Title

Neighborhood in Decay: Working Memory Modulates Effect of PhonologicalSimilarity on Lexical Access

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Neergaar, Karl David
Britton, James
Huang, Chu-Ren

Publication Date

2019

Peer reviewed

Neighborhood in Decay: Working Memory Modulates Effect of Phonological Similarity on Lexical Access

Karl David Neergaard (karl.neergaard@connect.polyu.hk)
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University

James Britton (james.britton@connect.polyu.hk)
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University

Chu-Ren Huang (churen.huang@polyu.edu.hk)
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University

Abstract

A mainstay of models that account for the access of lexical knowledge is that auditory words compete for selection based on form similarity, commonly seen in an inhibitory effect to greater phonological neighborhood density (PND). PND is a metric that states that two words are neighbors if they differ by the addition, deletion or substitution of a single phoneme. A drawback to this account is that there is competing evidence even among the European languages investigated thus far. We sought to verify whether the inhibitory effect of greater PND would hold for Mandarin Chinese in two auditory word repetition tasks with monosyllabic and disyllabic Mandarin words. Results of Experiment 1 showed a facilitative effect to greater PND. Experiment 2 added a non-verbal distractor task to lessen the putative effect of working memory load during the task. The facilitative effect to greater PND was confirmed along with a significant post-hoc interaction with memory decay, operationalized as the duration spent on the distractor tasks. The facilitative effects extend previous reports of differential behavior due to linguistic typology.

Keywords: Lexical access; phonological neighborhood density; memory decay; Mandarin Chinese

Introduction

Essential to the current models of lexical processing is that target words interact during selection with items in long-term memory based on their shared semantic, orthographic, and phonological similarity. Both orthographic and phonological similarity are most commonly calculated through the addition, deletion or substitution of a single letter or phoneme (Landauer & Streeter, 1973). According to this metric, known as neighborhood density, a target stimulus with many similar words in the lexicon, i.e., neighbors, resides in a dense neighborhood, while a word with few similar words in the lexicon resides in a sparse neighborhood. The contrasting of dense and sparse words has been used to model the structural organization of lexical knowledge. For example, in the recognition of orthography, according to both the orthographic and phonological metrics, words from dense neighborhoods have been shown to facilitate recognition (Orthographic; e.g., Coltheart, Davelaar, Jonasson, & Besnar, 1977; Phonological: e.g., Yates, Locker, & Simpson, 2004). This facilitation has motivated the claim that greater density results in greater overall activation, a defining feature that was later implemented in several computational models (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Grainger & Jacobs, 1996; Wagenmakers et al., 2004). However, when tasks are performed in the auditory modality, greater density has been reported to inhibit recognition (e.g., Luce & Pisoni, 1998; Vitevitch & Luce, 1998; Ziegler et al., 2003), motivating the construction of modality specific models of speech processing (Luce, Goldinger, Auer, & Vitevitch, 2000; McClelland & Elman, 1986; Norris, 1994). In response, Chen & Mirman (2012) constructed a connectionist model in an attempt to give a unified account of both facilitative and inhibitory neighborhood effects across both visual and auditory modalities, perception and production tasks, and orthographic, phonological and semantic neighborhood interactions. Their innovative approach, unfortunately, rests on the false assumption that there is a consensus in the literature for modality and task specific neighborhood effects.

To limit the discussion, we will consider only the effects known for phonological neighborhood density (PND). Two hypotheses of note have been advanced concerning differences in polarity from the body of behavioral evidence: psychotypology, and methodology.

