 or 0) for the following 13 features, where 1 denoted the presence of the feature and 0 denoted its absence: stop, affricate, fricative, nasal, liquid, aspirated, voiced, bilabial, labiodental, alveolar, palato-alveolar, alveolo-palatal, and velar.

For naming latencies, as a baseline, a model with participant and item as random effects was created. Adding initial phonemes as a fixed effect did not significantly improve model fit relative to the baseline model, $\chi^2(11) = 12.31$, $p > .05$, suggesting that the initial phonemes did not significantly affect naming latency. Nevertheless, we still included the initial phonemes into subsequent analyses in case they might have differential effects on naming across different experimental conditions. The results showed that the main effect of consistency was significant: adding consistency as a fixed factor improved model fit compared to a model with random effects of participant and item and fixed effects of neighborhood size and initial phonemes, $\chi^2(1) = 33.06$, $p < .001$, while the main effect of neighborhood size was not: adding neighborhood size as a fixed factor did not result in a significant improvement in model fit compared to a model with random effects of participant and item and fixed effects of consistency and initial phonemes, $\chi^2(1) = 0.02$, $p > .05$. However, adding the interaction between consistency and neighborhood size resulted in a significant improvement in fit compared a model containing random effects and main effects, $\chi^2(1) = 4.03$, $p < .05$, with a facilitatory effect of neighborhood size in the consistent condition, t = -2.57, $p < .05$, and an inhibitory effect in the inconsistent condition, t = 2.31, $p < .05$.

For the accuracy rate analysis, a baseline model with participant and item as random effects and with accuracy (correct or incorrect) as the dependent variable was created. Again, the inclusion of initial phonemes did not result in a significant change in model fit, $\chi^2(11) = 14.09$, $p > .05$, compared to the baseline model. The consistency effect was significant: adding consistency as a fixed factor improved model fit compared to a model with random effects of participant and item and fixed effects of neighborhood size and initial phonemes, $\chi^2(1) = 17.16$, $p < .001$, but neither the effect of neighborhood size nor the interaction between consistency and neighborhood size reached significance: adding neighborhood size as a fixed factor did not result in a significant improvement in model fit compared to a model with random effects of participant and item and fixed effects of consistency and initial phonemes, $\chi^2(1) = 0.80$, $p > .05$. Similarly, adding the interaction term between consistency and neighborhood size did not significantly improve model fit, $\chi^2(1) = 1.26$, $p > .05$, when both the random and main effects were included.

2.3. Discussion

The results of the preceding experiment demonstrate that inconsistent characters with large neighborhood size inhibited naming performance, which is congruent with the data reported by Li et al. (2011). However, when the characters were consistent and had many neighbors, a facilitatory effect was observed. This is congruent with the findings of previous studies in English (Andrews 1989, 1992; Balota et al., 2004; Carreiras et al., 1997; Grainger, 1990; Peereman & Content, 1995; Sears et al., 1995), suggesting that if characters come from a neighborhood with consistent mappings between orthography and phonology then the neighbors facilitate processing. Although the faciliatory effect of neighborhood size for consistent words appears to contradict the null effect found in the Li et al. (2011) study, it might be because the average consistency score in the high consistent condition in the present study is considerably higher than that used in their study (0.91 versus 0.6). Collectively, the results replicated previous findings of neighborhood size effect in Chinese character naming, and showed the effect was modulated by the consistency of the mappings between phonetic radicals and phonemes.

3. Simulations

Two computational models based on parallel distributed processing were developed to explore whether semantic processing contributes to the inhibitory effect of neighborhood size in Chinese character naming, particularly for inconsistent characters. The architecture of the models follows that of previous PDP models in English and in Chinese (Chang et al., 2012; Yang et al., 2013). The first model contained the reading pathway from visual-orthographic input (V) to phonology (P), termed *VP model*. In addition to the V->P pathway, the second model also contained the pathway from semantics (S) to phonology (P), termed *VSP model*, following Yang et al. (2013, Simulation 1), where the semantic input mainly functioned to provide additional semantic information about word identity to generate phonological output. As demonstrated by Plaut et al. (1996), the semantic pathway is particularly useful for the processing of words with inconsistent spelling-to-sound mappings. It should be acknowledged that this implementation of the semantic pathway is subject to some degree of simplification because it receives semantic input directly rather than visual-orthographic input, like the semantic pathway in a fully implemented connectionist model of reading developed by Harm and Seidenberg (2004). This characterization of the semantic pathway allows us to investigate whether the division of labor between the phonological and semantic pathways interacts with neighborhood and consistency in such a way as to account for the neighborhood effects in Chinese character naming.

The representations of orthography in the present models were not pre-specified; rather they were allowed to develop during the time course of training. This effectively mimics the development of orthography during reading acquisition in children (Chang et al. 2012). The training corpus included 3,621 characters taken from the Academia Sinica

Balanced Corpus (Huang & Chen, 1998). All characters were phonograms, which consisted of one semantic radical and one phonetic radical. To evaluate performance, both models were tested to see if they could account for the standard frequency and consistent effects (Lee et al., 2005). However, the key test was the ability of the models to simulate the neighborhood size by consistency interaction found in the present behavioral data.

3.1. Method

3.1.1. Model architecture

Both the VP and VSP models were feedforward networks. Figure 1 shows the architectures of both models. The VP model (Figure 1, left panel) consisted of five layers of units: 800 visual inputs, 60 hidden units, 100 orthographic hidden units, another 200 hidden units and 105 phonological units. The VSP model (Figure 1, right panel) had a phonological pathway and a semantic pathway. The phonological pathway for the VSP model consisted of 800 visual inputs, 60 hidden units, 100 orthographic hidden units, another 200 hidden units and 105 phonological units. The semantic pathway for the VSP model consisted of 200 semantic units and 100 hidden units. For each model, every unit in one layer was fully connected to the next layer. In this way, the VP model is a subset of the VSP model, which allows us to test whether the orthographic neighborhood size effect is dependent on the semantic pathway. However, the total number of units in both the VP and VSP models is not equal, where the VSP mode may have more capacity than the VP model. To exclude the possibility that the differential effects produced by the two models were due to a total capacity limitation, we have trained another VP model as a control model that has the same total number of units as that of the VSP model. As can be seen in Appendix A, the patterns produced by the control model are similar to that of the present VP model.

Insert Figure 1 about here

3.1.2. Visual image representations

The visual image representation scheme was adopted from the methodology used in Chang et al. (2012). Each character was represented by a bitmap image. The models were directly trained with those bitmap images as input. Two columns were used for the two constituent radicals of each character[2] and the size of each column was 20 x 20 pixels. Each character was positioned with its left radical on the first column and its right radical on the second column as that seen in print. An example of character can be seen in Figure 1.

3.1.3. Phonological representations

Phonological representations were generated according to Chinese phonetic features (Ministry of Education). Five phoneme slots were used to encode an initial consonant (C), a vowel (V), a medial (M), an ending consonant (C) along with a tone (T). Each of the first four phoneme slots consisted of a set of 25 phonological features and the last phoneme slot consisted of a set of 5 tonal features. All phonetic features were encoded as a binary value either 1 or 0. There were in total 105 features for each phonological representation. Figure 2 shows the phonological representation scheme used in the simulations.

Insert Figure 2 about here

----

[2] Although most characters used here have a semantic-phonetic structure (i.e. semantic radicals on the left and phonetic radicals on the right), there are some characters that have a phonetic-semantic structure (i.e. phonetic radicals on the left and semantic radicals on the right).

3.1.4. Semantic representations

According to previous simulation studies (Plaut, 1997; Plaut & Booth, 2000), semantic knowledge of words can be effectively represented by a set of systematically structured random features. This captures the arbitrary nature of the mappings between semantics and orthography, and between semantics and phonology. One would also expect that words within the same semantic category should share more features than words belonging to different categories. This scheme of constructing random but overlapping semantic representations was adopted in the present study. Semantic representations were constructed in respect of the knowledge of semantic radicals. It was assumed that characters having the same semantic radical would share some semantic features. Thus, a set of unique prototype patterns was first created for each semantic radical family. Each prototype consisted of 200 semantic features in which each feature had a probability of 0.1 being active: it was coded as 1 if a feature was active; otherwise it was coded as 0. Each prototype was then used to generate semantic vectors for each character within the same semantic radical family by regenerating the values of semantic features of its prototype with a 0.05 probability of changing zeros into ones or vice versa. A further constraint that Euclidean distance between two representations should be at least three was also applied. The average number of semantic features for each vector was 28.28 (SD=3.79). Illustrations of the semantic representations used in the VSP model are shown in Figure 3.

