

# The Dynamics of Lexical Competition During Spoken Word Recognition

James S. Magnuson<sup>a,b</sup>, James A. Dixon<sup>a</sup>,  
Michael K. Tanenhaus<sup>c</sup>, Richard N. Aslin<sup>c</sup>

<sup>a</sup>*Department of Psychology, University of Connecticut*

<sup>b</sup>*Haskins Laboratories, New Haven, Connecticut*

<sup>c</sup>*Department of Brain and Cognitive Sciences, University of Rochester*

Received 29 August 2005; received in revised form 25 May 2006; accepted 25 June 2006

---

## Abstract

The sounds that make up spoken words are heard in a series and must be mapped rapidly onto words in memory because their elements, unlike those of visual words, cannot simultaneously exist or persist in time. Although theories agree that the dynamics of spoken word recognition are important, they differ in how they treat the nature of the competitor set—precisely which words are activated as an auditory word form unfolds in real time. This study used eye tracking to measure the impact over time of word frequency and 2 partially overlapping competitor set definitions: onset density and neighborhood density. Time course measures revealed early and continuous effects of frequency (facilitatory) and onset-based similarity (inhibitory). Neighborhood density appears to have early facilitatory effects and late inhibitory effects. The late inhibitory effects are due to differences in the temporal distribution of similarity within neighborhoods. The early facilitatory effects are due to subphonemic cues that inform the listener about word length before the entire word is heard. The results support a new conception of lexical competition neighborhoods in which recognition occurs against a background of activated competitors that changes over time based on fine-grained goodness-of-fit and competition dynamics.

*Keywords:* Psycholinguistics; Spoken word recognition; Eye tracking

---

## 1. Introduction

Recognizing spoken words requires listeners to solve several perceptual challenges. Unlike written words or other visual objects, their components cannot be simultaneously presented; they do not persist in time, and so cannot be reexamined after their initial presentation. Instead, a series of transient acoustic events extending over a few hundred milliseconds must be mapped onto words in memory. This mapping must be rapidly achieved (conversational

---

Correspondence should be addressed to James S. Magnuson, Department of Psychology, University of Connecticut, 406 Babbidge Road, Unit 1020, Storrs, CT 06269–1020. E-mail: james.magnuson@uconn.edu

speech often reaches rates of 7 syllables per sec; Pollack & Pickett, 1964) without reliable cues to word boundaries (e.g., Cole & Jakimik, 1980). By analogy, imagine reading this page through a two-letter aperture as the text scrolled past, without spaces separating words, at a variable rate you could not control. Under these circumstances, efficiency would be improved if candidate words were generated as the text is revealed, rather than waiting until chunks of text larger than the aperture match precisely with stored lexical entries; indeed, if word boundaries are not marked and the entire stream is not or cannot be held in memory, there is no other way to segment the text, aside from finding lexical matches on the fly.

Therefore, the computational demands of spoken word recognition (SWR) require that those lexical representations that are acoustically similar to the unfolding input be partially activated, both to serve as a temporary memory of the input and to serve as a set of candidate hypotheses. Determining the nature of these so-called lexical neighborhoods is important for both practical and theoretical reasons. From an applied perspective, tests that take into account the effects of lexical neighborhoods are proving useful as measures of the efficiency of SWR in populations with hearing impairments (Kirk, Diefendorf, Pisoni, & Robbins, 1997; Sommers, Kirk, & Pisoni, 1997) and in devices that attempt to recognize speech by algorithm (Jurafsky & Martin, 2000). From a theoretical perspective, a proper definition of lexical neighborhoods will provide crucial constraints on models of SWR and underlying neural mechanisms. In this article, we use eye movements to measure neighborhood effects as a word unfolds over time. We show that the competitor set changes dynamically as a word is heard, with competitors that share onsets dominating early in the recognition process, and effects of global similarity emerging later.

Current models of SWR make different assumptions about which lexical competitors are activated as a word unfolds. Some models emphasize global similarity (Luce & Pisoni, 1998), whereas others emphasize onset-based similarity (Marslen-Wilson & Welsh, 1978; Norris, 1994). Models that emphasize onset-based similarity maximize the speed with which a lexical candidate is selected by activating a set of candidates that initially match the input, so-called *cohort* competitors, and strongly inhibiting candidates as soon as they mismatch the input. For example, *bat* will not be activated when the input is *cat*; but words like *cab*, *cattle*, *cavern*, and *catatonic* will be activated.

Evidence for onset-based neighborhoods comes from studies that have used priming to examine competitor activation (e.g., Marslen-Wilson & Zwitserlood, 1989). Such studies have consistently found strong evidence for cohort activation (*cat* primes *taxi*, an associate of *cab*), but little evidence for activations of words that mismatch at onset, such as rhymes (*cat* will not prime *vampire*, an associate of *bat*; but for trends toward priming when rhymes differ from targets by a single phonetic feature, see Andruski, Blumstein, & Burton, 1994; Connine, Blasko, & Titone, 1993; Marslen-Wilson, 1993).

The most influential model that emphasizes global similarity is the Neighborhood Activation Model (NAM; Luce, 1986; Luce & Pisoni, 1998). NAM predicts words will be activated by a spoken word (e.g., *cat*) when they differ by no more than one phoneme from the input<sup>1</sup> (whether by addition, deletion, or substitution; e.g., *cast*, *scat*, *at*, *bat*, *cot*, *cab*), whereas words that overlap at onset but then differ by several phonemes, such as *cavern* or *catatonic*, will not be activated. Models like NAM, which only make use of positive evidence (activation given full or partial phonemic matches, but no explicit mismatch inhibition), yield neighborhoods

based primarily on global similarity. The primary evidence for NAM comes from studies in which the make-up of the lexical neighborhood is inferred from how long it takes to recognize a word. Recognition time is predicted to be related to *frequency-weighted neighborhood probability*, which is the ratio of a word's frequency to the sum of its own frequency and the frequencies of its neighbors. The idea is that the more neighbors a word has, and the more frequently those neighbors occur, the harder that word will be to recognize. This measure provides the best prediction of recognition facility for large sets of words, accounting for approximately 20% of the variance (compared to about 5% for the next-best factor, word frequency; Luce & Pisoni, 1998).

It is important to note that the evidence for onset-based and global neighborhoods comes from different types of paradigms, each with strengths and weaknesses. Priming allows one to probe for a particular type of competitor, and it can be used to provide a snapshot of the activation of this competitor at different points in time after the prime is presented (as a function of the delay between prime and target). However, it does not easily lend itself to measuring the impact of entire neighborhoods on the recognition of a specific word (although one could measure priming to associates of an exhaustive set of competitors, and one might expect degree of priming to depend on competitor density).

Measures such as naming, lexical decision, or recognition in noise allow one to assess the global effects of competitor sets—that is, the impact on overall recognition—but do not provide information about how neighborhood effects might change as the word unfolds over time. Distinguishing among competing models of SWR requires a measure that is sensitive both to time course and to the overall effects of neighborhoods. Moreover, evaluation of competing models requires a metric that can distinguish neighborhoods with many onset competitors from neighborhoods with few.

The need for time-course measures is further highlighted by the results of Vitevitch (2002), who demonstrated that onset similarity has effects above and beyond neighborhood density. Vitevitch computed an “onset density” measure: the proportion of neighbors that overlapped in the first phoneme with the target. He compared words with a high proportion of onset neighbors (75.3%) with words with a low proportion (42%). Words with high-onset density neighborhoods were named more slowly (by 11 msec on average) and recognized more slowly in a lexical decision task (by 23 msec on average) than words with low-onset density neighborhoods. Because the onset metric was limited to neighbors (e.g., *cat*, *cab*, *cap*), it is not clear whether these results would generalize to onset competitors in general (e.g., *cat*, *cabin*, *cannibal*), or if it is limited to the influence of neighbor onsets.

The foregoing review of models of SWR highlights the need for a detailed evaluation of the set of lexical competitors over time as the target word is unfolding, rather than characterizing the competitor set as the set of words that should have been active at any point as the word was heard. This requires not only a dependent measure that has excellent temporal dynamics, but also a careful comparison of different metrics used to compute the acoustic–phonetic similarity of lexical items. In this study we compare neighborhood density<sup>2</sup> as defined by NAM with a conceptually parallel onset-based measure: frequency-weighted cohort density (i.e., the summed log frequency of a target word and all its cohorts). We chose this metric rather than the neighbor onset metric used by Vitevitch (2002) because that metric excludes many of the items predicted to compete most strongly by cohort-style models (i.e., it considers onset density only

for neighbors). Because words are neighbors if they differ by no more than one phoneme (accomplished by addition, deletion, or substitution of a phoneme), words such as *cask*, *castle*, *cabin*, *captain*, *café*, *camera*, and so forth, would not be counted by NAM as competitors of *cat*, whereas all of these words would be competitors of *cat* according to cohort density models. The different types of competitors are illustrated in Fig. 1. The two large ovals indicate neighbors and cohorts. The grey region identifies items that fit the criteria for both neighbors and cohorts. The dashed oval indicates the items that would be included in the Vitevitch onset metric.