The psychotypological argument holds that cognitive processes differ due to the linguistic differences between the languages being tested. The case, as it regards lexical access (Vitevitch & Rodríguez, 2004; Vitevitch & Stamer, 2006, 2009), was made based on evidence from both auditory recognition, and speech production. Dense phonological neighborhoods were inhibitory to speech recognition for English speakers (e.g., Luce & Pisoni, 1998), yet were facilitative for Spanish speakers (Vitevitch & Rodríguez, 2004), whereas, dense words were facilitative for English speakers in picture naming (Vitevitch, 2002), yet inhibitory for Spanish speakers (Vitevitch & Stamer, 2006, 2009). Vitevitch and colleagues speculated that the differences between the English and Spanish lexicons led to differences in polarity. Whereas English words have on average a greater number of shorter words with more phonological neighbors, the sparser Spanish vocabulary features words that are neighbors both phonologically and semantically, e.g., *niño/niña* (boy/girl).

The methodological argument to account for differences in polarity points towards the design and methods employed in PND related studies. Sadat, Martin, Costa, and Alario, (2014) posited that the contradictory findings could be amended through testing with larger stimulus sets and the use of mixed effects models. The re-analyses of PND studies done by Sadat and colleagues clarified important differences between F tests and mixed effects models in the analysis of a variable that is continuous in nature and thus best fit for regression rather than factorial designs.

A third possibility to account for the differences in polarity in PND studies is to investigate working memory, specifically as it concerns the size of the stimuli sets used. As a participant recognizes or names a word, that lexical item is temporarily stored in working memory. If participants are exposed to multiple words, memory load increases and with it reaction times (Cohen et al., 1997; Jha & McCarthy, 2000). If the participant sustains attention on one task then memory decay does not happen at the same rate compared to if they were given a pause or a distractor task that does not interfere with the domain or modality of the main task (Rae & Perfect, 2014). In the case of phonological information, Baddeley (1986) found that phonological memory decayed within roughly 2 seconds. This does not differ greatly from the recall of orthographic letters after doing simple math problems (Brown, 1958; Peterson & Peterson, 1959). Given that during PND related tasks the inter-stimulus pause tends to be between 500-1500ms, i.e., under the rate of decay known to exist for phonological information, it is feasible that words are subject to cumulative activation, i.e., that activations of multiple word representations overlap and contribute to participant performance.

The studies that have investigated neighborhood effects amongst adults have utilized a large range of different sized stimulus sets. In studies implementing auditory word repetition tasks the story is quite straightforward, wherein large stimulus sets (Luce & Pisoni, 1998: 400 words; Vitevitch & Luce, 1998: 240 words) led to an inhibitory PND effect. In lexical decision tasks, two experiments using large stimulus sets showed inhibitory PND effects (Luce & Pisoni, 1998: 610 words; Vitevitch, Stamer, & Sereno, 2008: 112 words) while one with a small set of stimuli showed a facilitative PND effect (Vitevitch & Rodríguez, 2004: 80 words). The picture naming literature is where we see inhibitory results with large and small stimulus sets (e.g., Sadat, et al., 2014: 533 pictures; Vitevitch & Stamer, 2006: 48 pictures), facilitative results with small stimulus sets (Baus, Costa, & Carreiras, 2008: 48 pictures; Marian, Blumenfeld, & Boukrina, 2008: 57 pictures; Pérez, 2007: 89 pictures; Vitevitch, 2002: 48 pictures; Vitevitch & Stamer, 2009: 48 pictures), and non-significant PND effects (Jescheniak & Levelt, 1994: 96 pictures repeated 3 times; Vitevitch et al., 2004: 44 pictures). Note that non-significant results might also have been due to issues unrelated to stimuli number, such as mixing photographic and hand-drawn stimuli or due to naming pictures that represent conceptual processes such as verbs (Newman & Bernstein Ratner, 2007; Tabak, Schreuder, & Baayen, 2010).

The role of PND in working memory has not been fully explored. The only studies to test their interaction found a facilitation of greater density in serial recall tasks with English speakers (Oberauer, 2009; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002). The facilitative effect was said to be due to redintegration, which can be described as the restoration of short-term memory traces due to long-term memory representations. In order to test cumulative activation, however, it is necessary to account for overall memory load rather than that of isolated words.

