Language and Speech

Word Length and Lexical Competition: Longer is the Same as Shorter

Michael S. Vitevitch, Melissa K. Stamer and Joan A. Sereno

The online version of this article can be found at:

Published by:

$SAGE

Additional services and information for *Language and Speech* can be found at:

Email Alerts: http://las.sagepub.com/cgi/alerts

Subscriptions: http://las.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.co.uk/journalsPermissions.nav

Citations http://las.sagepub.com/cgi/content/refs/51/4/361

Word Length and Lexical Competition: Longer is the Same as Shorter

Michael S. Vitevitch[1], Melissa K. Stamer[2], Joan A. Sereno[2]

[1] *Department of Psychology, University of Kansas*
[2] *Department of Linguistics, University of Kansas*

Key words

bisyllabic words

neighborhood activation model

neighborhood density

word recognition

Abstract

Neighborhood density refers to the number of words that sound similar to a given word. Previous studies have found that neighborhood density influences the recognition of spoken words (Luce & Pisoni, 1998); however, this work has focused almost exclusively on monosyllabic words in English. To investigate the effects of neighborhood density on longer words, bisyllabic words varying in neighborhood density were presented auditorily to participants in a perceptual identification task and a lexical decision task. In the perceptual identification task, words with sparse neighborhoods were more accurately identified than words with dense neighborhoods. In the lexical decision task, words with sparse neighborhoods were responded to more quickly and more accurately than words with dense neighborhoods. These results are similar to those found in studies examining the influence of neighborhood density on the recognition of monosyllabic words in English. In order to better understand lexical processing, models of spoken word recognition must account for the processing of words of all types.

— Language and Speech

1 Introduction

All current models of spoken word recognition (e.g., Luce & Pisoni, 1998; Marslen-Wilson, & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994) account for the activation of and competition among multiple, similar word-forms. A common method used to assess the number of competitors that a word has is to measure the *neighborhood density* of the word. Neighborhood density is simply a count of the number of words that are formed by the addition, substitution, or deletion of one phoneme in a target word (Greenberg & Jenkins, 1967; Landauer & Streeter, 1973; Luce & Pisoni, 1998). For example, the word *kit* has neighbors such as *skit* where a phoneme is added, *cot*, *lit*, or *kid* where a phoneme is substituted, and *_it* where a phoneme is deleted. A word, like *cat*, with many similar sounding words (e.g., *at*, *bat*, *mat*, *rat*, *scat*, *pat*, *sat*, *vat*, *fat*, *gnat*, *cab*, *cad*, *calf*, *cash*, *cap*, *can*, *cot*, *kit*, *cut*, *coat*) is said to have a dense neighborhood, whereas a word, like *pig*, with few similar sounding words (e.g., *fig*, *wig*, *big*, *peg*, *pin*, *pitch*) is said to have a sparse neighborhood. Note that each word has additional neighbors, but only a few are listed for illustrative purposes.

Laboratory-based experiments examining the process of spoken word recognition have shown in a variety of tasks that English words with a sparse neighborhood are responded to more quickly and more accurately than words with a dense neighborhood (Luce & Pisoni, 1998; see also Vitevitch & Luce, 1998, 1999). These results suggest that during spoken word recognition multiple word-forms are activated and compete with each other. (See Storkel, 2001, for the influence of neighborhood density on the acquisition of English words; Vitevitch, 1997, and Vitevitch & Sommers, 2003, for the influence of neighborhood density on the production of English words; and Vitevitch & Stamer, 2006, for the influence of neighborhood density on the production of Spanish words).

Evidence for the activation of and competition among multiple word-forms during spoken word recognition also comes from an analysis of a corpus of speech perception errors known as "slips of the ear" (Vitevitch, 2002b). In a slip of the ear, the "listener reports hearing, as clearly and distinctly as any correctly perceived stretch of speech, something that does not correspond to the speaker's actual utterance" (Bond, 1999, p. 1). Vitevitch (2002b) compared the words that were reported in slips of the ear to comparable words (i.e., content words with similar word length and familiarity rating as the slips) that were randomly drawn from the lexicon, and found that the words involved in slips of the ear had denser neighborhoods than words randomly drawn from the lexicon. This more ecologically valid examination of the process of spoken word recognition further demonstrates that multiple, similar word-forms are activated and compete during spoken word recognition, and that words that activate many similar sounding words are more likely to be misperceived than words that activate few similar sounding words.