To assess the time course of lexical competition as a function of neighborhood density and cohort density, we adapted the visual world eye-tracking paradigm (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) in a design similar to that used by Allopenna, Magnuson, and Tanenhaus (1998). Allopenna et al. monitored eye movements as participants followed spoken instructions to pick up and move one of four items displayed on a screen using a computer mouse (“pick up the beaker”). Critical trials included cohorts (e.g., *beetle*), rhymes (*speaker*), or both; as well as unrelated baseline items (e.g., *carriage*). The proportions of fixations to the displayed items mapped directly onto phonetic similarity over time: Targets and cohort competitor proportions increased and separated from the rhyme and unrelated baseline early on; as the input became more similar to the rhyme, its proportion separated from the unrelated baseline. Shortly after information disambiguating the target and cohort was available in the input, fixation proportions to the cohort began to drop off, returning to the unrelated baseline sooner than to the rhyme. Simulations of these data using TRACE (McClelland & Elman, 1986) accounted for nearly 90% of the variance in the time course fixation proportions to the target and competitors. TRACE adopts a middle ground between models that emphasize onset similarity and models that emphasize global similarity. TRACE does not include explicit mismatch inhibition. Therefore, similarity at any point can activate any word. However, lateral inhibition between words leads to an advantage for items overlapping at onset: Because they get activated early on, they inhibit items that are activated later, such as rhymes.

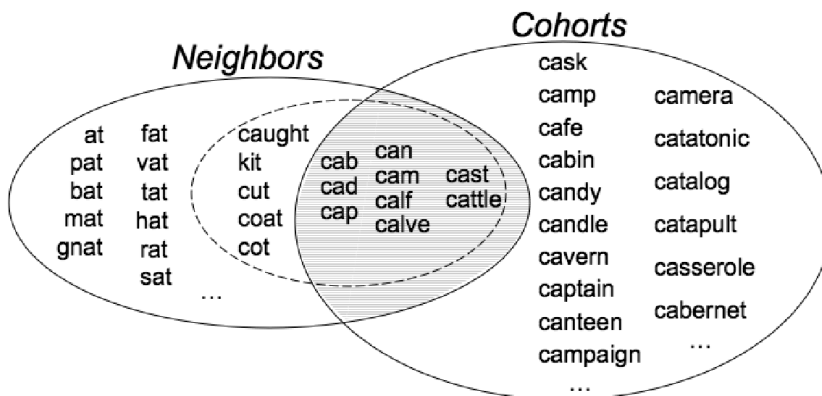


Fig. 1. Example of phonological neighborhood and cohort for the word *cat*. Neighbors are defined by a mismatch criterion: They differ from *cat* by no more than one phoneme. The cohorts are defined by a match criterion: They have the same onset as *cat*. The grey region of overlap indicates the items that meet both criteria. The dashed oval contains the items that would be included in the Vitevitch (2002) onset definition.

Although the Allopenna et al. (1998) paradigm provides fine-grained time course information, it suffers from the same limitation as priming measures (viz., it only measures activation for displayed items). In this study, we used conditions more similar to those in lexical decision or naming tasks. In our task using the visual world paradigm, we examined the recognition of single monosyllabic words that varied in frequency and competitor statistics by displaying a target picture among three unrelated distractors. In contrast to lexical decision studies, we were able to estimate the time course of lexical activation by measuring eye movements as target words were presented. Although competitors were never present in the same display as the target, we are able to infer how the set of activated lexical candidates changes as a word is heard by relating the time course of fixations to the target given the characteristics of nondisplayed competitors. Our results revealed that early in the word effects of word frequency and cohort density dominated, with global neighborhood effects emerging later in the recognition process. These results place strong constraints on models of SWR and highlight the importance of temporal dynamics in defining and measuring the effects of lexical neighborhoods.

## 2. Experiment

### 2.1. Method

#### 2.1.1. Participants

Fifteen native speakers of English who reported normal or corrected-to-normal vision and normal hearing were paid for their participation.

#### 2.1.2. Stimuli

The auditory stimuli consisted of 128 imageable, monosyllabic, English nouns. There were two levels (high and low) of frequency, neighborhood density, and cohort density. There were 16 items in each of the eight combinations of these levels. All of the items, along with several lexical characteristics, are included in the Appendix. Table 1 shows the mean levels of each factor in each condition (item-specific values are available from James S. Magnuson). Table 1 also includes four other measures that have previously been shown to have strong influences on SWR.

The first is duration. Although there are trends that correlate with some of the manipulations, none are reliable ( $p$  ranges from .12–.41), and the direction of trends varies (low frequency = 546, high = 538; low neighborhood = 549, high = 535; low cohort = 534, high = 550). The next two are the Vitevitch and Luce (1998) measures of phonotactic probability. *Phone psum* are the summed positional probabilities for each phoneme in a word (i.e., the independent probability of a phoneme occurring in its word position), and *Biphone psum* are the summed biphone probabilities. The fourth is a *max cohort density* onset metric similar to that used by Vitevitch (2002), where the criterion for cohort status is overlap in the first phoneme (although the set is not limited to neighbors). We used a two-phoneme definition because we expect it provides a closer analog to the original cohort model notion of cohorts overlapping in the first 150 to 200 msec. As can be seen in Table 1, the two cohort definitions give somewhat similar results ( $r^2 = .18$ ). We leave as a question for future research which of these measures provides the better estimate of onset-based competition.

Table 1  
Lexical statistics of the materials

Frequency	Neighborhood	Cohort	Log Frequency	Neighborhood Density	Cohort Density	Duration	Phone pSum	Biphone pSum	Max Cohort Density
Low	High	High	2.5	<b>101.8</b>	<b>307.7</b>	558	<b>0.203</b>	<b>0.015</b>	<b>2025.5</b>
High	High	High	<b>4.8</b>	<b>95.7</b>	<b>272.9</b>	538	<b>0.198</b>	<b>0.017</b>	<b>1874.1</b>
Low	High	Low	2.4	<b>99.6</b>	46.6	525	0.151	0.007	1367.2
High	High	Low	<b>4.5</b>	<b>108.7</b>	50.6	519	0.145	0.005	1266.9
Low	Low	High	2.1	25.0	<b>289.2</b>	552	0.15	<b>0.014</b>	<b>2217.3</b>
High	Low	High	<b>4.6</b>	26.1	<b>286.1</b>	553	<b>0.221</b>	<b>0.017</b>	<b>2410.6</b>
Low	Low	Low	2.2	23.6	43.3	550	0.158	0.008	1222.9
High	Low	Low	<b>4.8</b>	29.2	48.8	541	0.162	0.008	1490.2

*Note.* Frequency, neighborhood density, and cohort density were manipulated. Duration, probabilistic phonotactic measures (phone pSum = the sum of positional phoneme probabilities; biphone pSum = the sum of biphone positional probabilities), and max cohort density (with cohort operationalized as all items sharing the first segment with a target) are included to show that they do not provide additional explanatory power for these materials. High levels of each characteristic are shown in bold.



As can be seen in Table 1, the phonotactic probability measures also pattern with cohort density. This is somewhat surprising given their expected relation with neighborhood density. However, the phone measure correlates more strongly with cohort density ( $r^2 = .16$ ) than with neighborhood density ( $r^2 = .06$ ), as does the biphone measure ( $r^2 = .60$  with cohort density,  $r^2 = .23$  with neighborhood density). Given that all three measures (phone psum, biphone psum, and max cohort density) pattern with cohort density, we provisionally conclude that for these materials, they do not provide important information beyond that provided by our measures of interest (frequency, neighborhood density, and cohort density). The fact that biphone probability correlates with both neighborhood and cohort density might raise the concern that biphone probability is a proxy for both. However, as will become clear shortly, simple effects of neighborhood are found when cohort density is controlled and vice versa.

The auditory stimuli were produced by a male native speaker of English in a sentence context ("Click on the chef"). The stimuli were recorded using a Kay Lab CSL 4000 with 16-bit resolution and a sampling rate of 22.025 kHz. The mean duration of the "Click on the ..." portion of the instruction was 440 msec. Mean target duration was 538 msec.

The visual stimuli consisted of pictures of the 128 targets and 412 distractors. These came from a variety of sources, including the Snodgrass and Vanderwart (1980) pictures and a number of clip-art collections. We allowed as little variability as possible in realism, style, and other characteristics. The pictures are available on request from James S. Magnuson.