To test the possibility that cumulative working memory influences the directional effect of PND, we performed two experiments with a large number of stimuli. In Experiment 1 we presented the full stimuli set to our participants without sufficient time for decay in memory load. In Experiment 2 we inserted three nonverbal distractor tasks in order to introduce memory decay. In both experiments we implemented the auditory word repetition task due to it being the only task thus far without contradictory findings. The cumulative memory hypotheses, allows for the prediction of an inhibitory effect to greater PND in Experiment 1 and a facilitative or null effect in Experiment 2.

In the current study we also incorporate concerns brought by both previous hypotheses concerning methodology and psychotypology on differing PND effects. Through the use of a large stimuli set and mixed effects models we treated PND as a continuous variable. Meanwhile, our target language, Mandarin Chinese, was chosen due to its typological distinctness to either English or Spanish, allowing for a unique view on how the dimensions of the lexicon affects lexical access.

The Mandarin vocabulary differs from both English and Spanish in critical ways. Its syllable inventory, when including lexical tone, has roughly 1,300 items, which is a number far less than the 10,000+ English syllables. Unlike Spanish that boasts of a large proportion of multisyllabic words rich in morphological variation (Arbesman, Strogatz, & Vitevitch, 2010), roughly 72% of Mandarin's phonological words (i.e., in which all homophones are collapsed to one item) are disyllabic, and only 3.8% monosyllabic (Neergaard & Huang, 2019). Meanwhile, Mandarin words have little to no morphology. For instance, unlike Spanish, Mandarin verbs do not conjugate, and nouns do not note gender nor number.

To date, no results have been reported in the auditory word repetition task with Mandarin speakers. Despite this lack of prior evidence, the psychotypology account allows for certain predictions. Given the greater distance lexically from Spanish, particularly in relation to its on average longer word length and inflectional morphology, Mandarin shares greater similarity with English. English has shown inhibitory PND effects in both auditory lexical decision and word repetition, making it likely that lexical competition best accounts for the selection of dense phonological words in Mandarin.

Experiment 1

Methods

Participants Thirty-three native-Mandarin speakers participated in this experiment (Female: 21; Ages 18-35, M: 12, SD: 3.64). None of the participants reported speech, hearing, or visual disorders. All participants reported native-level proficiency in Mandarin.

The current study design was approved by The Hong Kong Polytechnic University's Human Subjects Ethics Sub-committee (reference number: HSEARS20140908002). The participants gave their informed consent and were compensated with 50HKD for their participation.

Stimuli The auditory stimuli for this experiment consisted of 154 Mandarin words (10 practice; 144 test). A female native-speaker of Mandarin from Fujian province produced all of the stimuli by speaking at a normal speaking rate into a high-quality microphone. Stimuli fell into 4 categories according to their syllable or segment length: 36 3-segment monosyllables with a CVN syllable structure (e.g., *san1*); 36 4-segment monosyllables with a CGVN syllable structure (e.g., *bian3*); 36 3-segment disyllables with a CV V syllable structure (e.g., *da4 yi1*); 36 4-segment disyllables with a CV CV syllable structure (e.g., *li4 shi3*). The 144 test stimuli were made from 20 syllable onsets, whose distributions were not significantly different in syllable length (p=1) or segment length (p=1). Eleven stimuli sets were constructed, each where stimuli were pseudo-randomized such that there were no consecutive presentations of items with the same onset or lexical tone (first syllables for disyllabic words). The stimuli list can be seen in Appendix 1.

Because the current stimuli consist of monosyllables and disyllables of both 3 and 4 segments in length, it was not possible to control for their durations along all 4 dimensions. Instead, stimuli were chosen in order to minimize durational differences between 3-Segment words (CV V, M: 609.25; SD: 11.59; CVN, M: 609.00; SD: 11.01) and between 4-Segment words (CVCV, M: 784.67; SD: 9.25; CVVN, M: 784.17; SD: 11.02). Stimuli did not differ within their respective segment length groups, but were significantly different across segment lengths (F=9433, p<0.001). Thus, while a significant difference in reaction times is expected between 3- and 4-Segment words, the same cannot be said between monosyllable and disyllables belonging to their respective segment lengths, which is critical in identifying whether monosyllables and disyllables are processed in an equivalent manner.

