COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 39 (2015) 538–558

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12156

Interaction Between Phonological and Semantic Representations: Time Matters

Qi Chen,[a,b] Daniel Mirman[b,c]

[a]*Center for Studies of Psychological Application and School of Psychology, South China Normal University*
[b]*Moss Rehabilitation Research Institute*
[c]*Department of Psychology, Drexel University*

Received 28 August 2012; received in revised form 26 August 2013; accepted 7 January 2014

Abstract

Computational modeling and eye-tracking were used to investigate how phonological and semantic information interact to influence the time course of spoken word recognition. We extended our recent models (Chen & Mirman, 2012; Mirman, Britt, & Chen, 2013) to account for new evidence that competition among phonological neighbors influences activation of semantically related concepts during spoken word recognition (Apfelbaum, Blumstein, & McMurray, 2011). The model made a novel prediction: Semantic input modulates the effect of phonological neighbors on target word processing, producing an approximately inverted-U-shaped pattern with a high phonological density advantage at an intermediate level of semantic input—in contrast to the typical disadvantage for high phonological density words in spoken word recognition. This prediction was confirmed with a new analysis of the Apfelbaum et al. data and in a visual world paradigm experiment with preview duration serving as a manipulation of strength of semantic input. These results are consistent with our previous claim that strongly active neighbors produce net inhibitory effects and weakly active neighbors produce net facilitative effects.

Keywords: Interactive activation and competition; Neighborhood effects; Lexical processing

1. Introduction

Across cognitive domains, including action (e.g., Botvinick, Buxbaum, Bylsma, & Jax, 2009; Cisek, 2006), perception (e.g., Palmeri, Wong, & Gauthier, 2004), memory (e.g., Polyn, Norman, & Kahana, 2009), and language (e.g., McClelland & Rumelhart, 1981), multiple similar representations are activated simultaneously and compete for final decision (Chen & Mirman, 2012). In the language domain, lexical neighborhood

Correspondence should be sent to Qi Chen, School of Psychology, South China Normal University, 510631 Guangzhou, China. E-mail: chenqi.research@gmail.com

effects—how recognition or production of a target word is affected by similar words—have been the focus of a lot of research (for reviews, see, e.g., Chen & Mirman, 2012; Magnuson, Mirman, & Myers, 2013), and these studies have gradually revealed the dynamics of graded, parallel activation of multiple similar representations. Different kinds of lexical similarity—orthographic, phonological, and semantic—have been investigated across a wide variety of tasks (e.g., word reading, word repetition, lexical decision, picture naming, semantic categorization, or judgment, etc.). Across these many studies, a striking pattern of consistent reversals has been revealed: The neighborhood effects are quite consistent for a particular task and neighbor type; however, the direction of the effect—facilitation versus inhibition—differs across tasks and neighbors types.

Briefly (for a more thorough review, see Chen & Mirman, 2012), differences as a function of neighbor type have been shown in visual word recognition, where orthographic and phonological neighbors generally facilitate word recognition (for a classic review, see Andrews, 1997; for more recent extensions, see, e.g., Yarkoni, Balota, & Yap, 2008; Yates, 2005), unless they are high-frequency neighbors (Davis, Perea, & Acha, 2009; Ferraro & Hansen, 2002; Grainger & Jacobs, 1996; for a review, see Grainger, 2008) or transposed-letter neighbors (e.g., *silver–sliver*; Acha & Perea, 2008; Andrews, 1996; Johnson, 2009), in which case they inhibit word recognition. Semantic neighbors provide another example of different effects of different neighbor types: Distant semantic neighbors (concepts that share a few semantic features) facilitate word recognition and production, whereas near semantic neighbors (concepts that share many semantic features) inhibit word recognition and production (Mirman, 2011; Mirman & Magnuson, 2008). Differences due to task are most obvious when comparing the effects of phonological neighborhood density: They are facilitative in visual word recognition (as just described), inhibitory in spoken word recognition (e.g., Dufour & Peereman, 2003a,b; Garlock, Walley, & Metsala, 2001; Goldinger, Luce, & Pisoni, 1989; Luce, 1986; Luce & Pisoni, 1998; Magnuson, Dixon, Tanenhaus, & Aslin, 2007), and facilitative in spoken word production (e.g., Kittredge, Dell, Verkuilen, & Schwartz, 2008; Vitevitch, 2002; Vitevitch & Sommers, 2003).

A variety of interactive activation and competition (IAC) models have been proposed to account for neighbor effects individually, but those models did not address why neighbors facilitate processing in some contexts and inhibit processing in others. In a recent computational modeling study (Chen & Mirman, 2012), we addressed this specific issue using the IAC framework. In the IAC framework, neighbors exert inhibitory effects through direct inhibitory connections (i.e., lexical competition) and facilitative effects through recurrent or resonant excitatory connections with units in other layers (sometimes called “gang effects”). The Chen and Mirman (2012) simulations captured the core qualitative patterns of orthographic, phonological, and semantic neighbor effects in word recognition and production tasks and revealed the core computational principle that determined whether neighbor effects were facilitative or inhibitory: *Strongly active neighbors had a net inhibitory effect, and weakly active neighbors had a net facilitative effect*. That is, for strongly active neighbors, their inhibitory effect outweighed their facilitative effect, but for weakly active neighbors, their facilitative effect outweighed their inhibitory effect. This pattern is consistent with the few previous attempts to explain why neighbors have opposite effects in different tasks (Dell & Gordon, 2003; Magnuson & Mirman, 2007).

Although the Chen and Mirman (2012) model implemented three processing levels (phonological/orthographic, lexical, and semantic), each of the reported simulations only considered the two processing levels relevant for a particular effect. So, for example, the spoken word recognition simulation used only the phoneme and lexical levels. Real-world spoken word recognition requires listeners to not only activate the word-form but also access its meaning (though not all laboratory tasks may require this). Processing at different levels (e.g., lexical and semantic) is thought to occur in a continuous or cascading fashion (e.g., McClelland, 1979); that is, semantic processing begins as soon as there is input to the semantic level, without waiting for lexical (or even sublexical) processing to complete. Particularly relevant to the study of neighborhood effects, one recent study showed that phonological neighborhood density influences the activation of semantic neighbors during spoken word comprehension (Apfelbaum et al., 2011).