### 2.1.3. Procedure

All 128 targets were presented in random order. Three visual distractors were chosen pseudo-randomly for every trial for every participant with the following constraints: Distractors could only appear in one trial per participant, distractors could not be cohorts or neighbors of each other or the target, and distractors could not be closely semantically related to each other or to the targets. Each picture was coded on 178 semantic classes such as *person*, *animal*, *vehicle*, *appliance*, *tool*, and *medical*. Items typically had two to four semantic codings (e.g., *bull* had *animal* and *farm*). Only one item from each class was permitted to appear in each display.

On each trial, the pictures appeared 100 msec after the participant clicked on a central fixation square. Concurrently, the auditory instruction began (e.g., "click on the yarn"). The trial ended 150 msec after the participant clicked on one of the pictures. The pictures were about  $1.5^\circ$  in diameter and were located approximately  $2^\circ$  from the central fixation square along the  $45^\circ$ ,  $135^\circ$ ,  $225^\circ$ , and  $315^\circ$  axes from the center of the display.

Eye movements were monitored using a SensoMotorics Instruments (SMI) EyeLink eye tracker, which provided a record of point-of-gaze in screen coordinates at a sampling rate of 250 Hz. The auditory stimuli were binaurally presented through headphones (Sennheiser HD-570) using standard Macintosh Power PC digital-to-analog devices. Saccades and fixations were coded from the point-of-gaze data using SMI's software.

## 2.2. Predictions

The predictions for main effects are straightforward because they are consistent across competing models and should mirror those found using other methods. First, there should be an ad-

vantage for high-frequency items compared to low (e.g., Howes, 1957; Savin, 1963), reflected in a steeper rise in target than nontarget fixation proportions (for frequency effects using the eye tracking paradigm, see Dahan, Magnuson, & Tanenhaus, 2001). Second, there should be an advantage for items with low neighborhood density compared to those with high (Luce & Pisoni, 1998) because neighbors should compete with a strength proportional to their own frequency (for neighborhood density effects on fixation proportions, see Magnuson, Tanenhaus, Aslin, & Dahan, 2003). Third, the same logic should hold for our cohort density measure; that is, there should be an advantage for items in low-density cohorts compared to those in high-density cohorts. Of primary interest is whether and how effects of cohort density and neighborhood density change as the target word unfolds over time.

### 2.3. Results

Fig. 2 shows the patterns of fixation for the main effects of frequency, neighborhood density, and cohort density. In each case, fixation proportions begin to depart from chance levels (.20, given 4 objects and the central fixation square) around 200 msec after target onset. As in previous studies (e.g., Allopenna et al., 1998), changes in fixation proportion were closely time locked to the speech signal, as it takes at least 150 msec to plan and launch an eye movement; and in tasks like ours, typical intersaccadic intervals are in the range of 200 to 300 msec (see Fischer, 1992; Saslow, 1967; Viviani, 1990). Unless otherwise noted, all analyses in this article are restricted to the window from 200 msec (the earliest point where we expect signal-driven changes in fixation proportions) to 1,000 msec (the point by which fixation proportions tend to asymptote in several studies using this technique). Predicted trends were observed for frequency (high-frequency advantage) and cohort density (low-density advantage), but the pattern for neighborhood density is strikingly different: The predicted low-density advantage emerges late and is preceded by a high-density advantage. This change in the neighborhood effect over time presents significant statistical challenges.

In the relatively brief history of using eye movements as an index of spoken language processing, a variety of statistical approaches have been used, but nearly all have used standard analyses of variance. A common approach has been to include time as a factor, typically by dividing time into some number of bins (e.g., Allopenna et al., 1998, used 8 successive 100-msec windows). However, because the value (i.e., fixation proportion) for some condition at time  $t$  is not independent of its value at time  $t-1$ , this approach violates the analysis of variance (ANOVA) assumption of independent observations. Another approach is to avoid this violation by finding a way to capture differences between conditions without including time, such as calculating average fixation proportion for each condition (Magnuson et al., 2003). This leaves us with a dilemma in approaching the results of this study, as the effect of neighborhood clearly changes over time. This led us to techniques developed specifically for examining change over time: growth curve analysis, particularly as it is applied in developmental psychology (Singer & Willett, 2003).

#### 2.3.1. Growth curve analysis

Growth curve analysis is a method for formally modeling variations in over-time trajectories. Conceptually, it is analogous to fitting separate, over-time regression models for each in-



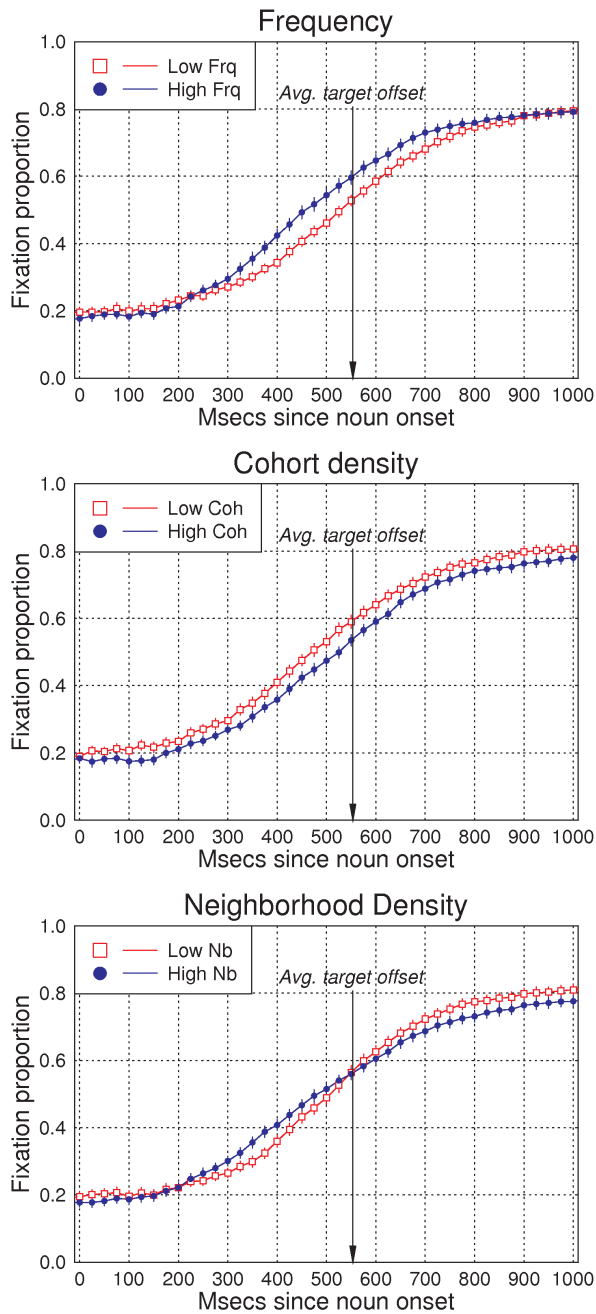


Fig. 2. Fixation proportions over time for frequency, cohort density, and neighborhood density conditions. Bars represent standard error.

dividual and then analyzing the resultant parameters as a function of some set of predictors (although the particular instantiation of the model accomplishes this through different means). In standard growth curve modeling, the over-time trajectory of the dependent measure is represented using general linear model parameters that define the trajectory (e.g., the intercept and slope), and time-invariant predictor variables impact the model through those parameters.

In this study, we used orthogonal power polynomials to capture the curvilinear form of the relation between the fixation proportions and time.<sup>3</sup> Specifically, we modeled fixation proportion as a function of polynomials that captured the linear, quadratic, and cubic effects of time.<sup>4</sup> We need these three parameters to fit the sigmoidal form typically found with fixation proportions over time (the quadratic term can describe a constant rate of change in slope—i.e., a single curve—whereas the cubic term can capture changes in that rate of change itself over time, which allows it to describe sigmoidal forms). Fig. 3 shows the main effect data with fitted growth curves superimposed. For each effect, we report the estimated parameters, standard errors, and the change in the deviance statistic resulting from adding the parameter to the model. The deviance statistic indexes the fit of the model (larger values indicate poorer fit). Change in the deviance statistic,  $\Delta D$ , is distributed as chi-square. Because the tests reported later all involve adding a single parameter to the model, the  $\Delta D$  tests are on 1 *df* (note that the critical chi-square value for 1 *df* is 3.84).<sup>5</sup> All three polynomial terms significantly contributed to the model:  $B$ s = 1.100, -2.083, -0.110;  $SE$ s = .028, .005, .005; for the linear, quadratic, and cubic components, respectively. The  $\Delta D$ s (1) were 313.8, 1064.2, and 369.4 ( $p < .001$  in each case), respectively. Items analyses correspond very closely to analyses by participants (see Table 2); therefore, in the interest of concision, we only report participant analyses. Note that the significance patterns are identical.