Lexical statistics for the stimuli were taken from the Database of Mandarin Neighborhood Statistics (Neergaard, Xu, & Huang, 2016), in which lexical frequency is derived from the wordlist of Subtlex-CH (Cai & Brysbaert, 2010) according to the summed subtitle frequency for each phonological word. All relevant statistics were calculated from 30,000 phonological words. In order to test the hypothesis that words of varying unit sizes are subject to the effect of phonological similarity during speech processing, it was necessary to use the fully segmented Mandarin syllable schema (C_G_V_X_T) because it allowed us to control for both segment and syllable length while distinguishing between words according to lexical tone. Stimuli did not differ in log10 lexical frequency for either 3-segment words (CVN, M: 3.08, SD: 0.40; CV V, M: 2.75, SD: 0.46) or 4-segment words (CGVN, M: 2.98, SD: 0.49; CV CV, M: 3.33; SD: 0.21) according to both segment length (p=0.869) and syllable length (p=0.981).

The remaining variables are of the density variety and include PND, log10 neighborhood frequency (NF, M: 3.11; SD: 1.06), and homophone density (HD, M: 1.67; SD: 1.26).

The goal in choosing stimuli according to PND, knowing that greater length negatively correlates with higher density, was to assure that there was sufficient spread for each group. For the syllable length group, disyllabic words had a spread of 0-11 neighbors (M: 3.71, SD: 2.45), while monosyllabic words had a spread of 4-25 neighbors (M: 13.29, SD: 5.05). For the segment length group, 3-segment words had a spread of 0-25 neighbors (M: 10.10, SD: 7.03), while 4-segment words had a spread of 0-17 neighbors (M: 6.90, SD: 4.85).

Procedure Participants sat in a quiet room in front of a computer running E-Prime 2.0 (Psychology Software Tools, 2012). They were instructed to repeat the words they heard over headphones into an attached microphone as fast as possible. Each trial began with a cross '+', in the center of the screen for 1000ms. Next, the onset of the target audio was presented concurrent with the exposure of a blank screen. A PST Serial Response Box was activated by the participants' voice, dependent on their response, which then led to a pause of 1000ms and the end of a trial. Stimuli were pseudo-randomized such that no two items were presented sequentially with the same onset or lexical tone. The entire experiment lasted roughly 10 minutes. Participants were given a practice set of 10 words prior to beginning the experiment.

Results and discussion

Reaction times were measured offline using SayWhen (Jansen & Watter, 2008). No participants were excluded due to excessive error rates, or deviant reaction times. Three stimuli were removed for error rates higher than 25% (*guang4, qing3, san4*). From the new total of 4,653 trials, 102 were removed due to production errors, accounting for 2.19%. A further 238 trials (5.23%) were removed for values below the duration of our shortest stimuli (577ms), and for values 2.5 standard deviations above the group mean. The final number of trials to be analyzed were 4,313 (M: 917ms; SD: 144ms).

As can be seen in Table 1, Subject and Item were placed in the random effects, while each of the density variables (PND, HD, and NF) were analyzed according to the two levels of segment length (SegLen: 3-seg, 4-seg). We also place syllable length (SyLen) into the fixed effects structure to evaluate whether there was a processing cost despite stimulus durations not being different between monosyllables and disyllables in each segment group.