It is important to note, however, that much of the empirical work—including studies of human behavior and computer simulations—examining the process of lexical activation has used monosyllabic words as stimuli. This is not only true for studies of spoken word recognition (e.g., Luce & Pisoni, 1998; Marslen-Wilson & Warren, 1994), but it is also true for studies of visual word recognition (e.g., Coltheart,

Rastle, Perry, Langdon, & Ziegler, 2001). In order to better understand the process of spoken (and visual) word recognition, it is important to understand how many different types of words, such as words with more than one syllable, are processed.

Most work examining the recognition of longer words, however, has focused on longer words that are also morphologically complex (e.g., Frost, Grainger, & Rastle, 2005; Greber & Frauenfelder, 1999; Hay, 2003; Schriefers, Zwitserlood, & Roelofs, 1991; Zwitserlood, 2004; however, see Mattys, Bernstein, & Auer, 2002). It is unclear what the implications of theories of morphological decomposition are for the processing of words that are monomorphemic. Similarly, it is unclear how current models of spoken word recognition—based primarily on work examining monosyllabic words—account for the processing of longer words. Given that Wiener and Miller (1946) found that longer words were identified in noise more accurately than shorter words, it is possible that longer words may influence the processing dynamics in a model of spoken word recognition differently than the way shorter words influence the processing dynamics of the model.

Indeed, recent work by Vitevitch and Rodríguez (2005) shows very different processing dynamics for the recognition of spoken bisyllabic words. The results of an auditory lexical decision task showed that bisyllabic words with dense neighborhoods were recognized more quickly and more accurately than bisyllabic words with sparse neighborhoods. It is important to note, however, that this work examined the recognition of bisyllabic *Spanish* words by native speakers of Spanish, not the recognition of bisyllabic English words by native speakers of English. Although the results of Vitevitch and Rodríguez suggest that multiple word-forms can facilitate the recognition of spoken words, it is unclear if the facilitative rather than competitive influence of neighborhood density on spoken word recognition could be better accounted for by a cross-linguistic difference in spoken word recognition processes—like that seen in the process of segmenting a spoken word from fluent speech (cf., Cutler & Norris, 1988; Mehler, Dommergues, Frauenfelder, & Segui, 1981)—or if the result reported in Vitevitch and Rodríguez was due to a different processing dynamic being employed for the recognition of longer versus shorter spoken words, as the work of Pitt and Samuel (2006) might suggest.

Pitt and Samuel (2006) conducted four phoneme identification experiments using monosyllabic and trisyllabic words to examine the influence of word length on lexical activation and on lexical competition/inhibition. In these experiments, participants heard one of two words (e.g., *wish* or *miss*, *abolish* or *arthritis*) and had to decide whether the last phoneme was either "s" (/s/) or "sh" (/ʃ/). The last phoneme in each word had been modified to create a continuum from more /s/-like to more /ʃ/-like tokens. Pitt and Samuel reasoned that the extent of the bias in labeling the ambiguous phonemes in the /s/–/ʃ/ continuum, or the *lexical shift*, was indicative of the amount of lexical influences on phoneme perception, such that a large lexical shift indicated a large influence of lexical activation on phoneme identification.

They hypothesized that longer words would have greater lexical activation than shorter words, leading to a larger lexical shift in the trisyllabic words than in the mono-syllabic words. They further hypothesized that words with more competitors would be subject to more competition (or to greater inhibition), resulting in decreased lexical

activation, and to a smaller lexical shift in the trisyllabic words with many competitors (as defined by a late uniqueness point in the word) than in the trisyllabic words with few competitors (as defined by an early uniqueness point in the word). Evidence from the four experiments appeared to support these hypotheses. For example, Pitt and Samuel (2006) showed that longer words and words with an early uniqueness point displayed the greatest lexical shifts. They also found that duration manipulations and task demands modulated the effect, most notably for the monosyllabic words. Pitt and Samuel concluded that a better understanding of lexical processing could only be obtained by examining words of various lengths, not just monosyllabic words as previous studies have used.