We added frequency, cohort density, and neighborhood density to the model by including the effects of these variables on the intercept and the linear time variable. The effect on the intercept tests whether the curve is shifted up as a function of condition, analogous to a main effect in ANOVA (e.g., on mean fixation proportion). The effect on the linear time variable tests whether the “slope” (i.e., the linear component of the trajectory) differs by condition. A significant effect of slope indicates that the trajectories change at different rates, which would be analogous to an interaction of condition and time (because differences in rate of change would lead to differences of varying magnitude at different points in time in a variable like mean fixation proportion).

*2.3.1.1. Frequency.* Frequency significantly affected the intercept,  $B = 0.037$ ,  $SE = .018$ ,  $\Delta D(1) = 4.00$ ,  $p = .046$ ; but not the slope,  $B = -0.060$ ,  $SE = .054$ ,  $\Delta D(1) = 1.20$ , *ns*. This confirms that the advantage for high word frequency apparent in the top panel of Fig. 2 was reliable, and the advantage did not depend on time (i.e., its magnitude did not change significantly in the analysis window).

*2.3.1.2. Cohort density.* There was also a significant effect of cohort density on the intercept,  $B = -0.040$ ,  $SE = .018$ ,  $\Delta D(1) = 4.74$ ,  $p < .029$ ; low-cohort density showed the predicted advantage, but there was no effect on the slope,  $B = 0.019$ ,  $SE = .054$ ,  $\Delta D(1) = .098$ , *ns*. This indicates that there was a reliable advantage for low-cohort density items (see the middle panel of Fig. 2) that did not change significantly over time.

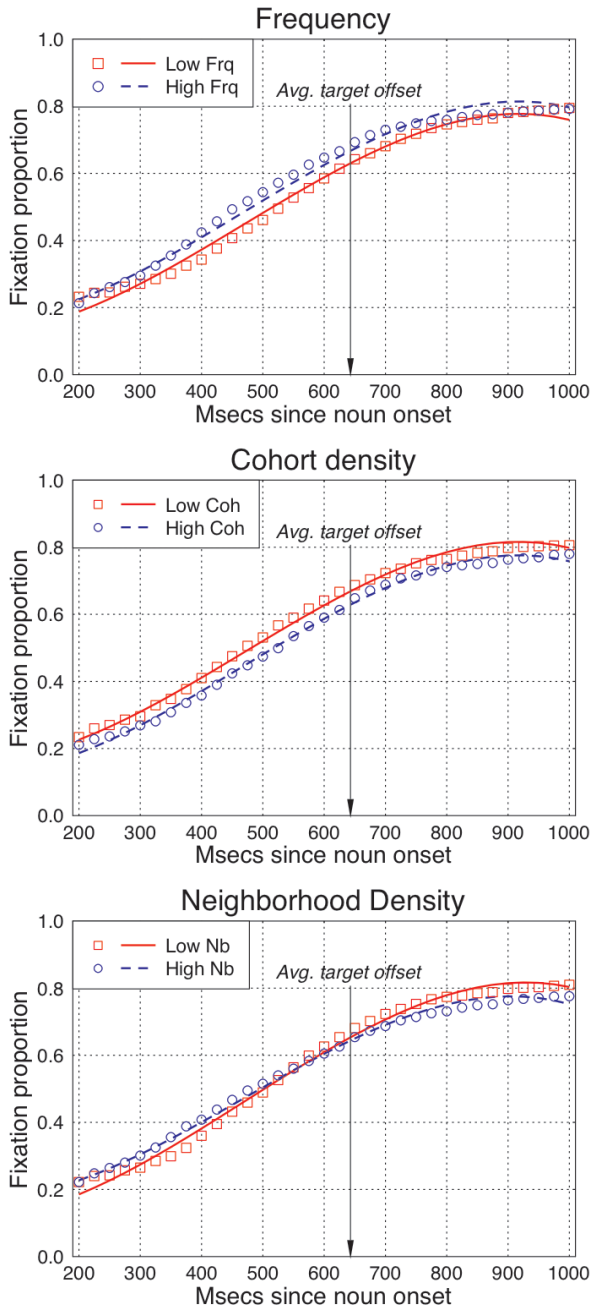


Fig. 3. Data from Fig. 2 with predictions from growth curve analyses superimposed; the graphs are restricted to a range of 200 msec to 1,000 msec as that is the range to which the models were applied.

Table 2  
Growth curve results for subject and item analyses

Effect	Subjects Analysis			Items Analysis	
	Estimate	SE	p (Chi Square)	Estimate	SE
Intercept	0.570	0.018	< .001	0.570	0.020
Linear (~slope)	1.203	0.054	< .001	1.208	0.056
Quadratic	-0.208	0.006	< .001	-0.201	0.056
Cubic	-0.111	0.006	< .001	-0.111	0.007
Frequency (intercept)	0.037	0.018	.046	0.037	0.020
Neighborhood (intercept)	-0.005	0.018	<i>ns</i>	-0.003	0.020
Cohort (intercept)	-0.040	0.018	.029	-0.047	0.020
Frequency (slope)	-0.061	0.054	<i>ns</i>	-0.050	0.057
Neighborhood (slope)	-0.159	0.054	.004	-0.161	0.056
Cohort (slope)	0.020	0.054	<i>ns</i>	0.019	0.056

*Note.* The effects of frequency, neighborhood, and cohort on intercept would be analogous to main effects on a measure like mean fixation proportion over the entire window of analysis. An effect on slope would be analogous to an interaction with time (due to significant change in magnitude or direction of a condition difference over time).

2.3.1.3. *Neighborhood density.* Neighborhood density did not have a reliable effect on the intercept,  $B = -0.004$ ,  $SE = .018$ ,  $\Delta D(1) = .098$ , *ns* (analogous to a null main effect, e.g., of mean fixation proportion, in an ANOVA). However, it did affect the slope,  $B = -0.159$ ,  $SE = .054$ ,  $\Delta D(1) = 8.23$ ,  $p = .004$ . Items with a lower neighborhood density showed steeper gains over time. As mentioned earlier, this is analogous to an ANOVA interaction between neighborhood density and time. As can be seen in the bottom panel of Fig. 2, there was an initial, unexpected advantage for high neighborhood density items that shifted to the expected low-density advantage in the latter portion of the time course. Given that this is analogous to a crossover interaction of neighborhood density with time (provisionally, early vs. late time course), a question arises: How can we explain the interaction statistically and theoretically?

Within growth curve analysis, techniques have been developed to assess whether the parameters underlying an over-time trajectory change or remain stable as a function of some external event (e.g., marriage, a new treatment) that occurs during the time course. However, to employ these techniques we must have a theory of the underlying cause of the change in rate to specify points in time where we predict a change.

The crossover from an early advantage for items in high-density neighborhoods to a late advantage for low-density items is surprising, given that our high- and low-density neighborhood items were matched on both frequency and cohort density. One possibility we entertained (suggested by Liina Pykkänen) is that the crossover reflects an initial benefit of high phonotactic probability correlated with high neighborhood density that is followed by an inhibitory neighborhood effect as lexical representations reach sufficient levels of activation to inhibit other lexical nodes significantly. Although this is a plausible explanation of the neighborhood pattern, it incorrectly predicts a similar *initial* disadvantage for high-density cohorts (i.e., an early difference between conditions has to be driven primarily by the initial segments of the target word). If high probability patterns in general have early facilitatory effects,

high-cohort density should also result in an early advantage; but even the earliest effects of cohort density are inhibitory. Our explanation has to do with the distribution of similarity over time within neighborhoods. Recall that neighbors and cohorts can overlap (see Fig. 1). Two words equated on cohort density might have quite different neighborhoods, with different amounts of overlap between their neighborhoods and cohorts. With this in mind, we computed the proportion of neighbors that were also cohorts for each condition (like Vitevitch's [2002] *onset density* measure, but operationalized as items sharing the first two phonemes; Magnuson [2001] introduced this cohort density measure, and also examined the proportion of neighbors that were also cohorts).

On average, the proportion of neighbors that were also cohorts was much higher for low-density neighborhoods (66%, with those items accounting for 67% of frequency weighted neighborhoods) than for high-density neighborhoods (37%, and 36% of the frequency weight of those neighborhoods; see Table 3 for details in all conditions). This explains the late advantage for low-density neighborhood items in the bottom panel of Fig. 2. In terms of the temporal distribution of similarity between a target word and its cohorts and neighbors, low-density neighborhoods were front loaded—most of their neighbors would be activated near word onset. Later in the word, their competition neighborhoods are relatively exhausted—the point of greatest neighborhood overlap has passed.

What about the early advantage for high neighborhood density items? If a higher proportion of low-density neighbors are cohorts, this suggests that there are differences in average cohort length (i.e., that more of the cohorts of low-density items are short, allowing them to fit the definition of neighbor). This is true: 65% of high-density items' cohorts were longer than one syllable (contributing 55% of the frequency weighted cohort density), whereas 47% of low-density items' cohorts were longer than one syllable (contributing 38% of the frequency weighted cohort density).