Table 1. Model estimates for Experiment 1

Random effects	Var.	SD			
Subject	0.003	0.057			
Item	0.012	0.111			
Residual	0.007	0.086			
Fixed effects	β	SE	df	t	p
Intercept	9.4^{e-1}	2.2^{e-2}	54	42.21	< 0.001
SyLen (di)	-5.2^{e-2}	1.9^{e-2}	132	-2.73	0.007
3-seg:PND	-3.3^{e-2}	9.4^{e-3}	132	-3.48	< 0.001
4-seg:PND	-5.2^{e-2}	1.4^{e-2}	132	-3.63	< 0.001
3-seg:HD	6.5^{e-3}	6.1^{e-3}	132	1.07	0.287
4-seg:HD	7.0^{e-4}	1.1^{e-2}	132	0.06	0.950
3-seg:NF	-1.6^{e-2}	1.1^{e-2}	132	-1.51	0.133
4-seg:NF	-7.8^{e-3}	1.1^{e-2}	132	-0.70	0.483

Results revealed that monosyllables (M: 903ms; SD: 15ms) were produced significantly faster than disyllables (M: 928ms; SD: 14ms). Greater PND was facilitative for both 3-segment (M: 867ms; SD: 138ms) and 4-segment (M: 965ms; SD: 137ms) items. No effects were found for either SegLen:HD or SegLen:NF. An estimate of r^2, using the 'r2glmm' package in R (Jaeger, Edwards, Das, & Sen, 2016), revealed that the model had a marginal r^2 of 0.224, and semi-partial r^2 of 0.145 for SegLen:PND, and 0.053 for SyLen.

The facilitative effect to greater PND for both monosyllabic and disyllabic words, rather than supporting a cumulative memory account, is suggestive that typological differences between English (majority inhibitory findings to greater PND) and Mandarin led to the differential performance.

Experiment 2

The premise of the cumulative memory account is that shorter stimulus sets in a naming task result in facilitative PND effects due to there being fewer lexical items stored in working memory when compared to a task with a large stimulus set. It is possible that working memory builds cumulatively leading to increased activation, but that due to the particular psychotypological features of Mandarin, a facilitative effect to greater PND is the outcome. The only way to verify the status of a facilitative effect, while also nullifying the cumulative account, is to provide participants with sufficient time for memory decay during naming.

Methods

Participants Forty-seven native-Mandarin speakers participated in this experiment (Female: 29; Ages 19-38, M: 24, SD: 4). None of the participants reported speech, hearing, or visual disorders.

Stimuli The same auditory stimuli from Experiment 1 were used in this experiment.

Procedure The current design differed from Experiment 1 in that the experiment was partitioned into 4 blocks of 36 trials each with three interleaved distractor tasks. Each distractor task included four basic math questions: e.g., "20*2=__". The distractor task was self-paced. Participants had to press a button to return to the following test block. The entire experiment took less than 15 minutes.

Results and discussion

Reaction times were again measured offline using SayWhen (Jansen & Watter, 2008). Three participants were excluded from the analysis; two for reaction times 2.5 standard deviations above the group mean, and one due to experimenter error in data acquisition. No participants were excluded due to excessive error rates; however, three stimuli were removed for error rates higher than 25% (*qing3, san4, sang1*). From the new total of 6,203 trials, 142 were removed due to production errors, accounting for 2.24%. A further 103 trials (1.66%) were removed for values below 577ms and above 1446ms, leaving our final number of trials to be analyzed at 6,100 (M: 1010ms; SD: 148ms).

The same model configuration from Experiment 1, as shown in Table 2, again found a significant SyLen effect between monosyllables (M: 995ms; SD: 148ms) and disyllables (M: 1027ms; SD: 148ms), and a significant facilitative effect to greater PND for both 3-segment (M: 951ms; SD: 139ms) and 4-segment (M: 1069ms; SD: 134ms) items, with no significant effects for SegLen:HD or SegLen:NF. The model's marginal r^2 was 0.202, with a semi-partial r^2 of 0.121 for SegLen:PND, and 0.042 for SyLen.