Apfelbaum et al. (2011) used eye tracking and the "visual world paradigm" (VWP; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; cf. Cooper, 1974): On critical trials in their experiment, four pictures were displayed—one target object, one object semantically related to the target, and two objects unrelated to the target—then the spoken name of the target object was presented and participants had to click on the named picture. Past studies have shown that listeners tend to fixate semantically related objects more than unrelated objects in this task (e.g., Huettig & Altmann, 2005; Mirman & Magnuson, 2009; Yee & Sedivy, 2006). Apfelbaum et al. replicated the general semantic competition effect and found that it was modulated by the phonological neighborhood of the target word. Specifically, semantic neighbors were activated more when the target had few phonological neighbors than when it had many phonological neighbors. That is, a reduction in phonological–lexical competition translated into increased activation of semantic neighbors.

In the present study, we extended our previous model to account for the findings of Apfelbaum et al. (2011) and thereby investigated the interactions between phonological and semantic representations. This investigation produced novel predictions from the model regarding how semantic input from picture preview modulates the effect of phonological neighbors on target word recognition, which we then evaluated with reanalyses of existing data and a new experiment. The simulations and experiment examine our proposed core computational principle—that strongly active neighbors have a net inhibitory effect and weakly active neighbors have a net facilitative effect—from a new perspective. Previous comparisons examined different items (e.g., high neighborhood density items vs. low neighborhood density items) or different tasks (e.g., word recognition vs. word production); here, we consider the same items in the same task, but under conditions where a particular set of neighbors is predicted to be strongly active versus weakly active.

The remainder of the paper is organized as follows: We begin by outlining the main features of our model and demonstrating that the model correctly accounts for the Apfelbaum et al. finding that phonological neighborhood density modulates semantic competition during spoken-word-to-picture matching. We then describe a novel prediction based on simulations of our model and test this prediction with reanalysis of the Apfelbaum et al. behavioral data. Finally, we further extend this prediction in additional simulations and test it using the VWP. We conclude by discussing the implications of our computational and behavioral findings.

2. Network architecture

The simulations were carried out using the same model architecture, implementation, and parameters as our previous simulations of orthographic, phonological, and semantic neighbor effects (see Appendix; Chen & Mirman, 2012). This model instantiated the basic computational principles of the IAC framework (a subset of the principles of parallel distributed processing); namely, that cognitive processing is an emergent property of interactions among simple processing elements and those interactions are governed by bidirectional weighted connections (for a review of the history and impact of parallel distributed processing, see Rogers & McClelland, 2014; and for a more specific focus on IAC see McClelland, Mirman, Bolger, & Khaitan, 2014). The original goal of the model was to examine neighborhood effects at orthographic, phonological, lexical, and semantic levels, so rather than implementing most of the language processing system, we chose a simple and abstract implementation that allowed us to evaluate the computational principles that give rise to qualitative behavioral patterns across these domains. The results were quite consistent across a wide range of parameters: the critical pattern that strongly active neighbors exert a net inhibitory effect and weakly active neighbors exert a net facilitative effect held over a wide range of parameter values. In the present simulations, we avoid major departures from the original model formulation in order to retain those benefits while adapting the basic structure to capture the properties of the VWP task. That is, we focus on a relatively specific set of findings in the present report, but by keeping the same architecture, neighbor definition, and parameters, we extend the model to capture new phenomena and make new predictions while ensuring that the model still accounts for that larger set of phenomena.

The model was an IAC network with simple processing units organized into three layers: units in the first layer corresponded to elements of word form (i.e., phonemes), units in the second layer corresponded to lexical elements (i.e., words in the model's lexicon), and units in the third layer corresponded to elements of meaning (i.e., semantic features of concepts denoted by the words). As in other IAC models (e.g., Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; McClelland & Elman, 1986; McClelland & Rumelhart, 1981), congruent units in different layers had bidirectional excitatory connections. That is, each word unit had bidirectional connections to its constituent phonemes and to its semantic features. To implement competition, each unit in the word layer was connected to each other unit in the word layer by bidirectional inhibitory connections with connection strength scaled by a sigmoid function of unit activation (i.e., weakly active word units had very little inhibitory effect on other word units, and strongly active words units had a very strong inhibitory effect on other word units). This non-linearity was implemented in order to allow initial parallel activation of many word candidates while still forcing the model to eventually settle to a single active representation. In exploratory simulations, we found that a linear inhibition function would either not allow sufficient activation of weak candidates (if the linear inhibition was high) or would not force the model to settle to a single active representation (if the linear inhibition was low).

To imitate the kind of recurrent connections that would emerge among semantic representations over the course of learning (e.g., Cree, McRae, & McNorgan, 1999; Rogers & McClelland, 2004), semantic units were generally connected by inhibitory weights, but this inhibition was reduced for each concept in which the semantic units (features) co-occurred (for evidence of facilitative effects of feature co-occurrence, see e.g., Cree & McRae, 2003; Rogers & McClelland, 2004). In other words, a semantic feature such as "has wings" was assumed to have inhibitory connections to unrelated features such as "has strings," excitatory connections to strongly (cor)related features such as "has feathers," and intermediate weights to weakly related features such as "eats fish."

All units followed the standard IAC activation function in which positive net input drives unit activation toward its maximum (1.0) and negative net input drives unit activation toward its minimum (0.0). All simulations reported here used a simple lexicon consisting of 11 two-phoneme words. Each word was represented by a unique unit in the lexical layer and each phoneme was represented by a unique unit in the phoneme layer. As in our previous simulations (Chen & Mirman, 2012), the phonological input was presented sequentially to the phoneme units: The first phoneme was activated for Time Steps 1 through 30, and the second phoneme was activated for Time Steps 25 and through 54 (the small amount of overlap was intended as a rough analog to co-articulation). This simple model was designed to account for single word processing tasks only, so a slot-based input representation was used (e.g., Dell et al., 1997; Plaut, McClelland, Seidenberg, & Patterson, 1996).

Each word was also associated with 10 semantic feature units. As in the previous simulations (Chen & Mirman, 2012), one word ("high density target") had three phonological neighbors, which all shared the first phoneme; another word ("low density target") had no phonological neighbors. In addition, the lexicon contained one semantically related "competitor" for each of the target words, each of which was phonologically unrelated to the target word but shared 5 out of 10 semantic features with the target word. Finally, the lexicon contained four words that were phonologically and semantically unrelated to any of the other words, which we used as "unrelated distractors." In the VWP task, pictures of objects are presented along with a spoken word and the task is to match them. Thus, the primary task is spoken word comprehension, but there is semantic input as well (for evidence of semantic activation during VWP preview, see Yee, Huffstetler, and Thompson-Schill, 2011).[1] To capture this combination of phonological and semantic input, a small amount of external input (0.01) was also provided to all semantic features of all four objects in the display (target, competitor, and two unrelated distractors). As in the Apfelbaum et al. experiment, none of the displayed objects were phonological competitors of the target.