Greater average cohort length could only explain the early high-density neighborhood advantage if listeners have access to cues to word length as they hear a word onset. Such cues exist: On average, vowel durations are longer in monosyllabic than multisyllabic words (Lehiste, 1972), with these differences increasing for strong positions in a prosodic domain (Ladd &

Table 3

Percentage of neighbors that are also cohorts in each condition, and percentage of "long" (more than 1 syllable) cohorts

Frequency	Neighborhood Density	Cohort Density	% Neighbors That Are Cohorts	% Neighbor		% Cohort Density Due to Long Cohorts
				Frequency Due to Cohorts	% Long Cohorts	
Low	High	High	48	46	82	76
High	High	High	48	50	68	57
Low	High	Low	27	23	58	48
High	High	Low	25	25	53	40
Low	Low	High	78	77	36	31
High	Low	High	81	81	44	35
Low	Low	Low	54	54	52	45
High	Low	Low	53	53	56	40

Campbell, 1991; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Four recent studies suggest listeners are indeed sensitive to these differences. First, there is greater priming between words of the same length (Davis, Marslen-Wilson, & Gaskell, 2002). Second, the time course of lexical access is altered significantly by approximately 20 msec differences in Dutch words between syllables analogous to the word *ham* and the first syllable in the word *hamster* (Salverda, Dahan, & McQueen, 2003). Third, when the target word is in utterance final position as it was in this study (e.g., *Click on the cap*), a monosyllabic cohort competitor such as *cat* is a stronger competitor than a multisyllabic competitor such as *captain*, although phonemic overlap is greater for the multisyllabic competitor (Salverda et al., in press). Finally, when all else is held constant, monosyllabic words with primarily monosyllabic cohorts are recognized more slowly than monosyllabic words with primarily multisyllabic cohorts, and vice versa (Magnuson & Strauss, 2006). Another way to put this is that because of durational differences, the relative goodness of fit of (*cap*, *cat*) is higher than that of (*cap*, *captain*). Therefore, in the initial consonant-vowel (CV)—the part of the word where low- and high-density neighborhood items are matched for cohort density—high-density items are at an advantage because their cohort competitors are longer, on average, than those of the low-density items; meaning their average cohort goodness of fit is lower, which results in less cohort competition.

The cohort proportions and cohort lengths for every condition are shown in Table 3. Two things stand out. First, cohort proportion is correlated with cohort density—it is higher in high-cohort density conditions. Second, we might expect a stronger influence of cohort density for low-density neighborhoods compared to high-density neighborhoods. At both levels of neighborhood density, a higher proportion of neighbors are cohorts at the high level of cohort density than at the low level. However, at the low level of neighborhood density, the majority of neighbors are cohorts at the high level of cohort density (about 75%). This means these neighborhoods are heavily “front loaded”: The majority of competitors overlap at onset.

Fig. 4 provides a schematic of the hypothesized impact of these relations in the competitor sets over time for items in high- and low-density neighborhoods. The competitors are broken into three groups: noncohort neighbors, short cohorts (which are also neighbors), and long cohorts. For low-density items, most neighbors fall into the short cohort group, which have a relatively large, early impact. The long cohorts are fewer and have lesser goodness of fit, and so make a smaller contribution to summed competitor activation. The noncohort neighbors are relatively few (only 1/3 of the neighborhood) and have a late, weak impact. By comparison,

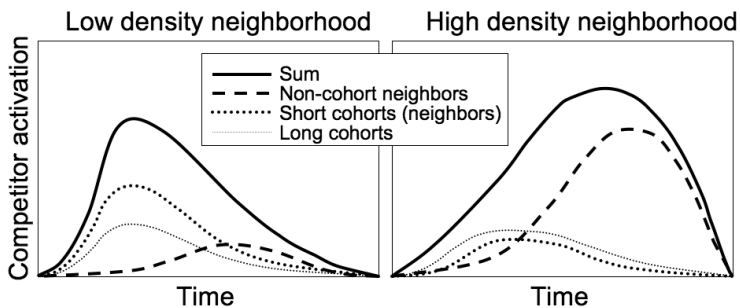


Fig. 4. Schematic of hypothesized changes in the impact of different types of competitors for low- and high-density neighborhood items used in this experiment.



most of the high-density items' cohorts are long (right panel). Despite their relatively large number, they are hypothesized to have a fairly weak impact. Their relatively few short-cohort neighbors have a modest impact. Most of their neighbors, instead, have a late impact (a larger impact of noncohort neighbors is anticipated both because there are more, but also because these items have denser neighborhoods).

We hypothesize that the differences in slope seen in the bottom panel of Fig. 2 are driven by the dynamically changing makeup of the competitor sets for these two conditions. The large, early competitor activation for low neighborhood items slows target activation initially, but once those items can be inhibited (once coarticulatory cues as to the second consonant are available "within" the vowel), little competition remains. In contrast, target activation is initially rapid for high-density items because the competitor set remains sparse, but then is impeded once noncohort neighbors are activated.

On this explanation, the vowel becomes the anchor for the switch from the early high-density advantage to the late low-density advantage. On the one hand, around vowel offset, cohort items can begin to be inhibited based on bottom-up mismatch. On the other, around vowel offset, noncohort neighbors begin to receive substantial bottom-up support. On average, vowel offset in our materials was approximately 330 msec after word onset. This provides the theoretical motivation required to explore the significant effect of neighborhood density on slope. As we noted earlier, it takes at least 150 msec to plan and launch an eye movement; and in tasks like ours, typical intersaccadic intervals are in the range of 200 to 300 msec. This means the earliest we would expect to see changes linked to average vowel offset would be approximately 530 msec after word onset, which is very close to the actual crossover in the bottom panel of Fig. 2 (550 msec).

For simplicity, we divided the time course into two epochs at the crossover point (but note that moving the dividing point back or forward as much as 100 msec changes neither the trends nor the patterns of significance we are about to report), which we call *pre-vocalic* and *post-vocalic* (by which we mean more precisely, pre-vowel offset and post-vowel offset). We created a variable that linearly indexed time within each epoch. Of interest was whether the effect of neighborhood density on slope was driven solely by the post-vocalic differences as opposed to both pre and post differences. To test this hypothesis, we included the effect of neighborhood density on pre-vocalic slope and post-vocalic slope. Both terms contributed significantly to the model,  $B_s = -0.138, -0.179$ ;  $SEs = .058, .058$ ;  $\Delta Ds(1) = 5.50, 9.00$ ;  $p_s = .019, .003$ , respectively; thus, both pre- and post-onset differences contribute to the overall effect of neighborhood density on slope (this is analogous to testing the simple effect of neighborhood in each epoch and finding that both the initial high-density advantage and the late low-density advantage are reliable).<sup>6</sup>

Further consideration of Fig. 2 and Table 3 suggests two additional predictions about the interaction of cohort density and neighborhood density. First, we should see a stronger effect of cohort density at the low level of neighborhood density than at the high level (because the majority of neighbors are cohorts at the high level of cohort density for the low level of neighborhood density). Second, we should see a stronger effect of neighborhood at the low level of cohort density than at the high level (because cohorts are less dominant at the low cohort density level).

The time course data in Fig. 5 confirm these predictions. As can be seen in the top row of Fig. 5, the effect of cohort density on intercept was stronger at the low level of neighborhood