Although we agree with Pitt and Samuel that a better understanding of lexical processing can only be obtained by examining words of various lengths, there are, unfortunately, several issues that make the results obtained by Pitt and Samuel (2006) difficult to interpret in such a straightforward manner, leaving the influence of word length on lexical processing unclear. First, consider that Pitt and Samuel (2006) defined lexical competition in terms of uniqueness points. Words with early uniqueness points were hypothesized to have few competitors, whereas words with late uniqueness points were hypothesized to have many competitors. This definition conflates uniqueness point with number of competitors (i.e., "cohort size," Pitt & Samuel, 2006, p.1122). Unfortunately, the relationship between uniqueness point and the number of competitors is not so straightforward. A word with a late uniqueness point may have many competitors as Pitt and Samuel suggest, or a word with a late uniqueness point may simply overlap one word, diverging only at the last phoneme, such as *absence* (/æbsəns/) and *absent* (/æbsənt/). That is, a word with a late uniqueness point may not have many competitors, but may actually have just one competitor.

The lack of a straightforward relationship between uniqueness point and the number of competitors is illustrated more clearly in Figure 1, where the computationally derived uniqueness point of each of the (approximately 20,000) words in the database used in Vitevitch and Luce (2004) are plotted as a function of the number of competitors of each word, as defined by the one-phoneme metric (Greenberg & Jenkins, 1967; Landauer & Streeter, 1973; Luce & Pisoni, 1998). As shown in the figure, many possibilities exist: a word with an early uniqueness point may have few competitors, a word with an early uniqueness point may have many competitors, a word with a late uniqueness point may have few competitors, and a word with a late uniqueness point may have many competitors. Because uniqueness point and the number of competitors are not linearly related, it is difficult to use uniqueness point as a consistent measure of number of competitors.[1] Although words with a later uniqueness point may be subject to greater inhibition, the study by Pitt and Samuel (2006) does not show that this influence is due to a greater number of competitors.

---

[1]   Also note that if a regression line was plotted for the data in Figure 1, the variance around the regression line would not be the same for all values of the predictor variable (those values on the x-axis), indicating that the assumption of homoscedasticity has been violated and making the results of such an analysis difficult to interpret.

Figure 1

Uniqueness point plotted as a function of the number of neighbors of each word in the database used in Vitevitch and Luce (2004)
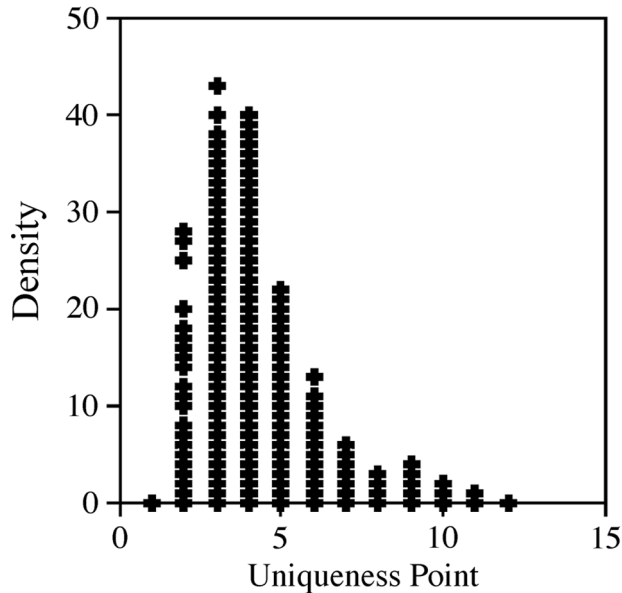
Second, Pitt and Samuel (2006) compared the size of the lexical shift obtained for the monosyllabic words to the lexical shift obtained for the trisyllabic words. In Experiment 1, they found that the monosyllabic words had a lexical shift of 26%, whereas the trisyllables (collapsed over uniqueness point) had a lexical shift of 40%; the difference between the lexical shifts (14%) was reported as statistically significant, $F(1, 27) = 15.70$. They interpreted the difference in lexical shift across word length as evidence that longer words generated more lexical activation than shorter words because longer words provided more perceptual information than shorter words. This interpretation is questionable because the number of competitors for the mono- and trisyllabic words (i.e., neighborhood density) was not controlled. In the stimuli used in Pitt and Samuel, we found a statistically significant difference, $t(14) = 5.56$, $p < .0001$, in the number of phonological neighbors between the monosyllabic words (mean = 13.8 neighbors, standard deviation = 7.1) and trisyllabic words (mean = 0.0 neighbors, standard deviation = 0.0), as defined by the one-phoneme metric (Greenberg & Jenkins, 1967; Landauer & Streeter, 1973; Luce & Pisoni, 1998). Recall that Luce and Pisoni (1998) suggested that words with dense neighborhoods experience more competition than words with sparse neighborhoods, accounting for the slower and less accurate responses to words with dense neighborhoods compared to words with sparse neighborhoods. Therefore, in the case of Pitt and Samuel (2006), it is unclear if the larger lexical shift that they observed for the trisyllabic words was due to greater activation among the longer words as they hypothesized, or if the observed difference in the lexical shift was due to greater competition among the monosyllabic words,