The spoken word-to-picture matching task used by Apfelbaum et al. (2011) involves integrating semantic and phonological information in order to make a decision—for both the final manual mouse click response and the intermediate eye fixations. In the model, the phoneme and semantic layers receive external input, so their activation patterns primarily reflect those external inputs. The integration of those inputs happens in the word layer, so model performance was evaluated at the word layer (see also, e.g., Kukona & Tabor, 2011; Mirman et al., 2013; Spivey, 2008). The word layer's localist representation also reflects the categorical nature of the behavioral responses and makes it possible to capture the fact that fixation proportions for distractors eventually drop to (essentially) 0 (the distributed word representations of the phonological and semantic layers mean that activation of related representations would never be completely eliminated). We report word unit activations to demonstrate model dynamics and explicitly model the constraints of the VWP task using the Luce (1959) choice rule ($k = 10$) with the four pictures as choice options, as in the other models of VWP studies (for complete specification of the Luce choice rule and examples of application to VWP data, see, e.g., Dahan, Magnuson, & Tanenhaus, 2001; Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Mirman, Yee, Blumstein, & Magnuson, 2011).

3. Simulation 1: Apfelbaum et al. (2011) data

The model's output is plotted in Fig. 1 along with the behavioral data from Apfelbaum et al. (2011). The model correctly fits the behavioral results reported by Apfelbaum et al. (2011): There was greater activation of and response (fixation) probability for semantic competitors of target words with low phonological neighborhood density than those with high phonological neighborhood density. Our simple IAC model captured this pattern

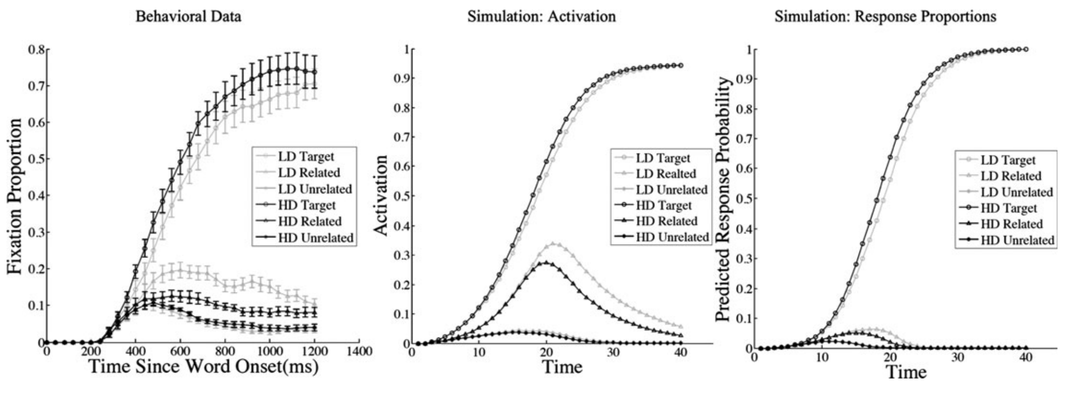

Fig. 1. Effects of phonological neighbors on semantic competition. The left panel (redrawn from Apfelbaum et al., 2011) shows the time course of fixation proportions for the target (circles), semantically related item (triangles), and unrelated picture (stars), for low phonological density stimuli (gray curves) and high phonological density stimuli (black curves). The middle and right panels show the corresponding simulation results (word unit activations and Luce choice probabilities, respectively).

because fewer phonological neighbors meant less competition in the lexical layer, allowing semantic neighbors to become more active.

The model also predicted a reversal of the canonical inhibitory effect of phonological neighbors on spoken target word recognition: Activation of the target word rose faster for the high neighborhood density word than the low density word. This pattern contrasts with the results of our previous simulation (Chen & Mirman, 2012; Simulation 3), which showed clear inhibitory effects of phonological neighbors on spoken word recognition when no semantic input was presented. The key difference was that the presence of semantic input provided additional excitatory input to the target word and the words corresponding to the other objects in the display, which were phonologically unrelated to the target. As a result, the phonological neighbors received more lateral inhibition (from the more active target and the partially activated distractors) so their activation was reduced. This allowed the phonological neighbors' facilitative effect on the target (due to recurrent excitation of shared phoneme units) to become stronger than their lateral inhibition of the target. Since the phonological neighbors do not share any phonemes with the semantic competitor, they had no facilitative effect on it, only an inhibitory effect due to lateral inhibition. This lateral inhibition effect was somewhat reduced when the phonological neighbors were less active, but for any level of phonological neighbor activation, fewer neighbors still meant less inhibition of the semantic competitor. Before examining this aspect of the model in more detail, we reanalyzed the target fixation data from Apfelbaum et al. (2011) in order to establish whether the reversal of the canonical inhibitory effect of phonological neighbors on spoken word recognition would find any support in their data.

4. Reanalysis of Apfelbaum et al. (2011) target data

Simulation 1 predicted a reversal of the canonical inhibitory effect of phonological neighbors on spoken word recognition: Target words with many phonological neighbors should be recognized faster than words with few phonological neighbors. As a first test of this novel model prediction, we reanalyzed target fixation time course data from Apfelbaum et al. (2011), which were not analyzed in their original report (Apfelbaum et al. were interested in competitor fixations and did not report analyses of target fixations). Growth curve analysis was used to analyze the fixation data (for details, see Mirman, 2014; Mirman, Dixon, & Magnuson, 2008) from the same analysis time window used by Apfelbaum et al. (300–1,000 ms after word onset). The overall time course of target fixations was captured with a third-order (cubic) orthogonal polynomial with fixed effects of Condition (low vs. high density) on all time terms, and participant and participant × condition random effects on all time terms. Statistical significance (p-values) for individual parameter estimates was assessed using the normal approximation (i.e., treating the t-value as a z-value; for discussion and evaluation of this approach, see Barr, Levy, Scheepers, & Tily, 2013).

The results revealed significantly higher overall target fixation probability for high phonological neighborhood targets than for low phonological neighborhood targets (*Estimate* = 0.0623, *SE* = 0.0154, $p < 0.0001$). Thus, the target fixation time course data were consistent with the model's novel prediction that, under the conditions tested in this experiment, spoken word recognition should be faster for words with high phonological neighborhood density than words with low phonological neighborhood density. Although they contrast with the well-replicated inhibitory effects of phonological neighbors in spoken word recognition, this model prediction and these behavioral results fit perfectly with our argument that strongly active neighbors have a net inhibitory effect, and weakly active neighbors have a net facilitative effect on target processing (Chen & Mirman, 2012). This prediction is also consistent with the "tug-of-war" between semantic and phonological factors described by Huettig and McQueen (2007), although in their case both semantic and phonological distractors were in the display.