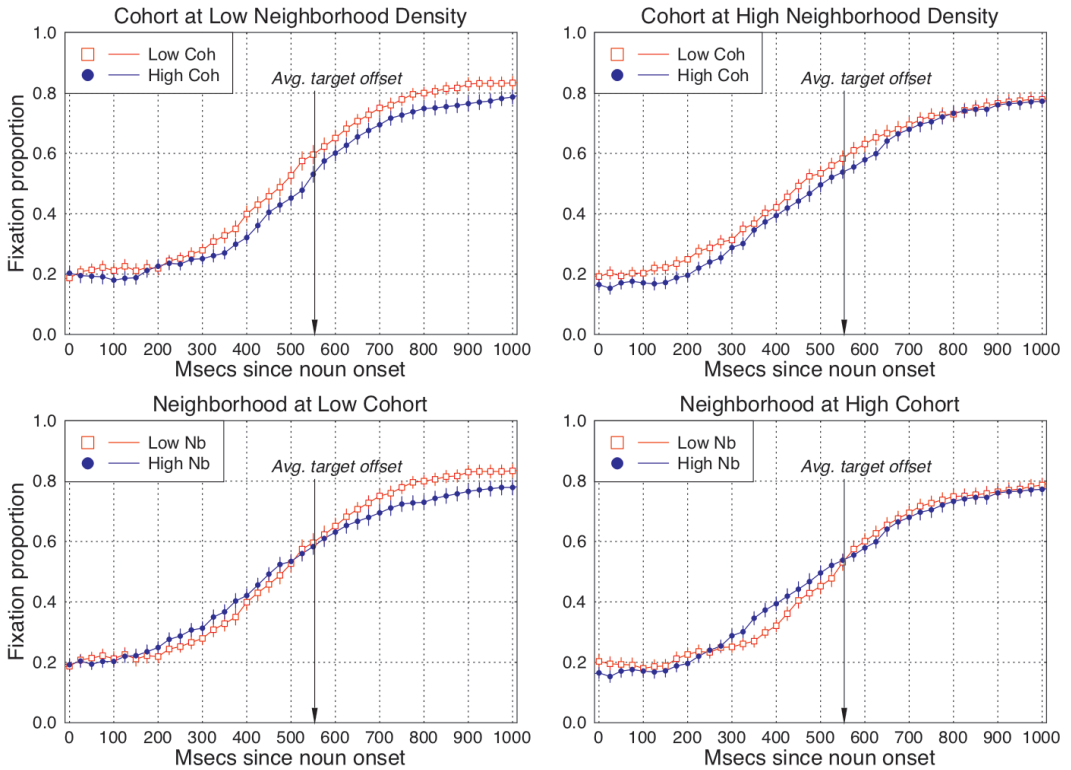


Fig. 5. Fixation proportions over time for simple effects of cohort density at each level of neighborhood density (upper panels) and simple effects of neighborhood density at each level of cohort density (lower panels). Bars represent standard error.

density than at the high level. The lower row in Fig. 5 shows the effects of neighborhood density at low and high levels of cohort density. As one would expect from the proportion of cohorts in low- and high-density neighborhoods, and the ratios of short and long cohorts, the disadvantage for low neighborhood density items was amplified at the high-density level of cohort, whereas the late advantage for low-density neighborhood items was amplified at the low-density level of cohort.<sup>7</sup>

### 3. Discussion and conclusions

These results both replicate and extend standard findings in SWR. First, we find clear effects of frequency and neighborhood density. Second, we confirm that competitor density based on the summed frequencies of items overlapping at onset (initial 2 phonemes)—onset cohorts—have strong effects on word recognition (as shown in Fig. 2). Third, this study shows the utility of applying a NAM-style frequency weighted density statistic (Luce & Pisoni, 1998) to this competitor type. Fourth, our results establish that, when used with unrelated distracters, the eye-tracking paradigm can be used to map out the time course of neighborhood effects.

Fifth, and most important, this study allowed us to evaluate the time course of the impact of target and competitor characteristics as the target word unfolded. None of the other measures used in SWR, by itself, could capture the fact that the competitor set changes as a word is heard (beyond identifying simple characteristics like uniqueness point), let alone the way the competitor set changes over time.

The effect of cohort density, for example, is apparent from the earliest signal-driven fixation proportions (around 200 msec after word onset), but the advantage observed for items in low-density neighborhoods does not begin until about 600 msec after word onset (and there was even an early advantage for high-density neighborhoods in our materials). This is consistent with findings like those of Allopenna et al. (1998) and Magnuson et al. (2003), where earlier on as the target unfolds, stronger competition is observed between targets and cohorts than between targets and rhymes. The cohort density metric only takes into account words overlapping at onset, whereas neighborhood density typically includes many items that mismatch at onset; therefore, the temporal distribution of overlap defined by each competitor metric has the potential to be substantially different.

Indeed, our most important finding from these results is the crossover from an advantage for high-density neighborhoods to a low-density advantage. A growing body of results suggest that fixation proportions over time in tasks like ours reflect the time course of lexical activation, as fixation proportions map extremely closely onto time course predictions from models like TRACE (Allopenna et al, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001) and simple recurrent networks (Magnuson et al., 2003). An important implication of these results is that if we were to link the time courses shown in Fig. 2 to predicted responses in tasks like lexical decision, we would draw different conclusions depending on the speed of the lexical decisions. We would also miss important interactions with time.

Consistent with this possibility, Newman, Sawusch, and Luce (1997) found effects of neighborhood density on phoneme identification for “medium” latency responses but not for fast responses. We can relate this result to the time course of neighborhood density effects shown in the lower panel of Fig. 2; if a participant were to respond quickly—that is, prior to the point where the relative advantage of low-density items kicks in—we would expect to see no effect of neighborhood density (or, for our materials, extremely early responses might suggest a high-density advantage).

We explained the early high-density neighborhood advantage as a function of listener sensitivity to subphonemic details that provide cues as to the length of the word that is being heard even during the first few segments. This is consistent with several recent studies documenting such sensitivity (Davis et al., 2002; Magnuson & Strauss, 2006; Salverda et al., 2003; Salverda et al., in press), and implies that competitor metrics must be based on finer grained (subphonemic) goodness of fit.

These results also have methodological and theoretical implications for the broader study of spoken language comprehension. The methodological implication is that it is essential to examine the time course of spoken language processing or risk missing complex interactions as spoken words unfold in real time. The theoretical implications are threefold. First, the persistence of cohort density effects throughout the recognition process, and the late emergence of neighborhood density effects, is problematic for models that incorporate strong bottom-up mismatch; including the Cohort model (Marslen-Wilson & Warren, 1994), Shortlist (Norris,

1994), and Merge (Norris, McQueen, & Cutler, 2000), as well as the distributed model described by Gaskell and Marslen-Wilson (1999) and models that ignore the temporal distribution of similarity such as NAM (Luce & Pisoni, 1998). Note that none of the lexical statistics we controlled (frequency, cohort density, neighborhood density) nor the others we measured (phonotactic probabilities and max cohort density) can by themselves account for the cross-over in neighborhood effects. An explanation requires explicit consideration of the temporal distribution of similarity, as in our account of the proportion of neighbors that are also cohorts. Second, the overlapping and nonoverlapping aspects of cohort and neighborhood sets, and the relative weight of cohort and neighborhood density, require further examination if we are to improve on existing metrics of spoken word similarity (e.g., Luce & Pisoni, 1998). Third, our results are inconsistent with the notion of static neighborhoods or recognition cohorts. Rather, the set of activated competitors is dynamic, and a full understanding of how processing neighborhoods change as a word is heard is needed to adequately constrain theories of spoken word similarity, processing, and recognition.

## Notes

1. There are also more complex neighborhood metrics that evaluate phoneme-by-phoneme similarity based on segmental confusion probabilities (Luce & Pisoni, 1998) or positional similarity ratings (Luce, Goldinger, Auer, & Vitevitch, 2000). The two make similar predictions (although for more subtle predictions that follow from more complex metrics, see Luce et al., 2000), and the short-cut metric is often used (e.g., Newman, Sawusch, & Luce, 1997). Crucially, both metrics are global: *cab*, *bat*, and *cot* are all considered roughly equally good neighbors of *cat*, despite large differences in the temporal distribution of overlap (although under the more complex metrics, not all phonemes are considered equal—e.g., a change from /k/ to /b/ is not necessarily equal to a vowel change)—and the complex metrics take time into account in a somewhat roundabout way to the degree that the metric is based on position-specific similarity).
2. Note that the term *neighborhood density* is sometimes used to indicate simply the number of neighbors (e.g., Vitevitch, 2002). We use it to indicate *frequency weighted neighborhood density*—the summed log frequencies of all items in the neighborhood (i.e., the denominator of the frequency weighted neighborhood probability rule; cf. Newman, Sawusch, & Luce, 1997).
3. We avoided nonorthogonal power polynomials because their terms are highly collinear. In the analysis presented here, the polynomial terms are orthogonal—that is, they are chosen so that they both capture the functional form and isolate independent components that underlie the form. Because they share no variance, they can be entered into the model simultaneously. Given these polynomials, the intercept is located in the center of the time series, rather than at its more traditional location at the intersection of the curve and y axis. Relocating the intercept is common practice in growth curve modeling because it allows one to test for differences in elevation at particular time points (see Singer & Willett, 2003).

4. We briefly note that there are other models that might also be fruitfully applied to these data. One potential alternative would be to compare parameters from curve fitting (cf. the logistic power peak analyses applied to condition difference curves by Scheepers, Keller, & Lapata, in press). However, growth curve analysis provides an approach that is both simpler (compare the 11-parameter model used by Scheepers et al. to the 4-parameter model used here) and for which well-developed significance test procedures are available. Another alternative would be to use a growth curve model that is nonlinear in its parameters, such as the logistic, to capture the curvilinear nature of the over-time trajectory. We chose the power polynomial approach, rather than the logistic, because (a) it is a mature, well-understood methodology that is directly analogous to ordinary least squares regression and analysis of variance; and (b) interpreting the parameters of “truly” nonlinear models is more complicated because such models are not dynamically consistent (Keats, 1983). In essence, on such account the effect of any parameter depends on the values of the other parameters. Therefore, interpreting the value of any parameter is only sensible in the context of the other parameters. One implication is that the averages of individual participant parameters typically will not equal the parameters of the average data, which complicates interpretation. On our approach, parameters are dynamically consistent.
5. We initially analyzed this data using standard analyses of variance (ANOVAs) and mean fixation proportion from 200 msec to 1,000 msec as the dependent variable. The results converge nearly completely with the results we report here; but, as we have just discussed, ANOVAs are not appropriate for this sort of data.
6. For all the analyses reported earlier, we included random effects for the intercept and slope. This creates a specific, but reasonably flexible, error covariance structure. We explored more complex error covariance structures, and, whereas some of these alternatives slightly improved overall fit of the model, they did not substantively change the estimated parameters reported earlier. Therefore, we retained the simpler error structure.
7. Statistically, although there were trends for each of these simple effects, only the largest effects in Fig. 5 (effect of cohort at the low level of neighborhood density and the effect of neighborhood at the low level of cohort) were reliable. Therefore, we present this analysis as suggestive evidence that the finer grained trends in simple effects are consistent with predictions that follow from the makeup of competitor sets over time.