which were from significantly denser neighborhoods than the trisyllabic words that they used.

In addition to the difference in neighborhood density between the tri- and monosyllabic words, some of the monosyllabic word pairs used in Pitt and Samuel (2006) also varied in neighborhood density (e.g., *wish* has 13 neighbors, whereas *miss* has 23 neighbors). Newman, Sawusch, and Luce (1997) found that phoneme perception was influenced by the neighborhood density of the stimuli at the ends of the continuum in the Ganong (1980) paradigm. The extent to which the results of Pitt and Samuel (2006) might also have been influenced by neighborhood density effects of this type is unclear.

Given the difficulty in interpreting the results of Pitt and Samuel (2006), the influence of word length on lexical processing is still unclear. To more directly test whether longer words are processed differently from shorter words, we examined how neighborhood density influences the recognition of spoken bisyllabic words in English. Luce and Pisoni (1998) proposed the neighborhood activation model to quantitatively account for the influences of stimulus word intelligibility, stimulus word frequency, neighborhood confusability, and neighborhood frequency on spoken word recognition. These influences on spoken word recognition were expressed in the Frequency Weighted Neighborhood Probability Rule (FWNPR) (see equation 6 in Luce & Pisoni, 1998):

$$p(ID) = \frac{\prod\limits_{i=1}^{n} p(PS_i|PS_i) * Freq_s}{\left\{\left[\prod\limits_{i=1}^{n} p(PS_i|PS_i)\right] * Freq_s\right\} + \sum\limits_{j=1}^{nn}\left\{\left[\prod\limits_{i=1}^{n} p(PN_{ij}|PS_i)\right] * Freq_{Nj}\right\}} \tag{1}$$

where *p(ID)* is the probability of correctly identifying the stimulus word, PS_i; is the probability of the *i*th phoneme of the stimulus word, PN_{ij} is the probability of the *i*th phoneme of the *j*th neighbor, *n* is the number of phonemes in the stimulus word and the neighbor, Freq_S is the frequency of the stimulus word, Freq_{Nj} is the frequency of the *j*th neighbor, and *nn* is the number of neighbors.

Note that in the FWNPR, *n*, the number of phonemes in the stimulus word and the neighbor, is not restricted in any way (e.g., to "short" words, or *n* can only be less than a certain value, etc.). Thus, the neighborhood activation model would predict that factors such as stimulus word intelligibility, stimulus word frequency, neighborhood confusability, and neighborhood frequency should influence the recognition of a long word the *same way* that those factors influence the recognition of a short word. In other words, there is not one set of word recognition processes used for words of one length, and a different set of processes used for words of another length.

Although most of the empirical tests of the neighborhood activation model have used monosyllabic words as stimuli, a few studies have examined the influence of neighborhood density on the processing of bisyllabic words in English (e.g., Charles-Luce, Luce, & Cluff, 1991; Cluff & Luce, 1990; Luce & Cluff, 1998; Vitevitch & Luce, 1999). However, certain characteristics of the stimuli and analyses employed in those

studies limit the ability to generalize the findings that were obtained. First, consider the stimuli used in Cluff and Luce (1990), Luce and Cluff (1998), and Vitevitch and Luce (1999). The stimuli in those experiments were so-called *spondaic* words, or words that contained two syllables (each a free morpheme) with equal stress, such as *baseball*. Most bisyllabic words in English do not have this structure. Rather, most bisyllabic English words are monomorphemic with a strong–weak stress pattern (Cutler & Norris, 1988; Sereno & Jongman, 1990), making it difficult to generalize the results of these previous studies of spondaic words to more typical bisyllabic words in English.