There is, however, an alternative account. Because looks to competitors are, by definition, not looks to targets, perhaps the facilitative effect of phonological neighborhood density on target fixation is simply an oculomotor consequence of the reduced semantic competition in this condition. That is, when the target word had fewer phonological neighbors, participants looked at the semantic competitor more, which meant that they looked at the target less. This alternative account requires a somewhat strange paradox: Apfelbaum et al.'s (relatively uncontroversial) interpretation was that the semantic competition was increased for low phonological neighborhood items because there was less phonological competition for these items. So this alternative account would have to be that the increase in semantic competition is larger than the reduction in phonological competition that produced it. This account also differs from our model's account in that the alternative account is purely oculomotor and does not reflect an interaction between phonological and semantic processing. As such, it predicts that strength of semantic input should monotonically modulate the effect of phonological neighborhood density effect on target fixation: Stronger semantic input produces more semantic competition, and therefore progressively larger reversals of the phonological neighborhood effect. Alternatively, on a strictly feed-forward interpretation, it might predict that there should be no effect of strength of semantic input if the fixation time course effects are only due to phonological competition. In Simulation 2, we further examine how the amount of semantic input modulates the effect of phonological neighbors in our model.

5. Simulation 2: Effects of varying the amount of semantic input

In the first simulation, semantic input to the four displayed objects reduced activation of phonological neighbors, making their net effect on target words facilitative. To examine these dynamics more thoroughly, the same basic simulation was repeated with different levels of semantic input. As expected, the effect of phonological neighbors on semantic competition reported by Apfelbaum et al. (2011) emerged across a wide range of semantic input strength because even reduced activation of phonological neighbors had an inhibitory effect on the semantic competitor. Fig. 2 shows the activation of phonological neighbors during spoken word recognition under varying levels of semantic input (low: 0.00001; medium: 0.01, as reported above; high: 0.05).[2]

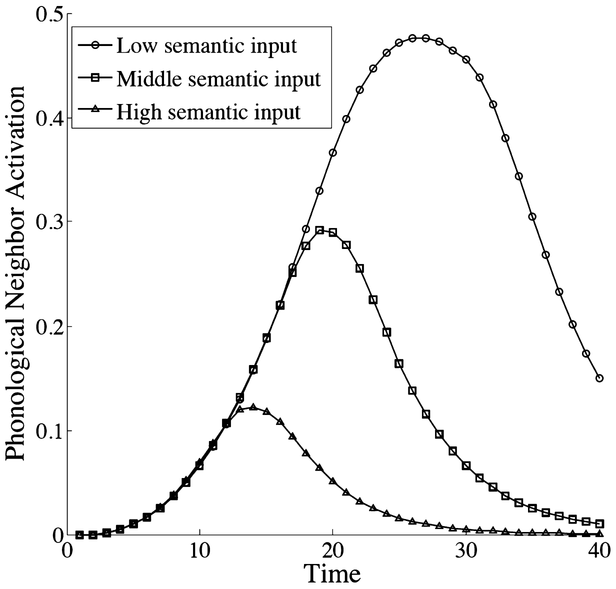

Fig. 2. The effect of semantic input on phonological neighbor activation. It shows activation of phonological neighbors with different level of semantic input: low semantic input (open circles), middle semantic input (open squares), and high semantic input (open triangles).

Fig. 3 shows how strength of semantic input modulated the effect of phonological neighborhood density on target word activation. With a low level of semantic input (Fig. 3, left), target words with many phonological neighbors were processed more slowly than the target words with few phonological neighbors. This result matched our previous simulation of spoken word recognition with semantic input (Chen & Mirman, 2012; Simulation 3) and was consistent with inhibitory effects of neighbors for spoken word recognition. At this low level of semantic input, phonological neighbors were strongly activated and target activation was substantially slowed by lateral inhibition from those neighbors. In contrast, moderate semantic input led to a reversal—a facilitative effect—of phonological neighbors (Fig. 3, middle). Moderate semantic input activated lexical representations of the visual distractors (which were not phonologically related to the target), which competed with the phonological neighbors, thus reducing—but not eliminating—their activation (see also Fig. 2). At this reduced level of activation, phonological neighbors exhibited a net facilitative effect on the target because their excitatory connections to shared phoneme units had a bigger effect than their lateral inhibitory connections to the target word unit.

When semantic input was further increased (Fig. 3, right), the visual distractors were even more strongly activated and further reduced activation of the phonological neighbors (see also Fig. 2); thus, the phonological neighbor effect was reduced (though still facilitative). This inverted-U-shaped pattern is due to the way semantic input to phonologically unrelated words modulates the activation of phonological neighbors (as shown in Fig. 2). Strongly active phonological neighbors (a low level of semantic input) had a net inhibitory effect, moderately active phonological neighbors (moderate semantic input)

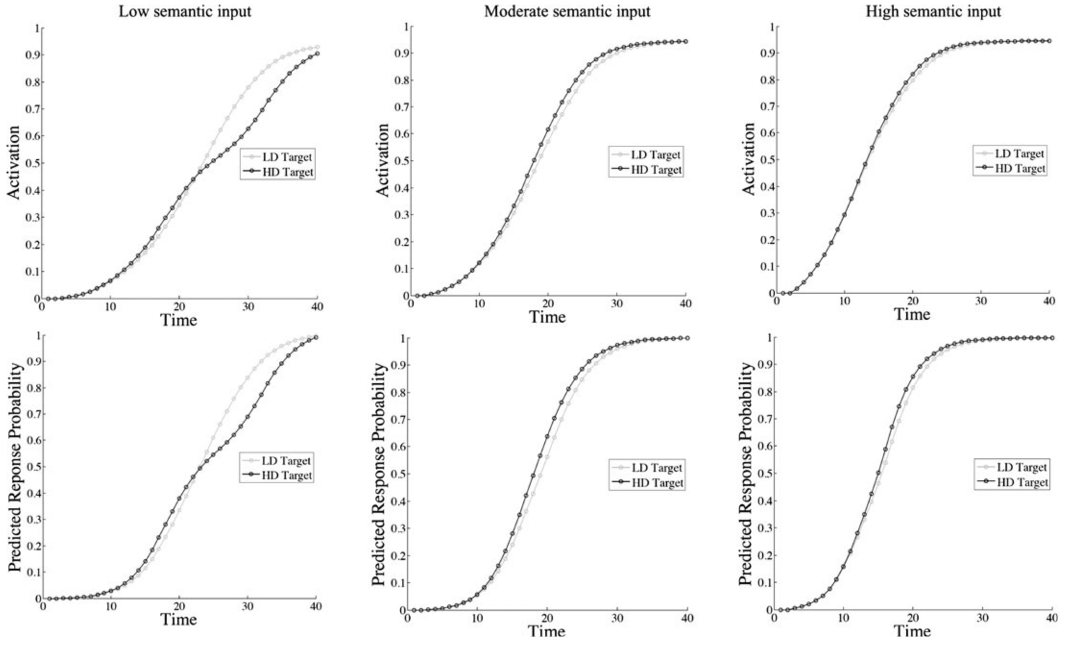

Fig. 3. The simulated time course of target processing for low phonological density stimuli (gray curves) and high phonological density stimuli (black curves) when the four displayed objects receive low semantic input (left), moderate semantic input (middle), and high semantic input (right). The top row shows word unit activations, and the bottom row shows Luce choice proportions (i.e., predicted fixation proportions).

had a net facilitative effect, and weakly active phonological neighbors (a high level of semantic input) had little effect on target word recognition. The following experiment was designed to test this model prediction.