## **Acknowledgments**

This study was supported by National Science Foundation (NSF) Grant SBR-9729095, and National Institute of Deafness and Other Communication Disorders Grant DC-005071 to Michael K. Tanenhaus and Richard N. Aslin; an NSF Graduate Research Fellowship, a Grant-in-Aid of Research from the National Academy of Sciences through Sigma Xi, and National Institute of Deafness and Other Communication Disorders Grant DC-005765 to James S. Magnuson; and National Institute of Child Health and Human Development Grant HD-01994 to Haskins Laboratories.



We thank Anne Pier Salverda for insightful comments, and Liina Pyykkänen for helpful discussions and for suggesting the hypothesis that the neighborhood cross-over effect might stem from early phonotactic probability facilitation and later lexical competition.

## References

- Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.
- Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163–187.
- Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32, 193–210.
- 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, 84–107.
- 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, 317–367.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Tracking the time course of subcategorical mismatches: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534.
- Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218–244.
- Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 31–45). New York: Springer-Verlag.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition and blending in spoken word recognition. *Cognitive Science*, 23, 439–462.
- Howes, D. H. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*, 29, 296–305.
- Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Upper Saddle River, NJ: Prentice Hall.
- Keats, J. A. (1983). Ability measures and theories of cognitive development. In H. Wainer & S. Messick (Eds.), *Principles of modern psychological measurement: A festschrift for Frederick M. Lord* (pp. 81–101). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Kirk, K. I., Diefendorf, A. O., Pisoni, D. B., & Robbins, A. M. (1997). Assessing speech perception in children. In L. L. Mendel & J. L. Danhauer (Eds.), *Audiologic evaluation and management and speech perception assessment* (pp. 101–132). San Diego: Singular.
- Ladd, D. R., & Campbell, W. N. (1991). Theories of prosodic structure: Evidence from syllable duration. *Proceedings of the 12th International Congress of Phonetic Sciences* (pp. 290–293). Aix-en-Provence, France.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018–2024.
- Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon* (Research on Speech Perception, Tech. Rep. No. 6). Bloomington, IN: Indiana University, Department of Psychology, Speech Research Laboratory.
- Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62, 615–625.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing*, 19, 1–36.



- Magnuson, J. S. (2001). *The microstructure of spoken word recognition*. Unpublished doctoral thesis, University of Rochester, New York.
- Magnuson, J. S., & Strauss, T. (2006). *Word-length biases in spoken word recognition: Goodness-of-fit and competition dynamics*. Manuscript submitted for publication.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The microstructure of spoken word recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, *132*, 202–227.
- Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In G. T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: The second sperlonga meeting* (pp. 187–210). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes and features. *Psychological Review*, *101*, 653–675.
- Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions during word-recognition in continuous speech. *Cognitive Psychology*, *10*, 29–63.
- Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576–585.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.
- Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 873–889.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.
- Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, *23*, 299–325.
- Pollack, I., & Pickett, J. M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, *3*, 79–84.
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.
- Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (in press). Effects of prosodically-modulated sub-phonetic variation on lexical neighborhoods. *Cognition*.
- Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*, 1030–1033.
- Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America*, *35*, 200–206.
- Scheepers, C., Keller, F., & Lapata, M. (in press). Competition vs. enriched composition in metonymic verbs: Evidence from the visual world paradigm. *Cognitive Psychology*.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.
- Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing*, *18*, 89–99.
- Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken-language comprehension. *Science*, *268*, 1632–1634.
- Vitevitch, M. S. (2002). Influence of onset density on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 270–278.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science*, *9*, 325–329.
- Viviani, P. (1990). Eye movements in visual search: Cognitive, perceptual, and motor control aspects. In E. Kowler (Ed.), *Eye movements and their role in visual and cognitive processes: Vol. 4. Reviews of oculomotor research* (pp. 253–393). Amsterdam: Elsevier.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*, 1707–1717.

**Appendix**

Word	Frq	Log Frq	Fam	No. Nbs	Nb Dens	FW-NPR	No. Cohs	Coh Dens	FW-CPR
Low frequency, low neighborhood density, low cohort density									
Couch	13	2.56	7	10	23.86	0.1075	35	61.28	0.0419
Cube	5	1.61	7	5	9.06	0.1777	19	33.39	0.0482
Fox	11	2.40	7	11	21.07	0.1138	57	87.41	0.0274
Goose	7	1.95	7	16	28.28	0.0688	8	6.38	0.3050
Hook	5	1.61	6.8	20	41.35	0.0389	9	11.45	0.1406
Pump	15	2.71	7	20	30.64	0.0884	38	64.45	0.0420
Thumb	14	2.64	7	27	43.65	0.0605	10	12.99	0.2032
Torch	4	1.39	7	4	6.68	0.2074	27	35.68	0.0389
Vice	25	3.22	6.8	5	10.20	0.3154	24	45.42	0.0709
Yarn	20	3.00	7	6	14.09	0.2126	11	15.50	0.1933
Keg	3	1.10	7	9	17.00	0.0646	29	44.22	0.0248
Bolt	9	2.20	7	15	32.65	0.0673	65	79.50	0.0276
Shield	8	2.08	7	9	22.07	0.0942	14	26.69	0.0779
Chef	9	2.20	6.8	10	20.64	0.1064	21	31.48	0.0698
Thread	20	3.00	7	15	35.07	0.0854	31	68.31	0.0439
Throne	6	1.79	7	11	20.77	0.0863	31	68.31	0.0262
<i>Ms</i>	10.88	2.21	6.96	12.06	23.57	0.1184	26.81	43.28	0.0864
Low frequency, low neighborhood density, high cohort density									
Clown	6	1.79	7	14	22.97	0.0780	190	289.81	0.0062
Crutch	7	1.95	6.4	11	15.90	0.1224	232	313.68	0.0062
Drill	21	3.04	7	11	16.45	0.1850	109	199.37	0.0153
Flag	18	2.89	7	14	28.57	0.1012	154	204.15	0.0142
Fork	20	3.00	7	11	37.94	0.0790	90	181.28	0.0165
Frog	2	0.69	7	8	10.21	0.0679	161	281.51	0.0025
Skate	1.001	0.00	7	15	36.52	0.0000	177	255.75	0.0000
Skull	5	1.61	7	8	19.97	0.0806	177	255.75	0.0063
Spire	8	2.08	4.1	17	32.02	0.0649	179	298.85	0.0070
Stump	7	1.95	1	5	7.36	0.2644	331	623.00	0.0031
Trunk	13	2.56	7	5	13.87	0.1849	205	349.81	0.0073
Wreath	11	2.40	7	19	36.47	0.0658	84	152.29	0.0157
Plug	23	3.14	7	16	18.46	0.1699	117	196.21	0.0160
Crown	19	2.94	7	19	37.89	0.0777	232	313.68	0.0094
Grill	11	2.40	7	18	33.80	0.0710	226	356.00	0.0067
Groom	5	1.61	6.9	15	31.38	0.0513	226	356.00	0.0045
<i>Ms</i>	11.06	2.13	6.40	12.88	24.99	0.1040	180.63	289.20	0.0086
Low frequency, high neighborhood density, low cohort density									
Bow	13	2.56	6.7	41	112.97	0.0227	15	23.91	0.1073
Bull	16	2.77	7	34	71.94	0.0385	48	59.47	0.0466
Saw	8	2.08	7	43	108.12	0.0192	31	37.75	0.0551
Tee	5	1.61	7	57	176.09	0.0091	27	39.11	0.0412
Cake	16	2.77	7	50	105.68	0.0262	38	61.65	0.0450
Cane	13	2.56	6.5	67	129.03	0.0199	38	61.65	0.0416
Goat	8	2.08	7	24	66.26	0.0314	27	38.47	0.0540
Nail	20	3.00	7	42	81.71	0.0367	24	50.34	0.0595
Nun	6	1.79	7	32	83.98	0.0213	22	35.62	0.0503
Sheep	24	3.18	7	30	80.53	0.0395	14	26.69	0.1191