Insert Figure 3 about here

3.1.5. Training procedures

The training parameters for the VP model and the VSP model were exactly the same. Both models were trained using a back-propagation algorithm with a learning rate of 0.1 and a weight decay of 1E-7. Error scores were computed on the basis of the cross entropy function (Plaut et al., 1996) and were used to adjust weight changes during training. The initial weights were set to random values ranging from -0.1 to 0.1. The frequency of each character was taken from psycholinguistic norms (Chang et al., 2016), and the score was compressed by using a square-root function with a cutoff frequency of 2000. The frequency of character presentation was implemented by scaling the error derivatives during the back-propagation procedure based on the compressed frequency (Plaut et al., 1996). Both the VP and VSP models were trained on the whole set of 3,621 characters. The task for the VP model was to learn the mappings from visual inputs to phonological outputs, and the task for the VSP model was to learn the mappings from both the visual and semantic inputs to phonological outputs. Each model was trained ten times with different random initial weights and different presentation order of the characters.

It might be argued that the VSP model should have been given some pretraining in the links between the semantic units and the phonological units to mimic the situation in reading development where children learn to speak before they learn to read. However, in children, making use of this pretrained pathway requires learning previously untrained mappings between orthography and semantics, which do not exist in our model. Thus, if we had taken this approach, the model would have been able to read via the semantic pathway before we started to train it, which is clearly not realistic.

3.1.6. Testing procedures

The testing procedures for both the VP and VSP models were exactly the same. The decoding procedure for phonology was based on the activation of the phonological units. Error score was measured by the sum of the squared differences between the activation of each output unit and its target activation. For accuracy, the activities of phonological units were first binarized: if unit activation was greater than .5, the unit was considered active;

otherwise it was classified as non-active. Accuracy of the model was then assessed by determining whether the phonological output pattern was the same as the target representation.

3.2. Results

The training procedures were halted after 18 million presentations for the VP model, and at this point it could correctly pronounce 99.12% of characters. For the VSP model, the training time was 20 million presentations, which was longer than that of the VP model. This was probably due to the interference caused by the semantic pathway before it learned to contribute coherently. At the end of training, the VSP model achieved an accuracy rate of 99.94% at the phonological level.

3.2.1. Frequency and consistency effects

Previous behavioral and simulation studies of Chinese character naming have shown an interaction between frequency and consistency (Lee et al., 2005; Yang et al., 2009), where the naming latencies for inconsistent characters are slower than for consistent characters, in particular for low frequency characters. It is thus important to verify whether both the VP model and VSP model could replicate the typical effect of frequency and consistency in Chinese character naming.

Both models were tested on six sets of characters, consisting of two levels of frequency (High or Low) and three levels of character type (Consistent Regular, Inconsistent Regular, or Inconsistent Irregular). Consistency is the ratio of the number of characters sharing a phonetic radical that have the same pronunciation, to the total number of characters sharing that phonetic radical; regularity is defined as whether a character is pronounced the same as its phonetic radical (Fang et al., 1986). Each set of stimuli comprised 20 characters, all taken from Lee et al. (2005, experiment 1). Error score was used as an analogy of participants' naming latencies (Seidenberg & McClelland, 1989; Plaut et al. 1996).

Linear mixed effects models were applied to analyze model performance with both item and simulation number (one to ten) as random factors and error score as a dependent variable. Error items and outliers (greater than two and a half standard deviations from the mean) were removed prior to analyses. This resulted in a removal of 6.58% of items in total for the VP model and a removal of 6.08% of items in total for the VSP model. Figure 4 displays the mean error score produced by both the VP model and VSP model on characters with different types as the function of frequency. The behavioral pattern reported in Lee et al. (2005) is also included in Figure 4. For the VP model, the frequency effect was significant: adding frequency as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of character type, $\chi^2(1) = 49.21$, $p < .001$. The effect of character type was also significant: adding character type as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of frequency, $\chi^2(2) = 8.12$, $p < .05$. There was also a significant interaction between frequency and character type, where the model fit was improved by adding the interaction between frequency and character type, $\chi^2(2) = 7.78$, $p < .05$, when both the random and main effects were considered. For the VSP model, both the effects of frequency and character type were significant: adding frequency as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of character type, $\chi^2(1) = 67.20$, $p < .001$, and adding character type as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of frequency, $\chi^2(2) = 8.87$, $p < .05$. The interaction between frequency and character type was also significant: the model fit was significantly improved by adding the interaction between frequency and character type, $\chi^2(2) = 7.07$, $p < .05$, when both the random and main effects were considered. These results demonstrated that both the VP model and VSP model were able to produce a similar pattern of frequency and consistency as seen in Lee et al. (2005).

However, as can be seen in Figure 4, in both models the frequency effect appears larger than observed in the behavioral data. It is possible that using a logarithmic transformation, or extending the training for longer would bring the effect sizes closer together.

Insert Figure 4 about here

3.2.2. Neighborhood size and consistency effects

The key test for the present study was to see whether the models could capture the interaction pattern between neighborhood size and consistency. Both the VP model and VSP model were tested on the same stimuli as in the preceding behavioral experiment. Linear mixed effects models were applied to analyze model performance with both item and simulation number (one to ten) as random factors and error score as the dependent variable. Error items and outliers (greater than two and a half standard deviations from the mean) were removed prior to analyses. This resulted in a removal of 6.75% of items in total for the VP model and a removal of 4.83% of items in total for the VSP model. The interaction patterns produced by the VP and VSP models along with the participants' interaction data are shown in Figure 5.

Insert Figure 5 about here

For the VP model, the consistency effect was significant: adding consistency as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of neighborhood size, $\chi^2(1) = 7.15$, $p < .01$, but the effect of neighborhood size was not: adding neighborhood size as a fixed factor did not result in a significant improvement in model fit compared to a model with random effects of simulation and item and a fixed effect of consistency, $\chi^2(1) = 0.29$, $p > .05$. The interaction between consistency and neighborhood size was also not significant: the model fit was not significantly improved by adding the interaction term between consistency and neighborhood size, $\chi^2(1) = 0.04$, $p > .05$, when both the random and main effects were included.

For the VSP model, the analysis revealed a significant main effect of consistency: adding consistency as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of neighborhood size, $\chi^2(1) = 6.16$, $p < .05$, while the main effect of neighborhood size was not: adding neighborhood size as a fixed factor did not result in a significant improvement in model fit compared to a model with random effects of simulation and item and a fixed effect of consistency, $\chi^2(1) = 0.83$, $p > .05$. Critically, the interaction between consistency and neighborhood size was significant: the model fit was significantly improved by adding the interaction term between consistency and neighborhood size, $\chi^2(1) = 5.88$, $p < .01$, when both the random and main effects were included. The interaction pattern was similar to participants' data in which a facilitatory effect of neighborhood size in the consistent condition, $t=-2.54$, $p < .05$, and an inhibitory effect was observed in the inconsistent condition, $t=3.01$, $p < .05$.

To confirm that there was a reliable difference between the VP model and the VSP model, we conducted additional linear mixed effects models to test the significance of a three-way interaction between consistency x neighborhood size x model type (VP or VSP). A control model with random effects of item and simulation, and fixed effects of consistency, neighborhood size and model type, and a two-way interaction between consistency and neighborhood size was created. Adding the three-way interaction between consistency, neighborhood size and model type significantly improved the model fit, $\chi^2(3) = 31.71$, $p < .001$, compared to the control model, indicating the difference between the VP model and the VSP model was statistically reliable.

3.3. Exploring the influence of semantic processing on the effect of neighborhood size

It seems that neighboring characters can facilitate the processing of a target character containing a consistent mapping between visual orthography and phonology. This was observed in both the VP model and the VSP model. However, only the VSP model could produce an inhibitory effect for inconsistent characters with large neighborhoods as seen in the participants' data. This suggests that the inclusion of a semantic pathway is the key to accounting for the correct interaction pattern between neighborhood size and consistency. To explore this we examined how the division of labor between the phonological pathway, vision to phonology (V->P), and the semantic pathway, semantics to phonology (S->P) was modulated by neighborhood size and consistency in the VSP model.

Several computational studies have utilized a lesion technique to explore the relevant contribution from different pathways to the activation of either semantics or phonology (Harm & Seidenberg, 2004; Welbourne, Woollams, Crisp, & Lambon Ralph, 2011). Following Welbourne et al. (2011), to obtain the contribution made by the phonological pathway (V->P), we first computed the error score in the phonological units after all the links between the hidden units in the semantic pathway and phonological units were removed. Similarly, for the contribution of the semantic pathway (S->P), the error score in the phonological units were computed after all the links between the hidden units in the phonological pathway and phonological units were removed. As the error score in the model indicated poor performance[3], the reciprocals of the two error scores were used to compute the proportional contribution made by the two pathways as a measure of division of labor.

We tested the VSP model on each group of characters used in the neighborhood size test by separately severing the links between the hidden units and the phonological units in each pathway. We then computed the division of labor between the pathways as described in

---

[3] The use of error scores rather than accuracy is because error scores can provide a more sensitive measure when the model is severely damaged (Welbourne, et al. 2011).

the previous paragraph. The results are shown in Figure 6. The phonological pathway made a larger contribution to the activation of phonology then the semantic pathway for all types of characters. However, the relative contribution for each type of character was different. The phonological pathway contributed more for consistent characters than for inconsistent characters. In contrast, the semantic pathway contributed more for inconsistent characters than for consistent characters, and most interestingly, the contribution was larger for inconsistent characters with small neighborhoods than for inconsistent characters with large neighborhoods. This is due to the fact that in the model the phonological pathway particularly specializes in learning consistent mappings and the strength of this effect is governed by neighborhood size. This in turn frees the semantic pathway to specialize in processing inconsistent characters with small neighborhoods. This graded specialization is what leads to the interaction between neighborhood and consistency.