Table 2. Model estimates for Experiment 2

Random effects	Var.	SD			
Subject	0.004	0.065			
Item	0.009	0.094			
Residual	0.008	0.091			
Fixed effects	β	SE	df	t	p
Intercept	1.04	1.9^{e-2}	115	55.22	< 0.001
SyLen (di)	-5.1^{e-2}	2.1^{e-2}	133	-2.41	0.017
3-seg:PND	-3.4^{e-2}	1.1^{e-2}	133	-3.19	0.002
4-seg:PND	-5.2^{e-2}	1.6^{e-2}	133	-3.25	0.001
3-seg:HD	6.1^{e-4}	7.0^{e-3}	133	0.09	0.931
4-seg:HD	-5.8^{e-3}	1.2^{e-2}	133	-0.46	0.643
3-seg:NF	-1.9^{e-2}	1.2^{e-2}	133	-1.57	0.118
4-seg:NF	-5.4^{e-3}	1.3^{e-2}	133	-0.43	0.667

In this experiment we confirmed the facilitative effect to greater PND for Mandarin. We have also shown that stimulus set sizes are not the likely candidates in the variability found in PND studies. We did not however account for how PND and working memory interact.

To investigate whether decay modulates the effect of PND, in a post-hoc analysis we operationalized memory decay as the time spent on the three interleaved distractor tasks. While each participant received the same basic math questions, they were given as much time as they saw fit to complete each task before returning to the repetition task. For the following analysis, it was necessary to exclude the trials belonging to the experiment's first block. In this way, each block under examination entailed auditory lexical processing after having received a limited time for memory decay from a previous session of auditory lexical processing.

The values for Decay ranged from as short as 3 seconds (3100ms) to as long as 41 seconds (41373ms). Visual inspection of Decay's token values revealed that it was not linearly distributed. We rescaled the variable using a Box Cox transformation (Tukey, 1977) to evenly distribute duration length of non-lexical processing during the distractor task.

Using the 'mcgv' package in R (Wood, Scheipl, & Faraway, 2013), a generalized additive model using tensor product smooths was constructed (with Subject and Item as random effects) in which Decay was added as an interaction to each level of PND, SegLen, and SyLen. With an adjusted r^2 of 0.651, Decay interacted significantly with PND ($F=23.94$; $p<0.001$); 3-segment ($F=28.75$; $p<0.001$) and 4-segment items ($F=15.62$; $p<0.001$); and both monosyllables ($F=6.32$; $p<0.001$) and disyllables ($F=8.28$; $p<0.001$). As can be seen in Figure 1, when Decay was shortest, the effect of PND was strongest, providing clear evidence that working memory is a determining factor of phonological neighborhood effects.

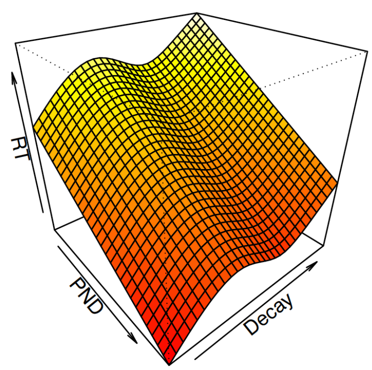

Figure 1: Interaction effect of phonological neighborhood density (PND) and time spent on the distractor task (Decay)

General Discussion

The purpose of the current study was to examine multiple hypotheses on the directional effect of PND through Mandarin Chinese, a language typologically distinct from the languages tested to date. We incorporated the methodological concerns brought by Sadat et al. (2014), through the testing of a large number of stimuli with mixed effects models wherein PND was treated as a continuous variable. We performed two auditory word repetition tasks, the only task to date that has not shown contradictory PND effects, in which our participants' rate of memory decay was manipulated to test whether differences in PND polarity have been due to cumulative memory. Finally, the testing of Mandarin participants allowed us to join the debate on psychotypology, i.e., whether the dimensions of a speaker's lexicon can result in differential behavioral outcomes.

In Experiment 1 we exposed our participants to the full stimuli set under the assumption that by not allowing for memory decay to occur our participants would produce an inhibitory effect to greater PND. Opposite our expectations, and in contrast to the previous English results, we found a facilitative effect to PND. In Experiment 2 we manipulated the task through the introduction of interleaved nonverbal distractor tasks. Changes in modality through distractor tasks have been shown to increase memory decay of the main task material (Rae & Perfect, 2014). The facilitative effect to greater PND was confirmed despite providing our participants with time for memory decay. A further post-hoc analysis illustrated that while working memory indeed modulates the phonological neighborhood effect, it can do so without lexical competition.