In addition to the unique stress pattern of the words used as stimuli in these previous experiments, the neighborhood density of the component syllables of the bisyllabic words—not the neighborhood density of the entire word—was manipulated in these previous studies. For example, previous studies examined how the neighborhood density of *base* and the neighborhood density of *ball* influenced processing, but did not examine how the neighborhood density of the entire word, *baseball*, influenced processing. Finally, Cluff and Luce (1990) and Luce and Cluff (1998) used words comprised of syllables that were *lexically easy* or *lexically hard* words. A lexically easy word is one with high word frequency and a sparse neighborhood, making it "easy" to recognize, whereas a lexically hard word is one with low word frequency and a dense neighborhood, making it "hard" to recognize. Thus, it is unclear how the overall neighborhood density of the word influences the recognition of a bisyllabic word when word frequency is not conflated with neighborhood density as it is in the easy–hard distinction.

In order to better understand how bisyllabic words are recognized in English, the present set of experiments used the same tasks—perceptual identification and auditory lexical decision—and the same definition of competitor set size—neighborhood density—that were used by Luce and Pisoni (1998) to examine the influence of lexical competition on the recognition of monosyllabic words in English. The stimuli of the present study, however, consisted of bisyllabic words that were not spondaic and that varied in neighborhood density across the entire word (word frequency and neighborhood frequency were controlled).

2 Experiment 1

In a perceptual identification task, a word mixed with white noise is presented auditorily over a set of headphones and the participant is asked to identify the word that was presented. Luce and Pisoni (1998) found, over a variety of signal-to-noise ratios (+15 dB, +5 dB, and –5 dB), that monosyllabic English words with sparse neighborhoods were identified more accurately than monosyllabic words with dense neighborhoods. A straightforward prediction can be derived from the neighborhood activation model developed by Luce and Pisoni: as in the case of monosyllabic words, bisyllabic words with sparse neighborhoods will be identified more accurately than bisyllabic words with dense neighborhoods (*ceteris paribus*). Note that the bisyllabic words used in the present experiment are fairly typical bisyllabic English words (i.e., words with a strong–weak stress pattern), and that only one signal-to-noise ratio was examined.

Language and Speech

2.1
Method

2.1.1
Participants

Thirty-seven native speakers of English (U.S. Midwestern dialect) participated in this experiment. The participants received credit toward an introduction to psychology class for their participation. All participants reported no history of hearing or speech disorders.

2.1.2
Materials

Fifty-six English words were used in this task. All the words were bisyllabic with a strong–weak stress pattern.

2.1.2.1 Neighborhood density   Phonological neighbors were defined as words that differed from the target word by the addition, substitution, or deletion of one phoneme (Greenberg & Jenkins, 1967; Landauer & Streeter, 1973; Luce & Pisoni, 1998). Half of the stimuli were classified as having dense neighborhoods with a mean of 11.71 phonologically similar words (SD = 1.58), and the other half of the stimuli were classified as having sparse neighborhoods with a mean of 4.43 phonologically related words (SD = 1.99). The difference between the two conditions was significant, $F(1, 54) = 229.88$, $p < .0001$.

In general, there are fewer neighbors for the bisyllabic words used in the present set of experiments than the monosyllabic words used in previous studies. This difference between our stimuli and the monosyllabic stimuli used in previous studies in terms of number of neighbors may be a result of the language-wide relationship between neighborhood density and word length: longer words tend to have fewer neighbors than shorter words (Amano, 1996; Frauenfelder, Baayen, & Hellwig, 1993).

2.1.2.2 Frequency of occurrence of the stimuli and their neighbors   Word frequency refers to the average number of times a word occurs in the language. The stimuli were controlled in terms of word frequency (dense words: mean = 46.89 occurrences per million, SD = 71.68; sparse words: mean = 52.89 occurrences per million, SD = 91.10). Neighborhood frequency refers to the mean word frequency of the neighbors of the target word. The stimuli were also controlled in terms of neighborhood frequency (dense words: mean = 92.22 occurrences per million, SD = 195.80; sparse words: mean = 123.86 occurrences per million, SD = 174.63). None of the differences in word frequency or neighborhood frequency were significant (all F's$(1, 54) < .5$, $p > .5$). All the frequency counts were based on the information contained in Kucera and Francis (1967).