Insert Figure 6 about here

3.4. Internal orthographic representations

In previous computational models of Chinese character naming, the orthographic representations have been predefined to encode the detailed orthographic structures of characters such as relative position of strokes or radicals, number of strokes or radicals in a character (Hsiao & Shillcock, 2004; Yang et al., 2009; Yang et al., 2013). In their analyses of the internal representations, Yang et al. (2009) showed that characters that shared the same phonetic radical tended to develop similar representations, suggesting the model was able to extract the critical functional unit in the translation from print to sound. In the present models, explicit orthographic information was not provided so it would be instructive to examine the internal representations of the orthographic hidden units in the present model to see what kind of orthographic features are represented. We conducted a principal components analysis on the representations of different characters sharing the same phonetic radical. The relationship between internal representations, visual similarity and consistency was also explored.

To test whether the VSP model had learnt to develop similar internal orthographic representations for characters in the same phonetic radical family, the activations of all orthographic hidden units for six representative phonetic radical families of characters were analyzed by principal components analysis. The first two principal components of the high-dimensional vectors formed by the activations of the orthographic hidden units were extracted. Figure 7 shows the components for the six phonetic radical families (皇 /huang2/, 半 /ban4/, 唐 /tang2/, 更 /qeng1/, 果 /guo3/, and 甫 /fu3/). The number of characters in each phonetic family ranged from four to ten. As can be seen, the representations formed distinctive clusters for each phonetic radical family. For instance, the representations for characters (徨, 煌, 蝗, 遑, 隍, 偟: /huang2/) sharing the same phonetic radical, 皇, formed a cluster in the orthographic space, which was clearly distinctive from those characters having different phonetic radicals. It is interesting to note that there was some variability in different phonetic radical families. It seemed that the representations for the phonetic radical's family members like 皇 had formed a tighter cluster while the representations for other phonetic radical's family members like 果 had formed a looser cluster.

Insert Figure 7 about here

To further explore the relations between internal representations and character consistency, we investigated whether the degree of similarity between orthographic representations in a given phonetic radical family depends on consistency, and how that is modulated by visual similarity. The entire training corpus was analyzed (in total 663 phonetic radical families). The average consistency score for each phonetic radical family was computed by averaging the consistency scores of all the family members, (M=0.54, SD=0.28). The visual similarity between two characters was measured by computing the cosine similarity between their visual representations. So the average visual similarity for each phonetic radical family was the average paired-wise cosine similarity of all the family members, (M=0.73, SD=0.06). The distance between internal representations for each phonetic radical family was measured by computing the paired-wise cosine distances between family members and then averaging the results, (M=0.012, SD=0.004). To explore the relation between visual similarity, consistency and orthographic hidden representations, we conduced a regression analysis with visual similarity, consistency and their interaction as predictors, and with cosine distance between representations as a dependent variable. The results showed that both visual similarity, $\beta = -0.44$, $p < .001$, and consistency, $\beta = -0.36$, $p < .001$ were significant; indicating that characters in the same phonetic radical family that are visually similar and have consistent pronunciations tend to develop similar orthographic hidden representations. Interestingly, the interaction between consistency and visual similarity was also significant, $\beta = 0.15$, $p < .001$ (see Figure 8), showing that the effect of visual similarity is stronger in inconsistent characters than in consistent characters. These results suggest the degree of similarity of representations between characters in the same phonetic radical family is not only determined by visual similarity of characters but also depends on the phonological properties of phonetic radicals.

Insert Figure 8 about here

----

4. General discussion

The primary aim of this study was to investigate the effect of orthographic neighborhood size in Chinese character naming. The behavioral data showed a facilitatory effect of neighborhood size for consistent characters whereas an inhibitory effect was observed for inconsistent characters (Figure 5). Two computational models of Chinese character naming were developed to explore the functional cause of the inhibitory effect observed in the behavioral data. Although both the VP and VSP models could simulate the frequency and consistency effect, only the model which included a semantic pathway (the VSP model) was able to produce the correct interaction pattern between neighborhood size and consistency (Figure 5), suggesting that semantic processing plays a key role in accounting for the effect of neighborhood size in Chinese character naming. It is worth noting that in this study the contribution of the semantic pathway is particularly strong for characters with small neighborhoods, which is contrary to what would be predicted from a purely lexical account (i.e., a larger effect of the lexical pathway for characters with large neighborhoods). This is because, on a PDP account, each pathway compensates for the properties of the other, whereas this is not true for purely lexical accounts.

Li et al. (2011) found an inhibitory effect of neighborhood size for inconsistent characters and they attributed the inhibition to high frequency neighbors. While the present finding of inhibitory effects for inconsistent characters is congruent with their data, we also observe a facilitatory effect for consistent characters. Analysis of frequency for all the items revealed that about 92% of the target characters had at least one high frequency neighbor so the high frequency neighbor hypothesis cannot explain all of our data. On this basis, we would suggest that it is the presence of neighbors with consistent pronunciations that drives the inhibitory effect for inconsistent items, though we acknowledge that the magnitude of the effect is likely driven by the summed frequency of those neighbors.

The interaction pattern between consistency and neighborhood size found in the present study indicates that if most neighbors of a target character share the same pronunciation with the character, the neighbors facilitate character naming; by contrast, if most neighbors and the target character have different pronunciations, the neighbors inhibit the naming latency. These results provide support to the phonological computation account of orthographic neighborhood size (Adelman & Brown, 2007; Peereman & Content, 1997). According to this account, the effect of neighborhood size is not limited to orthographic processing as proposed by Andrews (1989; 1992) but the effect also depends on the transformation of orthography-to-phonology. In particular, in a multiple regression study on four sets of large-scale word naming data, Adelman and Brown (2007) showed that orthographic neighbors of a word could greatly facilitate word naming when they were also phonological neighbors (i.e., phonographic neighbors) because they could provide the correct pronunciation for the target word. When this factor was taken into account, the measure of pure orthographic neighborhood size did not exert an additional unique facilitatory effect in accounting for naming latencies. Moreover, as indicated by Adelman and Brown (2007), if the orthographic neighbors of a target could not support the correct pronunciation, one might expect to see the inhibitory effect on naming. However, their regression results only showed the null or a little inhibitory effect of orthographic neighborhood size, suggesting the competitive inhibition between words in naming might be weak (Adelman and Brown, 2007). This may be because in alphabetic languages even inconsistent words have considerable consistent overlap with their inconsistent neighbors that would support the pronunciation of the onset and coda. In Chinese, phonetic consistency is a much more absolute concept. This may explain the clear inhibition effect of orthographic neighborhood size for inconsistent characters that has been observed in Li et al. (2011) and the current study. In addition, our simulation results suggest that the inhibition effect depends on semantic activation of phonology, presumably because the phonological route specializes in consistent mappings allowing the semantic pathway to compensate by learning the mappings for the inconsistent items. The lesion analyses of the VSP model also revealed that characters with small neighborhoods relied more on the semantic pathway compared with characters with large neighborhoods. Collectively, these results suggest that the division of labor between the semantic and orthographic activation of phonology is critical to explaining the inhibitory neighborhood effects in Chinese character naming.

This explanation can also shed light on the seemingly contradictory finding that neighborhood size in English has a facilitatory effect whereas it has an inhibitory effect in Chinese naming. In English semantic processing plays little role in naming words that have typical spelling-to-sound mappings, but is required for words that have atypical spelling-to-sound mappings (Strain, Patterson, & Seidenberg, 1995; Woollams, 2005). The majority of English words have middle to high consistency scores, as is evident from the average consistency scores in studies based on a large naming datasets. For instance, the average consistency score for words in Balota et al. (2004) is 0.9, and for words in Adelman and Brown (2007), it is 0.88. As the majority of English words are consistent in nature, a facilitatory effect of neighborhood size is expected because the contribution to phonology from the phonological pathway is dominant while the contribution to phonology from the semantic pathway is weak. In contrast, Chinese has a deep orthographic system where the spelling-to-sound information is less reliable. According to the Chinese naming database by Chang et al. (2016), about 89% of Chinese characters have pronounceable phonetic radicals. Only 46% of them share the same pronunciations with their phonetic radicals. This means that the pronunciation of a character derives directly from its phonetic radical may have 43% of chance to be wrong. Furthermore, if considering the degree of variation between pronunciations of the characters that share the same phonetic radical (regardless whether the phonetic radical is pronounceable), the average consistency score for Chinese characters is about 0.55 (Chang et al. 2016), which is much lower than the average consistency score (defined by similar concepts) in English. Given that the spelling-to-meaning information is not completely arbitrary, it is likely that semantic processing plays a much greater role in Chinese and indeed there is evidence showing that character naming performance is affected by a variety of semantic variables including imageability, concreteness, number of meanings, and semantic ambiguity (Chang et al., 2016; Liu, Shu, & Li, 2007).