Under the assumptions of the psychotypology account of PND (Vitevitch & Rodríguez, 2004; Vitevitch & Stamer, 2006, 2009), we predicted that our Mandarin participants would experience lexical competition due to greater PND, in line with previous English results and contrary to previous Spanish results. This assumption was built on the greater difference between the Spanish and Mandarin vocabularies compared to the differences between the English and Mandarin lexicons. While Spanish is rich in morphology and on average has longer words, Mandarin has on average shorter words and no inflectional morphology. Contrary to our prediction, our Mandarin speakers were facilitated by greater PND, revealing that word length and inflectional morphology are likely not a reason for why Spanish speakers also experienced facilitation by words from dense phonological neighborhoods.

Given our current negation of the cumulative memory account, further work would benefit from delving deeper into the psychotypology of lexical access. Evidence has been mounting for differences in brain areas during language process between English and Mandarin speakers, at the level of whole-brain maps (Wu et al., 2015), and targeting language processing areas during tasks such as rhyming judgments (Brennan, Cao, Pedroarena-Leal, McNorgan, & Booth, 2013). A comparison of the Spanish and Mandarin lexicons might reveal how similarities between typologically distinct languages can lead to outcomes that defy the current models of speech production and perception.

It is also possible that an influence other than phonological neighborhoods is responsible for the facilitative effects in Spanish and Mandarin. The work of Vitevitch and colleagues pointed to a possible candidate other than word length and inflectional morphology, namely, neighbors of target words that are of both phonological and semantic relations (i.e., 'boy/girl, *niño/niña*). Our search of the literature found evidence concerning possible effects of semantic neighbors during auditory lexical decision, but only for English (Goh, Yap, Lau, Ng, & Tan, 2016; Tucker et al., 2018). In line with the current predictions, semantic neighbors did not significantly predict reaction times. In contrast to English, it is possible that both Spanish and Mandarin feature a sufficient number of phono/semantic neighbors to lead to facilitation during an auditory task. While Mandarin does feature phono/semantic word pairs such as *bian1* ('side' 边) / *pian1* ('one-sided' 偏), it also entertains a uniquely high level of homophony (*bian1* = 9 homophones; *pian1* = 3 homophones), making future comparisons between the two languages challenging.

References

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). Comparative analysis of networks of phonologically similar words in English and Spanish. *Entropy*, *12*(3), 327–337. https://doi.org/10.3390/e12030327

Baddeley, A. D. (1986). *Working memory* (Oxford psy). Oxford: Oxford University Press.

Baus, C., Costa, A., & Carreiras, M. (2008). Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes*, *23*(6), 866–888.

Brennan, C., Cao, F., Pedroarena-Leal, N., McNorgan, C., & Booth, J. R. (2013). Reading acquisition reorganizes the phonological awareness network only in alphabetic writing systems. *Human Brain Mapping*, *34*(12), 3354–3368. https://doi.org/10.1002/hbm.22147

Brown, J. (1958). Some tests of the decay theory of immediate memory. *The Quarterly Journal of Experimental Psychology*, *10*, 12–21.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417–430. https://doi.org/10.1037/a0027175

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besnar, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555).

Coltheart, Max, Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud. *Psychological Review*, *108*(1), 204–256.

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M. R., & Tan, L. (2016). Semantic Richness Effects in Spoken Word Recognition : A Lexical Decision and Semantic Categorization Megastudy. *Frontiers in Psychology*, *7*(Article 976), 1–10.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review*, *103*(3), 518–565.

Jaeger, B. C., Edwards, L. J., Das, K., & Sen, P. K. (2016). An R 2 statistic for fixed effects in the generalized linear mixed model. *Journal of Applied Statistics*, *44*(6), 1086–1105.