2.1.2.3 Component segments   All stimuli contained three to five phonemes, with the mean phoneme length in the dense condition equal to 3.9 phonemes (SD = .5) and the mean phoneme length in the sparse condition equal to 4.0 phonemes (SD = .5). This difference was not statistically significant, $F(1, 54) = .68$, $p > .41$. Based on the computer readable transcriptions (see Vitevitch & Luce, 2004,

for more information regarding the transcription conventions in the database used to select the stimuli), stimulus items had the following syllable structures (C = consonant, V = vowel): CVV (e.g., *hurry*), CVCV (e.g., *movie*), CVCCV (e.g., *lumber*), CVCC (e.g., *muscle*), or CVCCC (e.g., *candle*). In the last two cases, the word-final C was a syllabic consonant.

Further note that equal numbers of the following phonemes appeared in each condition in the onset position: /b, d, k, f, h, l, m, p, ʃ, t, v, w/. For the vowels that appeared in the second position of each word, the dense words had the following vowels (with the number of occurrence in parentheses), æ (11), ɑ (1), ɛ (5), ʌ (4), ɝ (3), aI (1), e (1), i (1), o (1), and the sparse words had the following vowels (with the number of occurrence in parentheses), æ (5), ɑ (1), ɛ (2), ʌ (4), ɝ (3), aI (1), e (2), i (1), o (2), oI (1), I (1), u (1), au (3), ɔ (1). A chi-square analysis shows that there was no statistically significant difference in the distribution of vowels in the second position of each word between the two conditions ($\chi^2 = 11.20$, d.f. = 13, $p = .59$).

Finally, we counted the number of times a fricative appeared in any of the phoneme positions that followed the initial CV. For the dense words, three fricatives were found in one of the phoneme positions that followed the initial CV, and for the sparse words, seven fricatives were found in one of the phoneme positions that followed the initial CV. A chi-square analysis shows that this difference was not statistically significant ($\chi^2 = 1.60$, d.f. = 1, $p = .20$). It is important to carefully balance the phonological segments that appear in each condition because white noise differentially masks different types of phonemes. Given the similarities in the distribution of constituent phonemes in the two conditions, it is more likely that any difference we observe in the perceptual identification task is due to the difference in the independent variable (i.e., neighborhood density) than to differences in the constituent phonemes in the two conditions. The words, their phonemic transcriptions, and their lexical characteristics are listed in Appendix A.

2.1.2.4 Uniqueness points   The uniqueness points (Grosjean, 1980; Marslen-Wilson & Tyler, 1980) were determined computationally as in Luce (1986). Using the same computerized lexicon as Luce (1986), the uniqueness point of a word corresponded to the point at which the phonological transcription of the target word differed from all other transcriptions in the computerized lexicon (Nusbaum, Pisoni, & Davis, 1984). The mean uniqueness point was 4.29 phonemes for the dense words and 4.25 phonemes for the sparse words. This difference was not statistically significant, $F(1, 54) = .028$, $p > .80$. By controlling uniqueness point, but varying neighborhood density, we were able to more directly assess the influence of lexical competition on longer words (cf., Pitt & Samuel, 2006).

2.1.2.5 Recording the stimuli   All of the words were recorded from a list in isolation by a female native speaker of English from the Midwest (the second author) at a normal speaking rate in an IAC sound attenuated booth using a high-quality microphone. The stimuli were recorded onto a digital audiotape at a sampling rate of 44.1 kHz. The recordings were transferred directly to a computer hard drive via a sound card using ProTools LE software (Digidesign). Correct pronunciation of each word was verified, and the words were edited into individual sound files using

Language and Speech

Sound Edit 16 (Macromedia, Inc.). The amplitude of the sound files was adjusted with the Normalize function to amplify the words to their maximum peak value without clipping or distorting the sound and without changing the pitch of the words.