Sock	10	2.30	7	37	79.04	0.0291	52	72.24	0.0319
Vase	15	2.71	7	27	80.38	0.0337	13	24.43	0.1109
Vest	4	1.39	6.9	25	66.83	0.0207	37	54.21	0.0256
Chick	4	1.39	7	35	73.47	0.0189	28	39.23	0.0353
Knight	25	3.22	6.9	47	157.80	0.0204	37	51.16	0.0629
Net	24	3.18	6.9	38	120.08	0.0265	37	69.33	0.0458
Ms	13.19	2.41	6.93	39.31	99.62	0.0259	30.50	46.58	0.0583

Low Frequency, high neighborhood density, high cohort density

Match	24	3.18	7	27	60.61	0.0524	162	249.22	0.0128
Bear	24	3.18	7	67	178.10	0.0178	158	238.75	0.0133
Bell	23	3.14	7	63	137.35	0.0228	108	158.04	0.0198
Cap	22	3.09	7	51	106.71	0.0290	233	351.71	0.0088
Deer	13	2.56	7	53	136.86	0.0187	561	975.45	0.0026
Cone	15	2.71	7	56	115.61	0.0234	125	183.23	0.0148
Mop	2	0.69	7	34	70.00	0.0099	130	210.41	0.0033
Witch	13	2.56	7	38	97.68	0.0263	106	175.46	0.0146
Badge	6	1.79	6.9	26	67.17	0.0267	158	238.75	0.0075
Can	12	2.48	7	57	107.59	0.0231	233	351.71	0.0071
Grape	10	2.30	6.8	28	68.18	0.0338	226	356.00	0.0065
Pan	16	2.77	7	49	117.95	0.0235	136	189.82	0.0146
Patch	23	3.14	7	33	69.74	0.0450	136	189.82	0.0165
Pear	8	2.08	7	58	159.30	0.0131	136	189.82	0.0110
Cart	9	2.20	7	25	67.54	0.0325	325	513.15	0.0043
Calf	17	2.83	6.6	28	68.90	0.0411	233	351.71	0.0081
Ms	14.81	2.54	6.96	43.31	101.83	0.0274	197.88	307.69	0.0103

High frequency, low neighborhood density, low cohort density

Board	285	5.65	7	19	39.52	0.1430	65	79.50	0.0711
Child	620	6.43	7	7	15.14	0.4247	15	31.99	0.2010
Church	451	6.11	7	7	19.25	0.3175	5	10.83	0.5642
Dog	147	4.99	7	14	22.50	0.2218	32	32.17	0.1551
Fence	46	3.83	7	12	30.06	0.1274	41	54.45	0.0703
Food	198	5.29	7	15	35.97	0.1470	9	17.10	0.3093
Gift	45	3.81	7	9	21.59	0.1763	37	55.27	0.0689
Girl	374	5.92	7	24	29.25	0.2026	10	11.63	0.5093
Guard	63	4.14	7	14	39.24	0.1056	56	65.82	0.0629
Horse	203	5.31	6.8	6	18.94	0.2806	46	64.00	0.0830
Judge	81	4.39	7	6	12.41	0.3541	31	66.46	0.0661
Knife	86	4.45	6.8	12	36.55	0.1219	37	51.16	0.0871
Roof	64	4.16	7	24	49.85	0.0834	50	66.68	0.0624
Salt	52	3.95	7	19	35.81	0.1103	31	37.75	0.1047
Snake	70	4.25	7	12	22.60	0.1880	37	37.29	0.1139
Switch	63	4.14	7	15	38.47	0.1077	67	98.09	0.0422
Ms	178.00	4.80	6.98	13.44	29.20	0.1945	35.56	48.76	0.1607

High frequency, low neighborhood density, high cohort density

Dress	63	4.14	6.8	11	23.38	0.1772	109	199.37	0.0208
Truck	80	4.38	7	12	21.41	0.2047	205	349.81	0.0125
Cloud	64	4.16	7	14	30.01	0.1386	190	289.81	0.0144
Club	178	5.18	6.8	6	12.70	0.4081	190	289.81	0.0179
Desk	69	4.23	6.9	6	14.20	0.2983	117	190.58	0.0222
Scale	62	4.13	7	12	28.33	0.1457	177	255.75	0.0161
Screen	53	3.97	7	7	17.08	0.2325	177	255.75	0.0155

Card	61	4.11	7	17	46.44	0.0885	325	513.15	0.0080
Film	127	4.84	7	8	20.44	0.2370	146	229.77	0.0211
School	687	6.53	7	13	33.73	0.1937	177	255.75	0.0255
Bridge	117	4.76	6.9	8	20.22	0.2355	245	334.12	0.0143
Crowd	63	4.14	7	11	31.86	0.1301	232	313.68	0.0132
Frame	96	4.56	6.9	14	41.52	0.1099	161	281.51	0.0162
Class	292	5.68	6.9	19	29.37	0.1933	190	289.81	0.0196
Branch	63	4.14	6.8	11	18.80	0.2204	245	334.12	0.0124
Plant	182	5.20	7	8	28.29	0.1840	117	196.21	0.0265
Ms	141.06	4.64	6.94	11.06	26.11	0.1998	187.69	286.19	0.0173
High frequency, high neighborhood density, low cohort density									
Ball	123	4.81	7	46	104.44	0.0461	25	37.47	0.1284
Chair	89	4.49	7	39	111.55	0.0402	43	77.90	0.0576
Pool	129	4.86	7	35	80.14	0.0606	4	7.35	0.6616
Key	71	4.26	7	53	136.65	0.0312	24	35.50	0.1201
Shoe	58	4.06	6.9	50	164.84	0.0246	11	16.03	0.2533
Boat	123	4.81	7	47	123.49	0.0390	65	79.50	0.0605
Bone	53	3.97	7	43	101.30	0.0392	65	79.50	0.0499
Gun	142	4.96	7	36	86.01	0.0576	46	53.98	0.0918
Top	136	4.91	7	36	79.40	0.0619	46	67.42	0.0729
Chain	60	4.09	7	38	95.24	0.0430	14	30.05	0.1363
Wall	224	5.41	7	40	115.38	0.0469	58	87.19	0.0621
Moon	63	4.14	7	30	81.21	0.0510	14	34.25	0.1210
Goal	100	4.61	6.9	46	92.33	0.0499	27	38.47	0.1197
Knee	73	4.29	7	61	178.16	0.0241	34	56.55	0.0759
Sheet	71	4.26	7	29	88.20	0.0483	14	26.69	0.1597
Suit	64	4.16	7	41	100.78	0.0413	54	81.23	0.0512
Ms	98.69	4.51	6.99	41.88	108.70	0.0441	34.00	50.57	0.1389
High frequency, high neighborhood density, high cohort density									
Coat	52	3.95	7	46	119.02	0.0332	125	183.23	0.0216
Heart	199	5.29	7	27	62.06	0.0853	115	163.14	0.0324
Plane	138	4.93	7	29	82.56	0.0597	117	196.21	0.0251
Tree	160	5.08	7	28	68.74	0.0738	205	349.81	0.0145
Band	64	4.16	6.9	29	80.94	0.0514	158	238.75	0.0174
Bed	139	4.93	1	47	104.86	0.0471	108	158.04	0.0312
Car	393	5.97	7	43	96.91	0.0616	325	513.15	0.0116
Hat	71	4.26	7	53	159.17	0.0268	157	221.17	0.0193
Lip	87	4.47	7	49	83.92	0.0532	126	205.08	0.0218
Man	2110	7.65	7	53	118.28	0.0647	130	210.41	0.0364
Star	58	4.06	7	25	66.81	0.0608	331	623.00	0.0065
Train	86	4.45	7	25	65.79	0.0677	205	349.81	0.0127
Bag	51	3.93	7	47	97.55	0.0403	158	238.75	0.0165
Brain	64	4.16	7	36	73.24	0.0568	245	334.12	0.0124
Cup	58	4.06	7	26	72.81	0.0558	84	161.15	0.0252
Hair	160	5.08	7	58	177.91	0.0285	157	221.17	0.0229
Ms	243.13	4.78	6.62	38.81	95.66	0.0542	171.63	272.94	0.0205

*Note.* Frq = frequency; Fam = familiarity (as measured via 7-point ratings by Nusbaum, Pisoni, & Davis, 1984); Nb = neighbor; Dens = density; Coh = cohort; FW-NPR = frequency weighted neighborhood probability rule (Luce, 1986); FW-CPR = frequency weighted cohort probability rule. Each probability is the log frequency of the item divided by its neighborhood density (i.e., the summed log frequencies of its neighbors or cohorts). Two items were not included in the Nusbaum et al. familiarity ratings (*stump* and *bed*). Familiarities of 1 are presented in these tables for those items, although both are intuitively highly familiar.