Jansen, P. A., & Watter, S. (2008). SayWhen: an automated method for high-accuracy speech onset detection. *Behavior Research Methods*, *40*(3), 744–751.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824–843.

Jha, A. P., & McCarthy, G. (2000). The influence of memory load upon delay-interval activity in a working-memory task: an event-related functional MRI study. *Journal of Cognitive Neuroscience*, *12*, 90–105.

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior*, *12*, 119–131.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, *62*(3), 615–625. https://doi.org/10.3758/BF03212113

Luce, P. A., & Pisoni, D. B. (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing*, *19*(1), 1–36.

Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to Phonological Similarity Within and Across Languages. *Journal of Psycholinguisic Research*, *37*(3), 141–170.

McClelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech Perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

Neergaard, K. D., & Huang, C. (2019). Constructing the Mandarin phonological network: novel syllable inventory used to identify schematic segmentation. *Complexity*, (Article 6979830), 1–21.

Neergaard, K. D., Xu, H., & Huang, C.-R. (2016). Database of Mandarin neighborhood statistics. In *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*.

Newman, R. S., & Bernstein Ratner, N. (2007). The Role of Selected Lexical Factors on Confrontation Naming Accuracy, Speed, and Fluency in Adults Who Do and Do Not Stutter. *Journal of Speech, Language, and Hearing Research*, *50*(February), 196–213.

Norris, D. (1994). Shortlist - a Connectionist Model of Continuous Speech Recognition. *Cognition*, *52*(3), 189–234.

Oberauer, K. (2009). Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition*, *37*(3), 346–357.

Pérez, M. A. (2007). Age of acquisition persists as the main factor in picture naming when cumulative word frequency and frequency trajectory are controlled. *The Quarterly Journal of Experimental Psychology*, *60*(1), 32–42. https://doi.org/10.1080/17470210600577423

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*, 193–198.

Rae, P. J. L., & Perfect, T. J. (2014). Visual distraction during word-list retrieval does not consistently disrupt memory. *Frontiers in Psychology*, *5*(April), 362.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019–1034.

Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, *68*, 33.58.

Tabak, W., Schreuder, R., & Baayen, R. H. (2010). Producing in inflected verbs: A picture naming study. *The Mental Lexicon*, *5*(1), 22–46.

Tucker, B. V., Brenner, D., Danielson, D. K., Kelley, M. C., Nenadić, F., & Sims, M. (2018). The Massive Auditory Lexical Decision (MALD) database. *Behavior Research Methods*, 1–18.

Tukey, J. W. (1977). *Exploratory data analysis* (1st ed.). Pearson.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735–747.

Vitevitch, M. S., Armbruster, J., & Chu, S. (2004). Sublexical and lexical representations in speech production:effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 514–529.

Vitevitch, M. S., & Luce, P. A. (1998). When words complete: Levels of processing in perception of spoken words. *Psychological Science*, *9*(4), 325–329.

Vitevitch, M. S., & Rodríguez, E. (2004). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*, *3*(1), 64–73.

Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, *21*(6), 760–770.

Vitevitch, M. S., & Stamer, M. K. (2009). *The influence of neighborhood density (and neighborhood frequency) in Spanish speech production: A follow-up report*.

Vitevitch, M. S., Stamer, M. K., & Sereno, J. A. (2008). Word length and lexical competition: Longer is the same as shorter. *Language and Speech*, *51*(4), 361–383.

Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology*, *48*, 332–367.

Wood, S. N., Scheipl, F., & Faraway, J. J. (2013). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing*, *23*(3), 341–360.

Wu, J., Lu, J., Zhang, H., Zhang, J., Yao, C., Zhuang, D., … Zhou, L. (2015). Direct evidence from intraoperative electrocortical stimulation indicates shared and distinct speech production center between Chinese and English languages. *Human Brain Mapping*, *36*(12), 4972–4985. https://doi.org/10.1002/hbm.22991

Yates, M., Locker, L., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, *11*(3), 452–457. https://doi.org/10.3758/BF03196594

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