2.1.2.6 Durations   The mean total file duration, measured from the beginning of the sound file to the end of the sound file, for the dense words was 500 ms (SD = 48) and for the sparse words was 511 ms (SD = 52). The mean onset duration, measured from the beginning of the file to the onset of the stimulus word, for the dense words was 29 ms (SD = 10) and for the sparse words was 31 ms (SD = 9). The mean stimulus duration, measured from the onset of the word to the end of the word in the sound file excluding any silence before and after the word in the sound file, for the dense words was 432 ms (SD = 49) and for the sparse words was 446 ms (SD = 44). None of the differences in duration were statistically significant (all F's(1, 54) < .33, p > .55).

2.1.2.7 Noise   The stimuli in the perceptual identification task were presented at a +12 dB signal-to-noise ratio (S/N). Sound Edit 16 was used to create the degraded stimuli by adding to each sound file white noise that was equal in duration to each sound file and that was 12 dB less in amplitude than the mean amplitude of the sound file.

2.1.3
Procedure
Participants were seated in front of an iMac running PsyScope 1.2.2 (Cohen, MacWhinney, Flatt, & Provost, 1993) with a set of Beyerdynamic DT 100 headphones, and a computer keyboard. The computer program, PsyScope, controlled stimulus presentation and response collection.

Each trial proceeded as follows: the word READY appeared for 500 ms on the screen followed immediately by a randomly selected stimulus item (i.e., a word mixed with noise) presented auditorily through the headphones. Each stimulus was presented only once. The participant was instructed to use the computer keyboard to type the word that was heard through the headphones. After the participant pressed the Return key, the next trial began. Participants were able to see their responses while typing and were able to correct any typing error before pressing the Return key. Participants were instructed to provide their best guess for each word they heard. They were also instructed to press the Return key or "?" if they were absolutely unable to identify the word. Each participant received five practice trials before proceeding to the rest of the experiment. The practice trials were used to familiarize the participants with the task and were not included in the analyses.

2.2
Results and discussion
Words were scored as correct if a phonological transcription of the response matched the phonological transcription of the input. Misspellings and the transpositions of letters were also scored as correct responses. The omission of a single letter in a word was scored as a correct response only if the response did not form another English

word. Typographical errors and the addition of a single letter in the responses were scored as a correct response if the character was within one key from the target letter on the keyboard. Words were scored as incorrect if they did not meet the above criteria.

In the dense condition, 77.1% (SD = 7.5) of the words were correctly identified, whereas in the sparse condition, 80.3% (SD = 9.2) of the words were correctly identified. A dependent samples t-test was used with participants as a random factor to examine the difference in accuracy rates between the dense and sparse words. This analysis showed that the difference between the dense condition and the sparse condition was statistically significant, $t(36) = 7.08$, $p = .01$. The observed difference is considered an effect of medium size ($d = .44$) and has a high probability of being replicated ($p_{rep} = .81$, the lower limit for "replicable" is $p_{rep} \approx .55$, Killeen, 2005).

It is the current practice of psycholinguistic research to provide additional analyses of the items used as stimuli under the assumption that such an analysis rules out the possibility that the effects that were observed were due to a few "odd" items in the stimulus set (however, see Raaijmakers, 2003). In current practice, this concern is typically addressed with an independent samples t-test treating the stimulus items as a random factor. There is, however, much debate about the proper use and interpretation of such analyses (Baayen, 2004; Baayen, Tweedie, & Schreuder, 2002; Cohen, 1976; Keppel, 1976; Raaijmakers, 2003; Raaijmakers, Schrijnemakers, & Gremmen, 1999; Smith, 1976; Wike & Church, 1976). Although there are a number of concerns with the proper use and interpretation of such analyses, we have provided these analyses in order to be consistent with the current conventions of the field.

When collapsed across participants, the items in the dense condition had a mean accuracy rate of 77.10% (SD = 23.1), whereas items in the sparse condition had a mean accuracy rate of 80.30% (SD = 20.8). Note that the mean values are identical to those obtained in the analysis treating participants as a random factor. However, the standard deviations in the present analysis are two to three times those obtained in the analysis treating participants as a random factor. Given the increased variability in the present analysis, and the less sensitive independent samples t-test used in the present analysis, it is not surprising that the analysis failed to show a statistically significant difference, $t(54) = .54$, $p = .59$. To further address the concern that a few "odd" items were contributing to the observed effect, we provide several additional analyses.