6. Experiment

This experiment was designed to test the model's prediction that the phonological neighborhood density effect would follow an approximately inverted-U-shaped pattern as a function of amount of semantic input. The middle portion of this inverted-U was the critical prediction because it contrasts with previous findings (phonological neighbors predicted to have a facilitative rather than their standard inhibitory effect on spoken word comprehension) and because it was predicted to be the "odd-one-out" (i.e., to be different from the other two conditions). For this reason, our analyses focused on this condition and how it compared with the other two.

Previous studies have shown that longer preview durations produce more semantic processing of pictured objects (e.g., Yee et al., 2011; see also Huettig & McQueen, 2007), so the degree of semantic input was manipulated by varying the preview duration. Very few studies have examined the effect of preview duration, but previews of 750–1,500 ms have been frequently used in VWP studies and considered sufficient. On this basis, 1,000 ms was chosen as the long preview duration; 0 ms was chosen as the short preview duration because it is the minimum preview duration that still allows capturing the full time course of target word processing; and 500 ms was chosen as the middle preview duration because it was in the middle between those extremes. Although semantic activation is thought to proceed fairly quickly, we assume that within reasonable bounds (i.e., at least up to 1,000 ms), longer preview will produce more semantic activation.

6.1. Methods

6.1.1. Participants

Thirty University of Connecticut undergraduates completed the experiment for course credit. Data from two additional participants were excluded due to excessive track loss (>30%). All participants were native speakers of American English and reported normal hearing and either normal or corrected-to-normal vision.

6.1.2. Materials

The materials, including all object images and target word audio files, were identical to those used by Apfelbaum et al. (2011). These included 72 critical trials (36 low density and 36 high density), an additional 36 filler trials, and 4 practice trials.

6.1.3. Procedure

Stimulus presentation and response recording were conducted by E-Prime Professional (Version 2.0, Psychology Software Tools, Inc. Sharpsburg, PA, USA.) experimental design software and a remote (desktop-mounted) Eyelink 1,000 eye tracker was used to record gaze position at 250 Hz. The experiment began with a standard nine-point calibration procedure, followed by four practice trials on which feedback was provided. Each trial began when the participant clicked a cross (+) in the center of the screen. This caused four images to appear, each near one of the screen corners. After a delay of 0, 500, or 1,000 ms (constant during the session, manipulated between-participants, $N = 10$ for each preview duration), participants heard the target word through headphones and had to click on the image corresponding to the spoken word. Trial order was randomized and the four object positions were randomly assigned on each trial.

6.1.4. Data analysis

As in the reanalysis reported above, growth curve analysis (Mirman, 2014; Mirman et al., 2008) was used to analyze the target gaze data from 300 to 1,000 ms after word onset. The overall time course of target fixations was captured with a third-order (cubic) orthogonal polynomial, with fixed effects of Condition (low vs. high density; within-participants), Preview Duration (0, 500, or 1,000 ms; between-participants), and Condition × Preview Duration interaction on all time terms. The model also included participant and Participant × Condition random effects on all time terms. In the first analysis phase, we used model comparisons to test whether adding the fixed effects of each factor (Condition, Preview Duration, and their interaction) on all time terms improved model fit. Our primary interest was whether the Condition × Preview Duration interaction would improve model fit, that is, whether preview duration modulated the effect of phonological density on the time course of spoken word recognition. In the second phase, the Condition × Preview Duration interaction parameter estimates for individual time terms were statistically evaluated in order to understand the effect of each factor on the time course of target fixation. These interaction term parameter estimates capture the effect of Preview Duration on the target fixation differences between low-density and high-density conditions; that is, the effect of Preview Duration on the low density advantage. Effects on the intercept term reflect differences in overall fixation probability, effects on the linear term reflect differences in the linear slope of the fixation curves, effects on the quadratic term reflect differences in the curvature of the fixation curves, and effects on the cubic term tend to capture effects in the tails of the fixation curves.

6.2. Results

Overall accuracy was very high (>96% correct in all conditions) with a slight numerical advantage for the High-Density condition (M = 99.4%, SE = 0.4%) over the Low-Density condition (M = 97.3%, SE = 1.0%) and minimal differences between preview durations (0 ms: M = 97.5%, SE = 1.0%; 500 ms: M = 98.9%, SE = 0.8%, 1,000 ms: M = 98.8%, SE = 0.6%). Logistic regression revealed no statistically significant effects of density, preview duration, or density × preview–duration interaction on accuracy (all $p > 0.7$). Only correct response trials were included in the fixation analysis.

6.2.1. Target fixation

The time course of target fixation is shown in Fig. 4. The Condition × Preview Duration interaction improved model fit ($\chi^2(8) = 19.67$, $p < 0.05$), indicating that preview

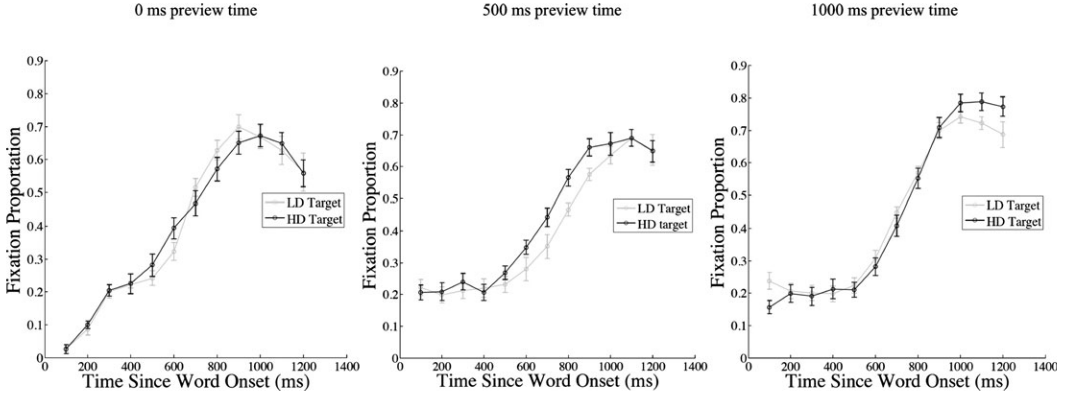

Fig. 4. Results of behavioral study. The panels show the time course of fixation proportions for the target with different preview times: 0 ms (left panel), 500 ms (middle panel), and 1,000 ms (right panel) for low phonological density stimuli (gray curves) and high phonological density stimuli (black curves). Error bars indicate ± SE.