The models developed in the present study can be considered as an extension of a previous model of Chinese character naming (Yang et al. 2009), sharing the same statistical learning principles. However, a unique feature of the present model is that the training starts from visual processing. Although the model did not explicitly encode the orthographic structures of characters, the internal representation analyses showed that the model was able to extract sensible orthographic representations through learning. Characters in the same phonetic radical family tended to develop similar internal representations, and the within-family distances of the internal representations were not merely determined by visual similarity of characters, but also determined by the degree of the characters sharing the same pronunciation. However, it should be noted that the present model did not include a pathway from vision to semantics. So the internal orthographic representations here were only modulated by phonological properties of characters. One might expect that the representations would also be sensitive to semantic properties of characters if the model had been trained on the mappings between vision and semantics in addition to the mappings between vision and phonology. As argued by Feldman and Siok (1999), both phonetic and semantic radicals are important processing units in Chinese. Thus future work should be conducted to explore the development of orthographic representations and the degree of sensitivity in orthographic processing to phonetic and semantic radicals.

5. Conclusion

The present study investigated how characters sharing the same neighboring structures (phonetic radicals) affect the naming latencies by using the behavioral experiment and computational modeling. The behavioral data showed orthographic neighbors could aid character naming when they shared the same pronunciation with the target; otherwise the inhibitory effect was observed. This provides support to the phonological computation account of orthographic neighborhood size. Simulation results showed that the model required both the orthography-to-phonology and the semantics-to-phonology processing routes to simulate this behavioral pattern.

Acknowledgements

This research was supported by grants under the National Science Council - Taiwan NSC96-2628-H-001-058-MY3 and NSC102-2420-H-001-006-MY2, and Academia Sinica - Taiwan AS-102-TP-C06. This research was also supported by Academia Sinica postdoctoral fellowship awarded to Ya-Ning Chang. We are grateful to Chun-Hsien Hsu for his help in preparing experimental stimuli, and to Wei-Fan Chen and Pei-Chun Chao for their help in conducting the naming experiment. We thank the Editor, David Plaut and the other anonymous reviewer for helpful comments.

References

Adelman, J., & Brown, G. A. (2007). Phonographic neighbors, not orthographic neighbors, determine word naming latencies. *Psychonomic Bulletin & Review, 14*(3), 455-459. doi: 10.3758/BF03194088

Andrews, S. (1989). Frequency and neigbhorhood effects on lexical access - activation or search. *Journal of Experimental Psychology-Learning Memory and Cognition, 15*(5), 802-814. doi: 10.1037/0278-7393.15.5.802

Andrews, S. (1992). Frequency and neigbhorhood effects on lexical access - lexical similarity or orthographic redundancy. *Journal of Experimental Psychology-Learning Memory and Cognition, 18*(2), 234-254. doi: 10.1037/0278-7393.18.2.234

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology-General, 133*(2), 283-316. doi: 10.1037/0096-3445.133.2.283

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi:10.1016/j.jml.2007.12.005

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighborhood in visual word recognition: cross-task comparisons. *J Exp Psychol Learn Mem Cogn, 23*(4), 857-871. doi: 10.1037/0278-7393.23.4.857

Chang, Y. N., Furber, S., & Welbourne, S. (2012). "Serial" effects in parallel models of reading. *Cognitive Psychology, 64*(4), 267-291. doi: 10.1016/j.cogpsych.2012.01.002

Chang, Y. N., Hsu, C. H., Tsai, J. L., Chen, C. L., & Lee, C. Y. (2016). A psycholinguistic database for traditional Chinese character naming. *Behavior Research Methods, 48*(1), 112-122. doi: 10.3758/s13428-014-0559-7

Chen, X., Shu, H., Wu, N., & Anderson, R. C. (2003). Stages in learning to pronounce Chinese characters. *Psychology in the Schools, 40*(1), 115-124. doi: 10.1002/pits.10073

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). New York: Academic Press.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance VI* (pp. 535-555). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review, 108*(1), 204. doi: 10.1037/0033-295X.108.1.204

Fang, S. P., Horng, R. Y., & Tzeng, O. J. L. (1986). Consistency effects in the chinese character and pseudo-character naming tasks. In H. S. R. Kao & R. Hoosain (Eds.), *Linguistics, Psychology, and the Chinese Language* (pp. 11-21). University of Hong Kong: Hong Kong: Center of Asian Studies.

Feldman, L. B., & Siok, W. W. T. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language, 40*(4), 559-576. doi: 10.1006/jmla.1998.2629

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12*(6), 627-635. doi: 10.1016/S0022-5371(73)80042-8

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *J Exp Psychol Learn Mem Cogn, 22*(3), 696-713.

Glushko, R. J. (1979). Organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology-Human Perception and Performance, 5*(4), 674-691. doi: 10.1037/0096-1523.5.4.674

Grainger, J. (1990). Word-frequency and neigbhorhood frequency - effects in lexical decision and naming. *Journal of Memory and Language, 29*(2), 228-244. doi: 10.1016/0749-596x(90)90074-a

Grainger, J., Oregan, J. K., Jacobs, A. M., & Segui, J. (1989). On the role of competing word units in visual word recognition: The neighborhood frequency effect. *Perception & psychophysics, 45*(3), 189-195. doi: 10.3758/BF03210696

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological review, 106*(3), 491-528. doi: 10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological review, 111*(3), 662-720. doi: 10.1037/0033-295x.111.3.662

Hsiao, J. H., & Shillcock, R. (2004). *Connectionist modeling of Chinese character pronunciation based on foveal split-ting.* In K. Forbus, D. Gentner, & T. Regier (Eds.), Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society (pp. 601-606). Mahwah, NJ: Lawrence Erlbaum Associates.

Hsu, C. H., Lee, C.-Y., & Tzeng, O. J. L. (2014). Early MEG markers for reading Chinese phonograms: Evidence from radical combinability and consistency effects. *Brain and Language, 139*(0), 1-9. doi: http://dx.doi.org/10.1016/j.bandl.2014.09.008

Hsu, C. H., Tsai, J. L., Lee, C. Y., & Tzeng, O. J. L. (2009). Orthographic combinability and phonological consistency effects in reading Chinese phonograms: An event-related potential study. *Brain and Language, 108*(1), 56-66. doi: 10.1016/j.bandl.2008.09.002

Huang, C. R., & Chen, K. J. (1998). Academia sinica balanced corpus.

Hue, C.-W. (1992). Recognition processes in character naming. In H.-C. E. Chen & T. O. J. L (Eds.), *Language processing in Chinese* (Vol. 90, pp. 93-107): Amsterdam: North-Holland.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59*(4), 434-446. doi: 10.1016/j.jml.2007.11.007

Lee, C. Y., Tsai, J.-L., Su, E. C.-I., Tzeng, O. J., & Hung, D. L. (2005). Consistency, regularity, and frequency effects in naming chinese characters. *Language and Linguistics, 6*, 175-197.

Li, Q.-L., Bi, H.-Y., Wei, T.-Q., & Chen, B.-G. (2011). Orthographic neighborhood size effect in Chinese character naming: Orthographic and phonological activations. *Acta Psychologica, 136*(1), 35-41. doi: 10.1016/j.actpsy.2010.09.012

Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods, 39*(2), 192-198. doi: 10.3758/BF03193147

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological review, 88*(5), 375-407. doi: 10.1037/0033-295X.88.5.375

Ministry of Education, Taiwan (1986). Mandarin Phonetic Symbols II.

Peereman, R., & Content, A. (1995). Neighborhood size effect in naming: Lexical activation or sublexical correspondences? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 409. doi: 10.1037/0278-7393.21.2.409

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: not all neighbors are equally influential in orthographic space. *Journal of Memory and Language, 37*(3), 382-410. doi: 10.1006/jmla.1997.2516

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 112*(1), 43-59. doi: 10.1037/0033-295x.112.1.43

Perfetti, C. A., & Tan, L. H. (1999). The constituency model of Chinese word identification. In A. Inhoff, H.-S. Chen, & J. Wang (Eds.), Reading Chinese script: A cognitive analysis (pp. 115-134). Mahwah, NJ: Erlbaum.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological review, 114*(2), 273-315. doi: 10.1037/0033-295x.114.2.273

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12*(5-6), 765-805. doi: 10.1080/016909697386682

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: empirical and computational support for a single-mechanism account of lexical processing. *Psychological review, 107*(4), 786-823. doi: 10.1037/0033-295X.107.4.786

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological review, 103*(1), 56-115. doi: 10.1037/0033-295X.103.1.56

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 876. doi: 10.1037/0096-1523.21.4.876

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review, 96*(4), 523-568. doi: 10.1037/0033-295X.96.4.523

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1140. doi: 10.1037/0278-7393.21.5.1140

Taft, M., & Zhu, X. (1997). Submorphemic processing in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 761. http://dx.doi.org/10.1037/0278-7393.23.3.761.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26*(6), 608-631. doi: 10.1016/0749-596X(87)90105-7

Welbourne, S. R., Woollams, A. M., Crisp, J., & Lambon Ralph, M. A. (2011). The role of plasticity-related functional reorganization in the explanation of central dyslexias. *Cognitive neuropsychology, 28*(2), 65-108. doi: 10.1080/02643294.2011.621937

Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: convergence and divergence. *J Exp Psychol Learn Mem Cogn, 31*(5), 878-890. doi: 10.1037/0278-7393.31.5.878

Yang, J. F., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language, 61*(2), 238-257. doi: 10.1016/j.jml.2009.05.001

Yang, J. F., Shu, H., McCandliss, B. D., & Zevin, J. D. (2013). Orthographic influences on division of labor in learning to read Chinese and English: Insights from computational modeling. *Bilingualism: Language and Cognition, 16*(02), 354-366. doi: 10.1017/S1366728912000296

Zhao, J., Li, Q.-L., & Bi, H.-Y. (2012). The characteristics of Chinese orthographic neighborhood size effect for developing readers. *PLoS ONE, 7*(10), e46922. doi: 10.1371/journal.pone.0046922

Zhou, Y. G. (1978). To what degree are the "phonetics" of present day chinese characters still phonetic? *Zhongguo Yuwen*, 146(3), 172-177.

Appendix A: The VP control model

To demonstrate the differential effects produced by the VP and VSP models were not simply due to a capacity or training difference, we trained an additional VP model to serve as a control model. This VP control model had an equal number of units to that of the VSP model. It also received the same amount of training presentations as did the VSP model. Figure A.1 shows the architecture of the VP control model. It consisted of five layers of units: 800 visual inputs, 60 hidden units, 300 orthographic hidden units, another 300 hidden units and 105 phonological units. Every unit in one layer was fully connected to the next layer. All the training and testing procedures were otherwise identical to that descried in Simulation section 3.1.

The training was halted after 20 million presentations, and at this point the VP control model could correctly pronounce 99.5% of characters. We then examined whether the VP control model could produce both the effect of frequency and consistency (Lee et al. 2005) and the effect of orthographic neighborhood size and consistency obtained in the present behavioral experiment (Section 2.2). Importantly, we aimed to test whether the patterns produced by the VP control model were similar to those produced by the VP model.

For the effects of frequency and consistency, the VP control model was tested on the all the stimuli taken from Lee et al. (2005, experiment 1). Linear mixed effects models were applied to analyze model performance with both item and simulation number (one to ten) as random factors and error score as a dependent variable. Error items and outliers (greater than two and a half standard deviations from the mean) were removed, resulting in a removal of 5.75% of items in total. The statistical analysis results showed that the frequency effect was significant: adding frequency as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of character type, $\chi^2(1) = 52.36$, $p < .001$. The effect of character type was also significant: adding character type as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of frequency, $\chi^2(2) = 6.71$, $p < .05$. There interaction between frequency and character type was marginally significant, where the model fit was improved by adding the interaction between frequency and character type, $\chi^2(2) = 5.07$, $p = .079$, when both the random and main effects were considered. As can be seen in Figure A.2, the resulting pattern was similar to that of the VP and the VSP model in Figure 4.

For the effect of orthographic neighborhood size and consistency, the VP control model was tested on the same set of stimuli as those in the behavioral experiment. Again, linear mixed effects models were applied to analyze model performance with both item and simulation number (one to ten) as random factors and error score as the dependent variable. Error items and outliers (greater than two and a half standard deviations from the mean) were removed prior to analyses, resulting in a removal of 3.58% of items in total. The statistical analysis results showed that the consistency effect was significant: adding consistency as a fixed factor improved model fit compared to a model with random effects of simulation and item and a fixed effect of neighborhood size, $\chi^2(1) = 11.43$, $p < .01$, but the effect of neighborhood size was not: adding neighborhood size as a fixed factor did not result in a significant improvement in model fit compared to a model with random effects of simulation and item and a fixed effect of consistency, $\chi^2(1) = 0.18$, $p > .05$. The interaction between consistency and neighborhood size was also not significant: the model fit was not significantly improved by adding the interaction term between consistency and neighborhood size, $\chi^2(1) = 0.008$, $p > .05$, when both the random and main effects were included. As can be seen in Figure A.3, the interaction pattern produced by the VP control model was similar to that of the VP model (see Figure 5). However, the interaction pattern was different from either that of the VSP model or that of participants (see Figure 5). In sum, these results provided strong evidence that the difference between the VP and VSP models was attributable to the different architectures of the two models and could not be attributed to capacity limitation or differing lengths of training.

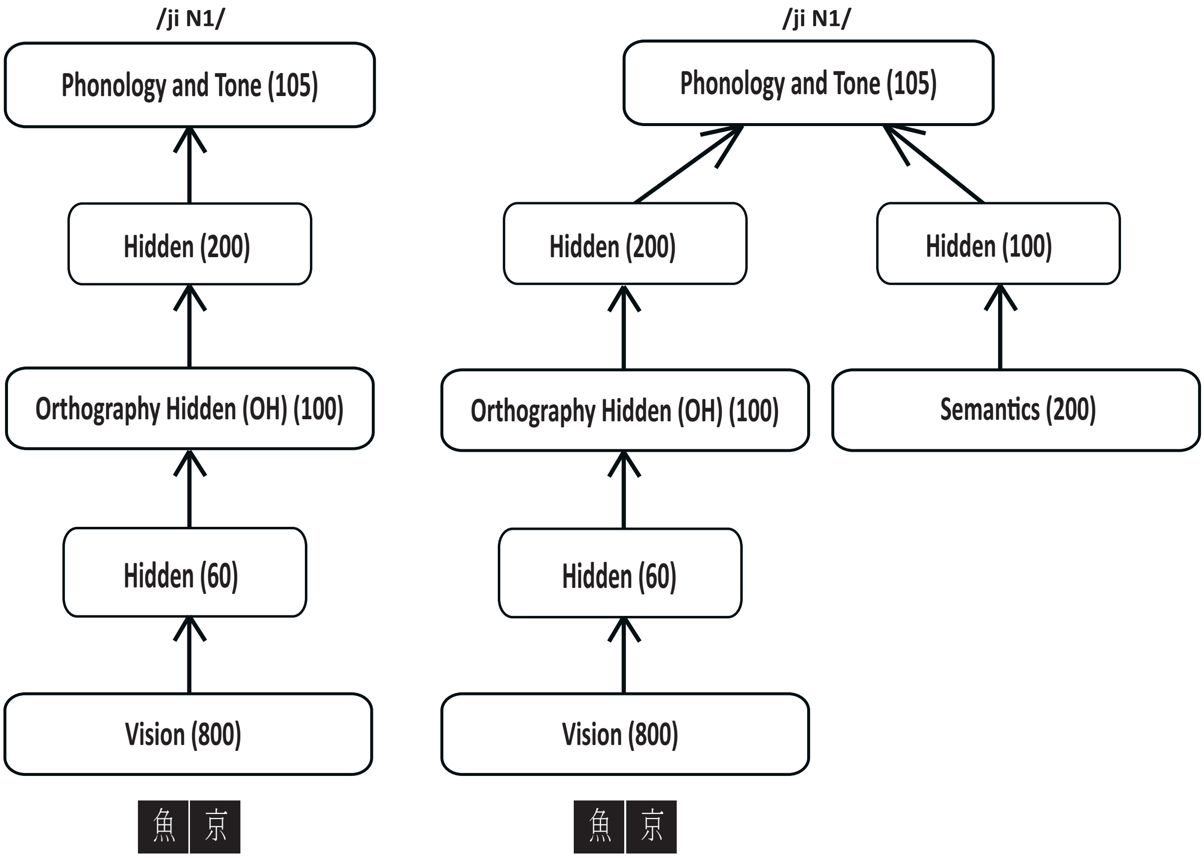

Figure 1. The architectures of the VP model (left panel) and the VSP model (right panel).

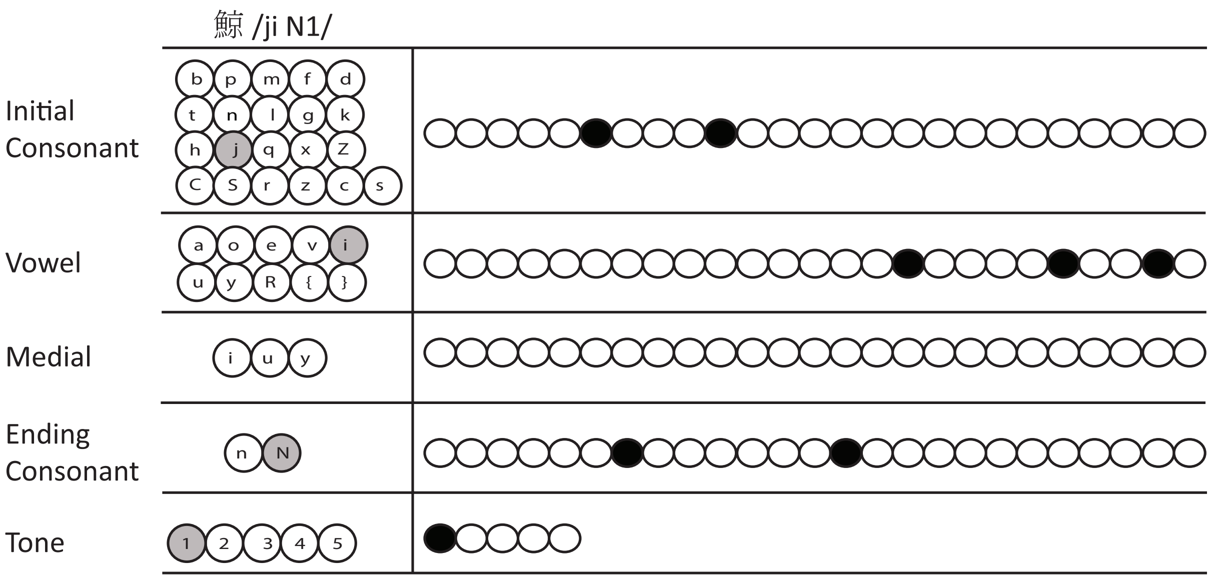

Figure 2. The phonological representation scheme based on Chinese phonetic features.