duration did modulate the effect of phonological density. Evaluation of the individual parameter estimates for the interaction factor effect on the intercept term (Table 1) revealed that the low-density advantage was significantly smaller after a 500 ms preview than a 0 ms preview (*Estimate* = −0.0545, *SE* = 0.0254, $p < 0.05$) and was marginally larger after a 1,000 ms preview than a 500 ms preview (*Estimate* = 0.0461, *SE* = 0.0254, $p = 0.07$). That is, as predicted by the model, there was an approximately inverted-U-shaped pattern such that a 500 ms preview duration produced significantly less low-density advantage than a 0 ms preview and marginally less than a 1,000 ms preview. This inverted-U-shaped pattern is plotted in Fig. 5, with the intercept term parameter estimate used as the phonological neighbor effect size.

The critical novel prediction from the model was that there should be a reversal of the canonical inhibitory effect of phonological neighbors on spoken word recognition in the

Table 1
Parameter estimates (*SE* in parentheses) for Condition × Preview Duration interaction effects. Parameter estimates are for the difference in the low-density advantage (i.e., low-density condition relative to the high-density condition) between each pair of preview duration conditions

	Preview Duration Comparison		
Time Term	0 ms vs. 500 ms	0 ms vs. 1,000 ms	500 ms vs. 1,000 ms
Intercept	−0.0545 (0.0254)*	−0.0084 (0.0254)	0.0461 (0.0254)~
Linear	−0.0636 (0.0664)	−0.1109 (0.0664)~	−0.0472 (0.0664)
Quadratic	0.01063 (0.0469)*	−0.0758 (0.0469)	−0.1821 (0.0469)***
Cubic	0.0822 (0.0409)*	0.0943 (0.0409)*	0.0121 (0.0409)

Note. ~$p < 0.1$, *$p < 0.05$, ***$p < 0.001$.

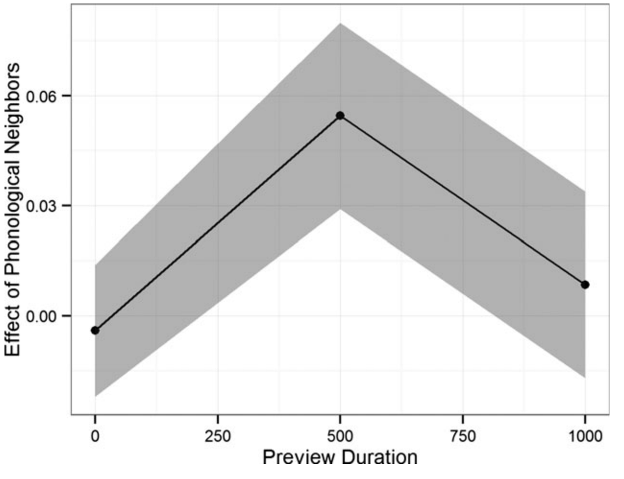

Fig. 5. The inverted U-shaped pattern of effects of phonological neighbors as a function of preview duration. Positive effect indicates a facilitative effect of neighbors. The effect size is based on the effects of condition and preview duration on the intercept term. The gray region represents ± *SE*.

500 ms preview condition. Analysis of the effect of phonological neighborhood density in just the 500 ms preview condition confirmed that there was a significant facilitative effect of phonological neighborhood density (i.e., a reversal of the canonical inhibitory effect) on both the intercept (*Estimate* = −0.0505, *SE* = 0.0182, $p < 0.01$) and the quadratic (*Estimate* = 0.110, *SE* = 0.0351, $p < 0.01$) terms, just as in the reanalysis of the Apfelbaum et al. data.

The inverted-U-shaped pattern of phonological neighbor effects is consistent with the model prediction and conflicts with the alternative account of the target fixation data from Apfelbaum et al. (2011). Under that alternative account, the facilitative effect of phonological neighborhood on target fixation should either be constant across preview durations or should increase monotonically. The (marginal) reduction in the facilitative effect of phonological neighborhood on target fixation from the 500 ms preview duration to the 1,000 ms preview duration conflicts with both interpretations of this alternative account.

6.3. Discussion

These results replicate our reanalysis of the Apfelbaum et al. (2011) data and indicate that it is possible to reverse the established pattern that phonological neighbors inhibit spoken word recognition (e.g., Luce & Pisoni, 1998). The eye data revealed that the effect of phonological neighborhood density was modulated by preview duration: It was significantly more facilitative after a 500 ms preview than no preview (i.e., 0 ms preview duration) and marginally more facilitative than after a 1,000 ms preview. This approximately inverted-U-shaped pattern was predicted by simulations of our computational model (Fig. 3) on the assumption that longer preview duration allows for more semantic processing of the pictured objects in parallel with the phonological processing. When there is no preview, and consequently little semantic processing, phonological neighbors should be strongly activated and low-density words have an advantage because they have fewer competitors. After a moderate preview (500 ms), moderate semantic processing constrains the lexical competition, allowing less activation of phonological neighbors, so their net effect should be facilitative (due to recurrent facilitation of shared phonemes) rather than inhibitory; thus, high-density words have an advantage because they have more neighbor facilitation. After a long preview (1,000 ms), semantic input dominates the lexical competition and there should be little effect of the number of phonological neighbors.

6.4. Relation to previous VWP studies of neighborhood effects

6.4.1. Apfelbaum et al. (2011)

The target fixation data from Apfelbaum et al. (2011) were similar to our 500 ms preview duration data although, arguably, their participants had more time to preview the display. In the Apfelbaum et al. study, the objects were presented for 500 ms, then participants were prompted to click on a fixation dot at the center to initiate the auditory input. The overall average duration from picture onset to onset of the target word was approximately 1,360 ms (B. McMurray, pers. comm., August 21, 2012). Under these conditions, it is difficult to define the exact duration of the preview period (indeed, this is why we eliminated the need to click to hear the target word), but given the structure of their preview period, it is possible that participants spent approximately 500 ms previewing the objects and then approximately 800 ms responding to the prompt to initiate the trial. Thus, we consider their data to converge with our 500 ms preview duration data.