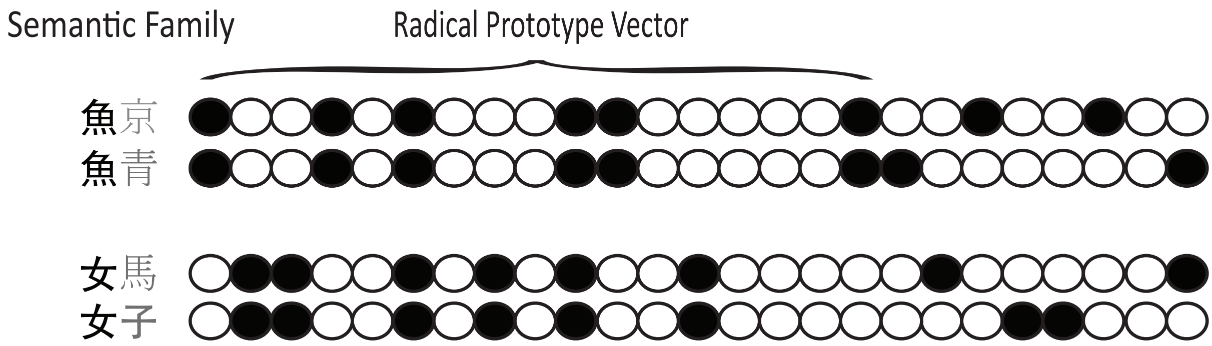

Figure 3. Illustrations of semantic representations used in the VSP model.

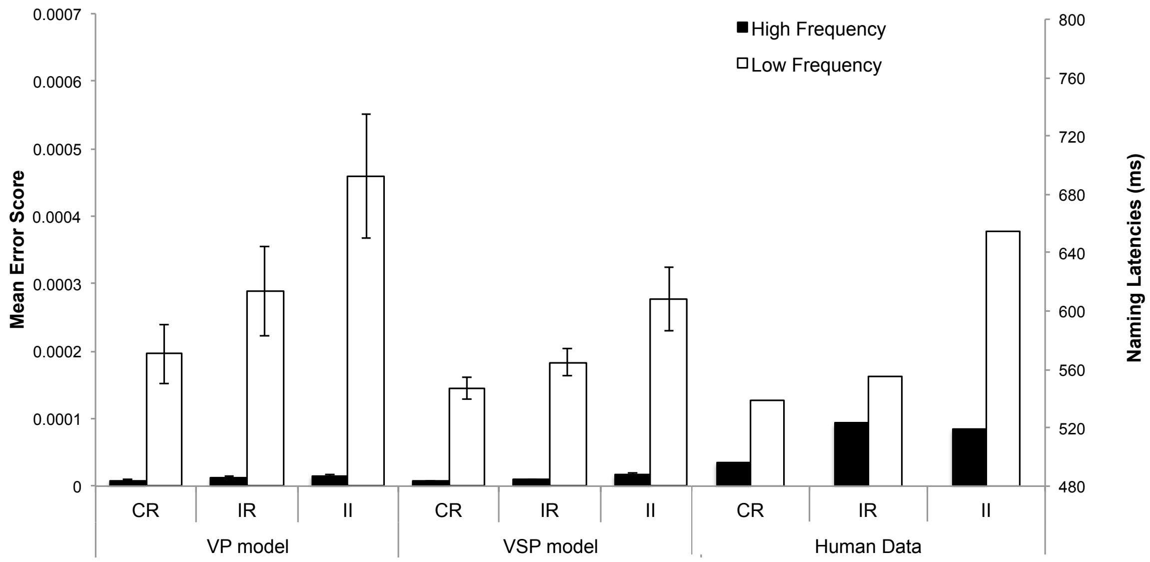

Figure 4. Mean error scores produced by the VP model and the VSP model for high and low frequency characters with different levels of character type along with the behavioral data reproduced from Lee et al. (2005; experiment 1). Error bar represents ±1 standard error. CR: Consistent Regular; IR: Inconsistent Regular; II: Inconsistent Irregular.

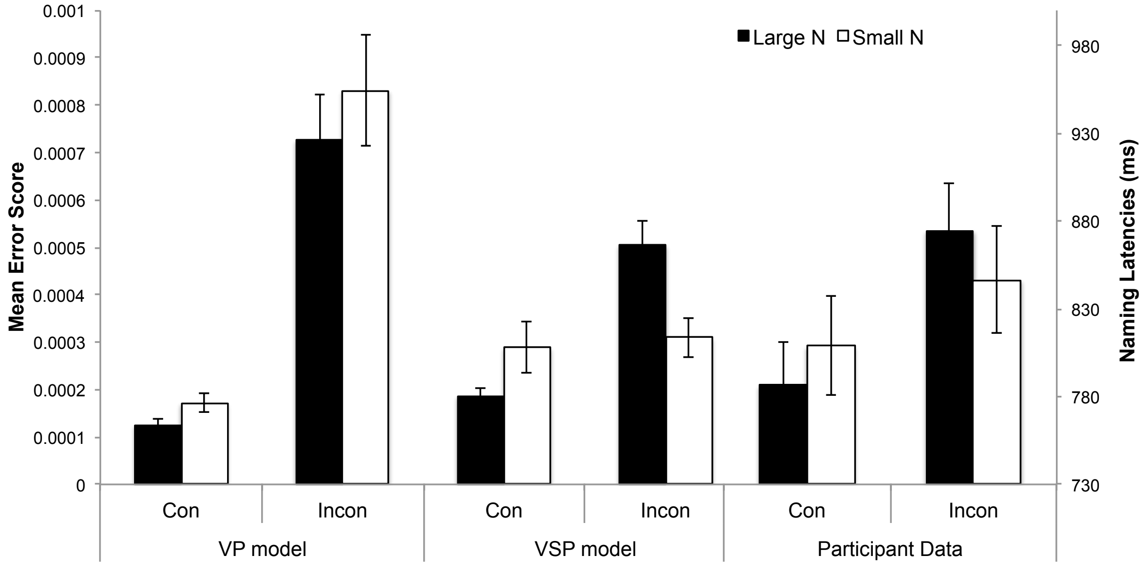

Figure 5. Mean error scores produced by the VP model and the VSP model for characters with large and small neighborhood as a function of consistency along with the participants' data. Error bar represents ±1 standard error. Con: consistent; Incon: inconsistent; Large N: large neighborhood; Small N: small neighborhood.

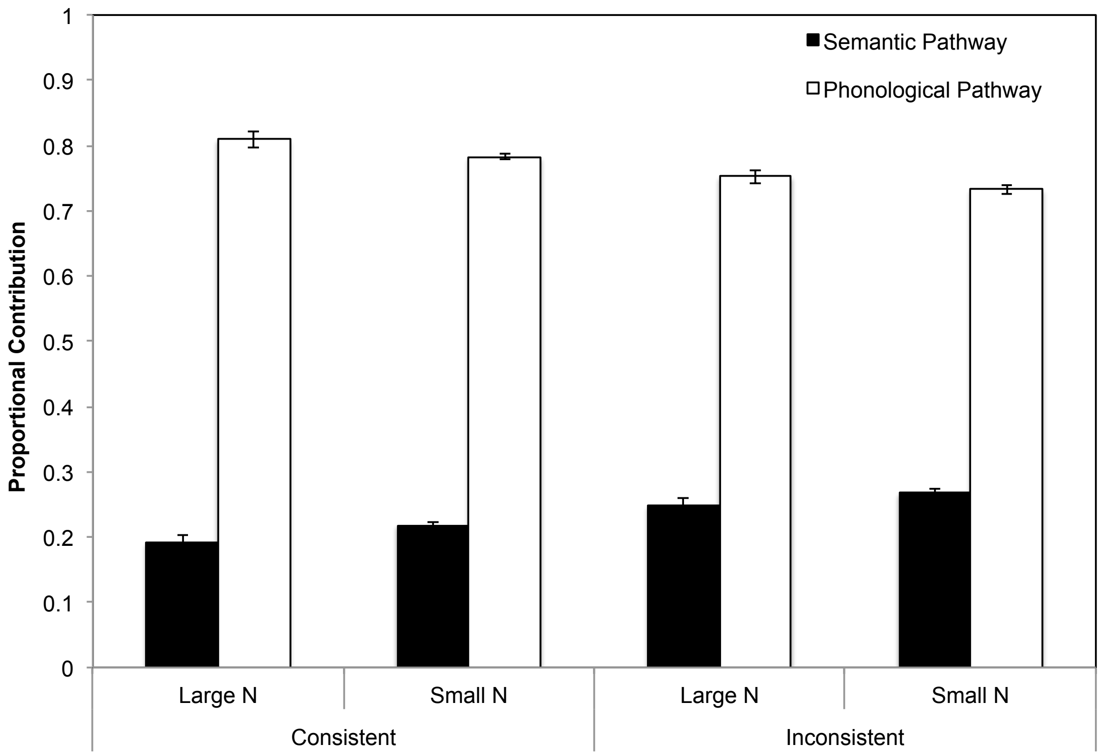

Figure 6. Division of labor in the computation of phonology. Effects of consistency and neighborhood size on the phonological pathway and the semantic pathway in the VSP model. Error bar represents ±1 standard error. Large N: large neighborhood; Small N: small neighborhood.

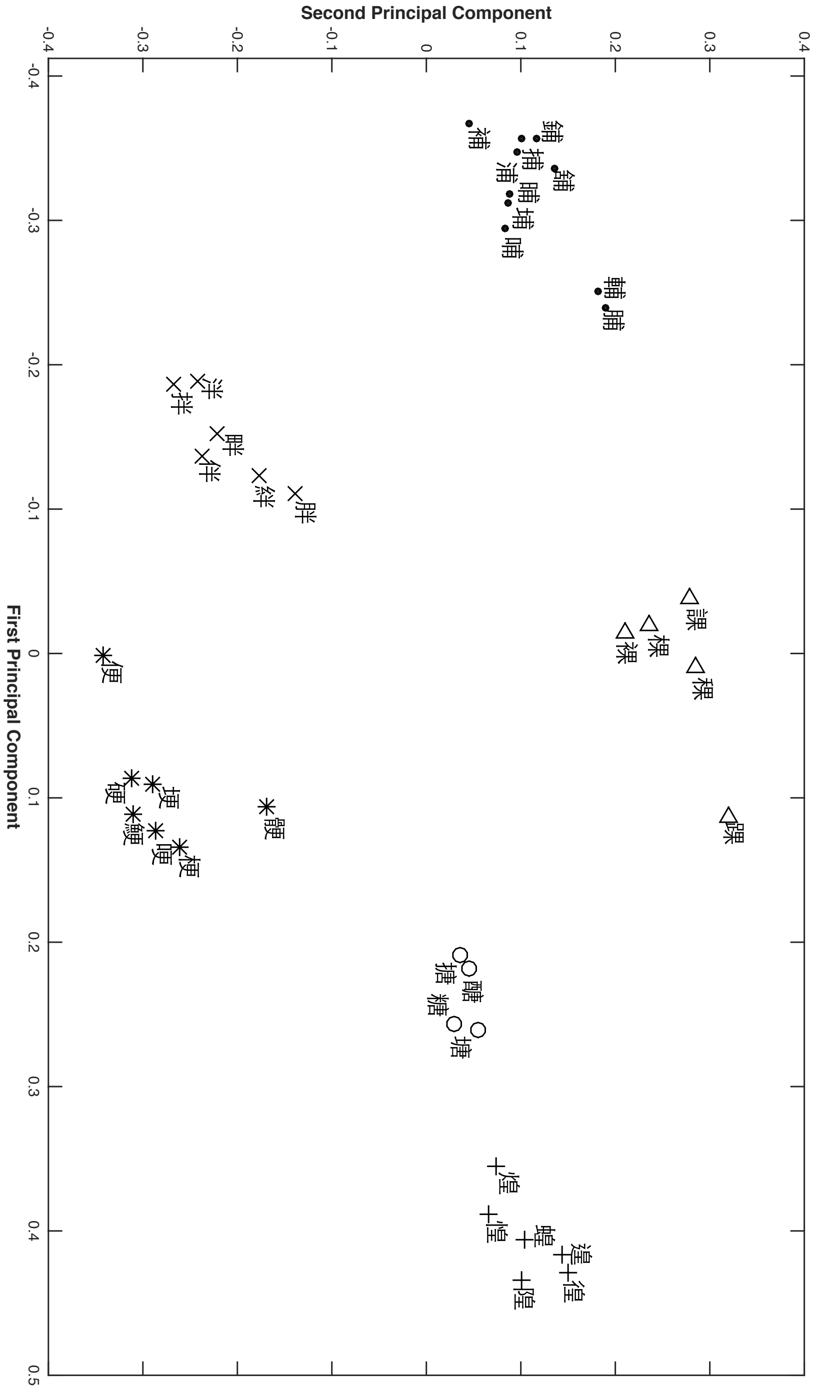

Figure 7. The internal representations of six representative phonetic radical families (皇 /huang2/, 半 /ban4/, 唐 /tang2/, 更 /qeng1/, 果 /guo3/, 甫 /fu3/).

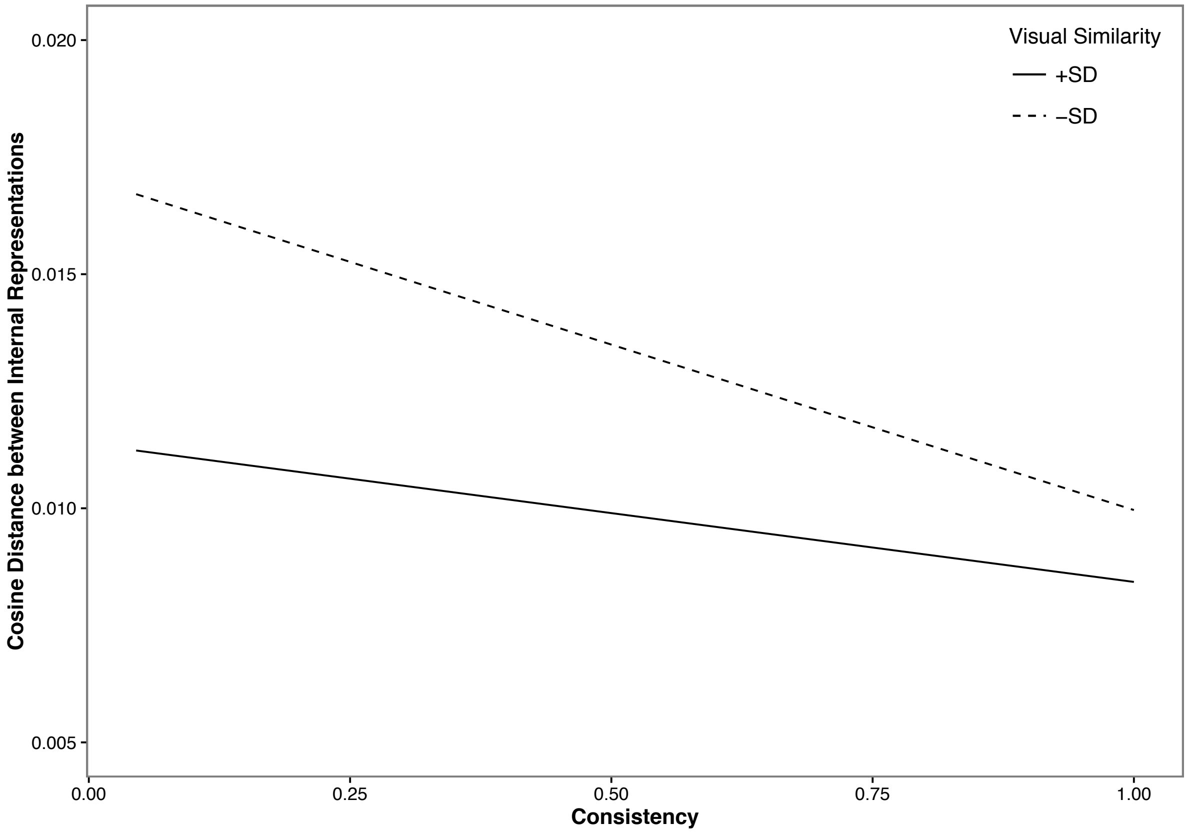

Figure 8. Cosine distance between internal representations for all phonetic radical families as a function of consistency and visual similarity.