6.4.2. Magnuson et al. (2007)

Using an almost identical paradigm, Magnuson et al. (2007) reported inhibitory effects of cohort density and a cross-over pattern for phonological neighborhood density, such that fixation proportions for low-density targets initially rose more slowly, then more quickly, than for high-density targets. The onset of the display was simultaneous with the onset of a carrier phrase ("click on the [target]"). The duration of the "click on the" portion was 440 ms, slightly less than our 500 ms preview duration. Because of the auditory input, participants in the Magnuson et al. study may not have used this 440 ms to preview the objects (or used it less than during a silent 500 ms preview). On this view, the cross-over effect in our 0 ms preview duration is an exact replication of their finding. In addition, there were no semantic competitors in the Magnuson et al. study. The presence of a semantic competitor may increase attention to semantic information, thus increasing its impact. As a result, our short preview condition may correspond to the same amount of semantic processing as the moderate preview duration in Magnuson et al.

7. General discussion

We extended our model of lexical neighborhood effects (Chen & Mirman, 2012) to account for eye-tracking data showing effects of phonological neighborhood density on semantic activation in spoken word recognition (Apfelbaum et al., 2011). This simulation also made a novel prediction based on the model's core principle that strongly active neighbors have a net inhibitory effect and weakly active neighbors have a net facilitative effect. According to the model, phonological neighbors are normally strongly active during spoken word recognition, but the visual input in this task (preview of objects) provided an additional (semantic) input to phonologically unrelated words, which competed with the phonological neighbors, thereby reducing their activation and leading to a net facilitative effect on target recognition. A new analysis of existing data showed that the target fixation data were consistent with the prediction that, when both phonological and semantic input are presented, phonological neighbors will have a facilitative effect on spoken word recognition. Further simulations explored how the strength of semantic input modulated the effect of phonological neighbors on the time course of target word activation predicted an approximately inverted-U-shaped pattern with a maximal high-density advantage at an intermediate level of semantic input. A new experiment manipulated preview duration as a proxy for strength of semantic input and found an approximately inverted-U-shaped pattern of phonological neighborhood effects on spoken word recognition, as predicted by the model.

An important aspect of these simulations and experiments is that there were no phonological competitors in the display. This meant that the target was the only item that received both phonological and semantic input. At the intermediate semantic input strength (500 ms preview duration), the semantic competitors were sufficiently active to reduce the activation of the phonological competitors, flipping their net effect to be facilitative rather than inhibitory. If one or more of the displayed pictures were phonological neighbors, then those items would receive both phonological and semantic input, making them even stronger competitors and leading them to have a net inhibitory effect.

It is also important that the word-to-picture matching task requires selecting a single response that matches both the phonological spoken form and its semantic representation. In this task context, semantic and phonological competitors are equivalent with respect to the task demands (they reflect neighborhood properties of the target, but they need to be eliminated in order to make a correct response). However, this equivalence may not be a general property of the language system. Other tasks, such as phoneme detection, lexical decision, or semantic categorization, which have differential phonological and semantic demands, may produce somewhat different effects of neighbors.

The simulations reported here are a straightforward extension of our earlier model (Chen & Mirman, 2012), which was itself based on previous IAC models of word recognition and production (e.g., Dell & Gordon, 2003; Dell et al., 1997; Magnuson & Mirman, 2007; McClelland & Elman, 1986; McClelland & Rumelhart, 1981). Consistent with IAC principles and other recent computational models for the VWP (e.g., Kukona & Tabor, 2011; Mirman et al., 2013; Spivey, 2008), our model embodies the idea that VWP can be understood as an integration or decision process melding two sources of information: visual and auditory. In fact, one of the consistent findings across many VWP studies is the online integration of visual and linguistic sources of information to, for example, resolve syntactic ambiguity (Chambers, Tanenhaus, & Magnuson, 2004) and phonological ambiguity (Magnuson, Tanenhaus, & Aslin, 2008). Particularly relevant to the present study, Magnuson et al. (2008) suggested that phonological neighbor activation depends on whether the visual display allows or disallows the neighbor. Our finding further shows that manipulating the strength of semantic input (preview duration of visual display) can modulate phonological neighbor activation.

Similarly, Huettig and McQueen (2007) demonstrated that preview duration can modulate the relative importance of phonological, shape, and semantic information in the VWP. Our simulations provide a concrete computational framework for investigating this "tug of war." Huettig and McQueen found that shape information—a semantic feature that is particularly visually salient—was particularly important at short preview durations. Although outside the scope of our simple model (which was designed to investigate the basic computational principles of interactions between semantic and form neighbors across tasks), such differences in the time course of activation of different semantic features are a growing and fruitful line of research (e.g., Kalénine et al., 2012; Yee et al., 2011).

In sum, using a combination of computational model simulations and behavioral experiments, we showed that visual (semantic) input modulates the effect of phonological neighbors during spoken word recognition. We began by extending our model to account for recent evidence that when there is less phonological competition, semantic competitors become more active. Those simulations also revealed that semantic competition modulates the effect of phonological neighbors on target word processing. The results provided new evidence consistent with our proposal that strongly active neighbors have a net inhibitory effect and weakly active neighbors have a net facilitative effect (Chen & Mirman, 2012). Finally, we manipulated the degree of semantic processing by manipulating the duration of the preview period—the length of time the visual objects were displayed before the spoken target word was presented (see also Huettig & McQueen, 2007; Yee et al., 2011). This experimental approach provides a promising method for investigating the time course of interactions between semantic and phonological processing.

Acknowledgments

This research was supported by R01DC010805 to DM and by the Moss Rehabilitation Research Institute. The authors thank Keith Apfelbaum, Sheila Blumstein, and Bob McMurray for making their behavioral data available for reanalysis and providing their experimental stimuli for our experiment. We also thank Sergey Kornilov for his help collecting the behavioral data.

Notes

1. Yee et al. (2011) tested relatively long preview durations (1,000 and 2,000 ms) in the context of specific semantic relations (function and form). In general, semantic activation from pictures is thought to proceed very quickly (e.g., Dell'Acqua & Grainger, 1999; Huettig & McQueen, 2007). For our purposes, it is only important that the pictures provide some semantic input.
2. This approach was chosen because it is relatively neutral with respect to theories of semantic processing and activation. Alternatively, it is possible that some semantic features are activated before others (e.g., Kalénine, Mirman, Middleton, & Buxbaum, 2012; Yee et al., 2011). In the context of our simple model, sequential activation of individual features and gradual simultaneous activation of all features both lead to gradually increasing activation of the visually presented object names, so they are equivalent in terms of their consequences at the lexical layer.