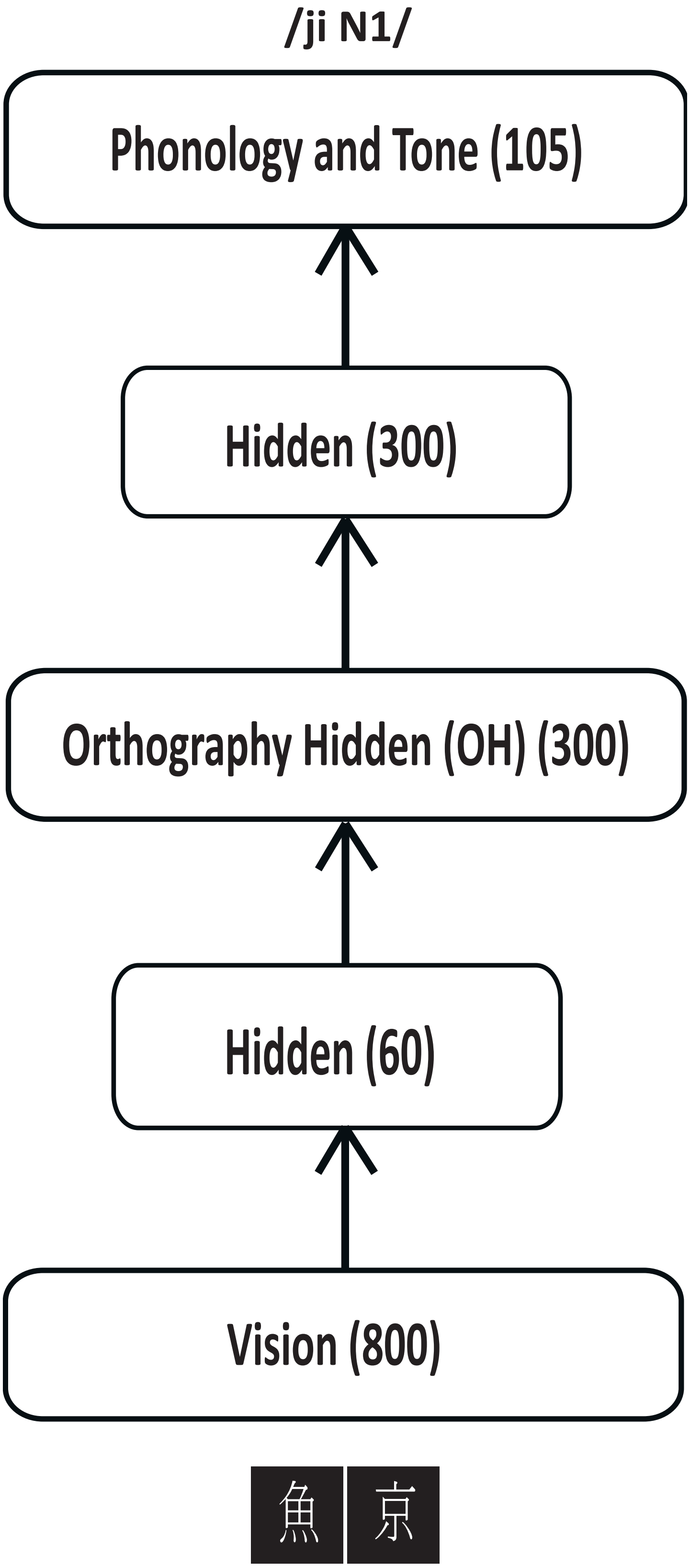

Figure A1. The architecture of the VP control model.

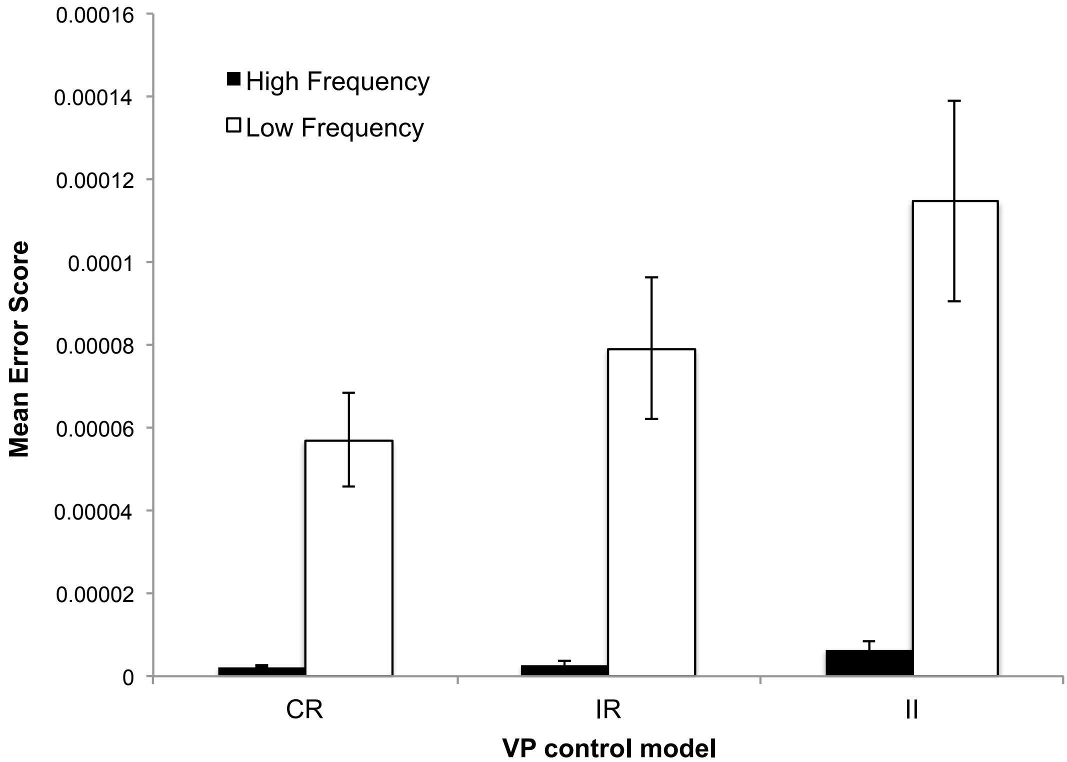

Figure A.2. Mean error scores produced by the VP control model for high and low frequency characters with different levels of character type. Error bar represents ±1 standard error. CR: Consistent Regular; IR: Inconsistent Regular; II: Inconsistent Irregular.

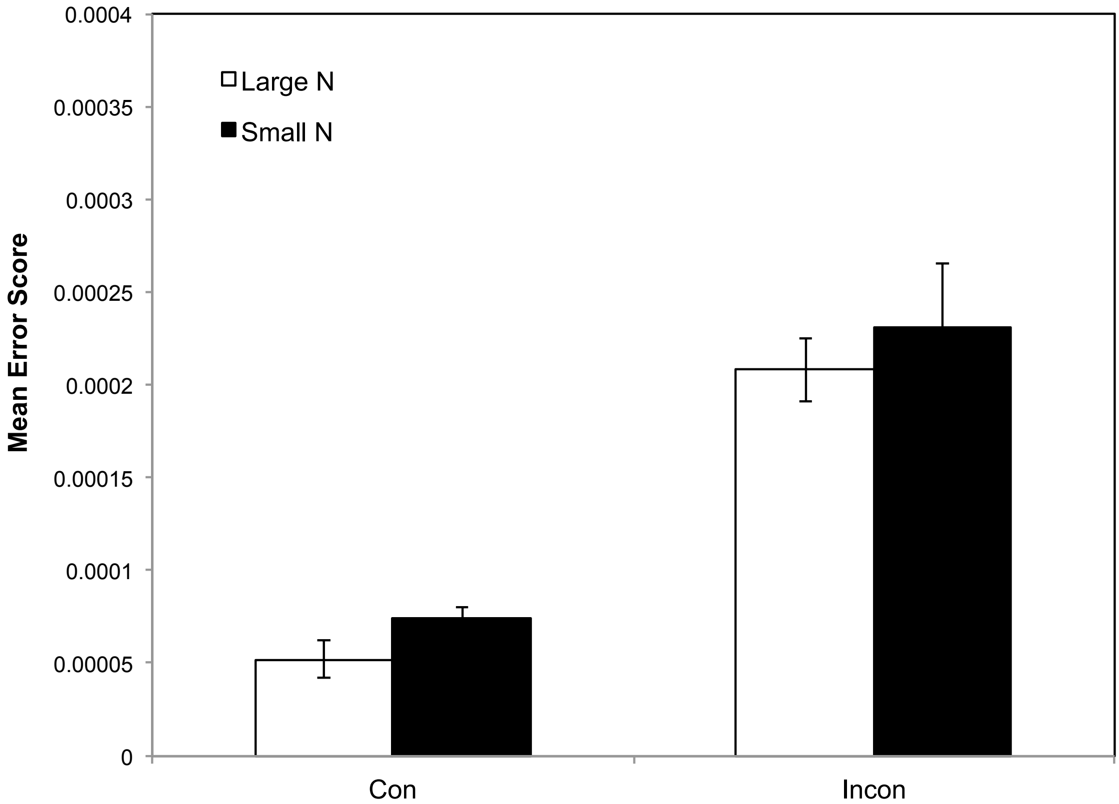

Figure A.3. Mean error scores produced by the VP control model for characters with large and small neighborhood as a function of consistency. Error bar represents ±1 standard error. Con: consistent; Incon: inconsistent; Large N: large neighborhood; Small N: small neighborhood.