References

Acha, J., & Perea, M. (2008). The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. *Cognition*, *108*, 290–300.

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language*, *35*(6), 775–800.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin & Review*, *4*, 439–461.

Apfelbaum, K. S., Blumstein, S. E., & McMurray, B. (2011). Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. *Psychonomic Bulletin & Review*, *18*(1), 141–149.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Botvinick, M. M., Buxbaum, L. J., Bylsma, L. M., & Jax, S. A. (2009). Toward an integrated account of object and action selection: A computational analysis and empirical findings from reaching-to-grasp and tool-use. *Neuropsychologia*, *47*(3), 671–683.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of experimental psychology. Learning, Memory, and Cognition*, *30*(3), 687–696.

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.

Cisek, P. (2006). Integrated neural processes for defining potential actions and deciding between them: A computational model. *The Journal of Neuroscience*, *26*(38), 9761–9770.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*(2), 163–201.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*(3), 371–414. doi:10.1016/S0364-0213(99)00005-1.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*(4), 317–367.

Davis, C. J., Perea, M., & Acha, J. (2009). Re(de)fining the orthographic neighborhood: The role of addition and deletion neighbors in lexical decision and reading. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1550–1570. doi:10.1037/a0014253.

Dell, G. S., & Gordon, J. K. (2003). Neighbors in the Lexicon: Friends or foes? In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production: Differences & similarities* (pp. 9–37). Boston, MA: Mouton de Gruyter.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*(4), 801–838.

Dell'Acqua, R., & Grainger, J. (1999). Unconscious semantic priming from pictures. *Cognition*, *73*(1), B1–B15.

Dufour, S., & Peereman, R. (2003a). Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the prime word. *Cognition*, *88*(3), B33–B44.

Dufour, S., & Peereman, R. (2003b). Lexical competition in phonological priming: Assessing the role of phonological match and mismatch lengths between primes and targets. *Memory & Cognition*, *31*(8), 1271–1283.

Ferraro, F. R., & Hansen, C. L. (2002). Orthographic neighborhood size, number of word meanings, and number of higher frequency neighbors. *Brain and Language*, *82*(2), 200–205.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, *45*(3), 468–492.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory & Language*, *28*, 501–518.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, *23* (1), 1–35.

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

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Johnson, R. L. (2009). The quiet clam is quite calm: Transposed-letter neighborhood effects on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35* (4), 943–969.

Kalénine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional action knowledge during auditory comprehension of artifact words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1274–1295. doi:10.1037/a0027626.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, *25*(4), 463–492.

Kukona, A., & Tabor, W. (2011). Impulse Processing: A dynamical systems model of incremental eye movements in the Visual World Paradigm. *Cognitive Science*, *35*(6), 1009–1051.

Luce, R. D. (1959). *Individual choice behavior*. Oxford, England: John Wiley.

Luce, P. A. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception, Technical Report No. 6*. Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University.

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

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*, 1–24.

Magnuson, J. S., & Mirman, D. (2007). *Neighborhood effects in word recognition: It's not where you live, it's how you get home*. Paper presented at the 48th Annual Meeting of the Psychonomic Society. Long Beach, CA.

Magnuson, J. S., Mirman, D., & Myers, E. (2013). Spoken word recognition. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 412–441). New York: Oxford University Press.

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition*, *108*(3), 866–873.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*(2), 202–227.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

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

McClelland, J. L., Mirman, D., Bolger, D. J., & Khaitan, P. (2014). Interactive activation and mutual constraint satisfaction in perception and cognition. *Cognitive Science*, *38*(6), 1139–89.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I An account of basic findings. *Psychological Review*, *88*(5), 375–407.

Mirman, D. (2011). Effects of near and distant semantic neighbors on word production. *Cognitive, Affective, and Behavioral Neuroscience*, *11*(1), 32–43.

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: Chapman & Hall/CRC.

Mirman, D., Britt, A. E., & Chen, Q. (2013). Effects of phonological and semantic deficits on facilitative and inhibitory consequences of item repetition in spoken word comprehension. *Neuropsychologia*, *51*(10), 1848–1856.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 65–79.

Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition*, *37*(7), 1026–1039.

Mirman, D., Yee, E., Blumstein, S. E., & Magnuson, J. S. (2011). Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling. *Brain and Language*, *117*(2), 53–68. doi:10.1016/j.bandl.2011.01.004.

Palmeri, T. J., Wong, A. C.-N., & Gauthier, I. (2004). Computational approaches to the development of perceptual expertise. *Trends in Cognitive Sciences*, *8*(8), 378–386.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. E. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.

Rogers, T. T., & McClelland, J. L. (2014). Parallel distributed processing at 25: Further explorations in the microstructure of cognition. *Cognitive Science*, *38*(6), 1024–77.

Spivey, M. J. (2008). *The continuity of mind*. New York: Oxford University Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 632–634.

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., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition*, *31*(4), 491–504.

Yarkoni, T., Balota, D. A., & Yap, M. J. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*(5), 971–979. doi:10.3758/PBR.15.5.971.

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1385–1397.

Yee, E., Huffstetler, S., & Thompson-Schill, S. L. (2011). Function follows form: Activation of shape and function features during word recognition. *Journal of Experimental Psychology: General*, *140*(3), 348–363. doi:10.1037/a0022840.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14.

Appendix

All parameters were kept the same as in our previous model (Chen & Mirman, 2012), except the following. First, the sigmoid inhibition function within the word layer was changed slightly to have a later "cross-over" point, thus allowing the model have more time to accumulate information to make a decision. This change reflects our intuition that decision-making is slower when information is integrated across two modalities. The sigmoid function used for all simulations reported here was:

$$y = \frac{7}{1 + e^{-\beta(x - x_0)}}$$

with $\beta = 13$ and $x_0 = 0.65$. This sigmoid scaling allows multiple word units to be activated initially (inhibition is weak when unit activation is low) and forces the model to rapidly settle to a single active word unit (rising activation causes a fast increase in inhibition strength; see also, Cisek, 2006).

The second change was to increase the small positive weights for feature units that sometimes occurred together and sometimes separately. In the original model, this value was 0.002, and for the current simulations the value was 0.018. This weight was increased because we were testing a more complex interaction of semantic and phonological neighbors, so this semantic weight needed to also balance the phonological weights—a constraint that was not relevant in our original model. Note, however, that with these two changes the model still replicated all results reported with our original model (Chen & Mirman, 2012).