If a few "odd" items were responsible for the observed effect, one might expect to see the (numerically) higher accuracy for sparse words compared to dense words to disappear or reverse when the items are randomly split into two groups of sparse and dense words, as the "odd" item would only affect half of the data. When half of the dense words and half of the sparse words were randomly separated into two lists (List A and List B, for ease of discussion), the numerical advantage in accuracy rates for sparse over dense words was found in both lists. In List A, the items in the dense condition had a mean accuracy rate of 74.9% (SD = 23.8), whereas items in the sparse condition had a mean accuracy rate of 77.6% (SD = 23.7; $t(26) = .30$, $p = .76$). In List B, the items in the dense condition had a mean accuracy rate of 79.3% (SD = 23.1), whereas items in the sparse condition had a mean accuracy rate of 83.0% (SD = 17.9;

$t(26) = .47$, $p = .64$). Although neither of these analyses is statistically significant, both lists of randomly assigned stimuli (each containing half of the dense items and half of the sparse items) show the expected result: sparse words are responded to (numerically) more accurately than dense words. If the result observed in the participant analysis were due to a few "odd" items, it is unlikely that *both* lists of randomly separated items would continue to show the expected result.

To further rule out the possibility that a few "odd" items were responsible for the observed effect, we compared the frequency distributions of the accuracy rates for the dense and sparse items. If only a few sparse items were responded to accurately, producing a spurious advantage for sparse words, one would expect a positively skewed distribution for the sparse items. Similarly, if only a few dense items were responded to with low accuracy, producing a spurious disadvantage for dense words, then one would expect a negatively skewed distribution for the dense items. A chi-square analysis comparing the number of dense and sparse words that occurred in each bin of 10% increments from 0–100% accuracy shows that the two distributions do not differ from each other ($\chi^2 = 6.692$, d.f. $= 8$, $p = .57$; Preacher, 2001). The number of responses in each accuracy bin is provided in Appendix B. Given the similarity in the shape of the distributions for the dense and sparse items, it is unlikely that a few "odd" items in either the dense or sparse condition were responsible for the observed effect.

The result from the present perceptual identification task shows that bisyllabic English words with sparse phonological neighborhoods are identified more accurately than bisyllabic English words with dense phonological neighborhoods. These results are consistent with the predictions of the neighborhood activation model, and are similar to the results of Experiment 1 from Luce and Pisoni (1998). Across three S/N ratios (–5 dB, +5 dB, +15 dB), Luce and Pisoni found a significant, positive correlation between the predicted probability of identifying a stimulus word (these values were derived from the FWNPR, which takes into account the influences of stimulus word intelligibility, stimulus word frequency, neighborhood confusability, and neighborhood frequency) and identification scores obtained in the perceptual identification tasks, providing clear evidence that monosyllabic English words with sparse neighborhoods were identified more accurately than monosyllabic English words with dense neighborhoods. The results of the present experiment suggest that the mechanisms that influence the recognition of short words also influence the recognition of long words, contrary to the claims of Pitt and Samuel (2006), who found processing differences between short and long words.

Although a perceptual identification task has much ecological validity—we often experience background or street noise in addition to the signal produced by an interlocutor—we, like Luce and Pisoni (1998), wished to further examine the influence of neighborhood density on the recognition of spoken words in English. Like Luce and Pisoni (see their Experiment 2), we performed an auditory lexical decision task using the biyllabic words we used in the present experiment. As Luce and Pisoni (1998) pointed out, an auditory version of a lexical decision task enables us to examine the influence of neighborhood density without degrading the stimuli (as in the perceptual identification task). That is, a lexical decision task will enable us to "demonstrate that the effects of neighborhood structure generalize beyond words that are purposefully made difficult to perceive" (Luce & Pisoni, 1998, p. 15), thereby providing evidence

from another task that is convergent with the evidence obtained in the perceptual identification task.

3 Experiment 2

In an auditory lexical decision task, a word or a nonword is presented without noise over a set of headphones and the participant must decide as quickly and as accurately as possible if the stimulus is a real word in English or a made-up, nonsense word. Using this task, Luce and Pisoni (1998; Experiment 2) found that monosyllabic words with a sparse neighborhood were recognized more quickly than words with a dense neighborhood. To further explore how neighborhood density influences the recognition of longer words, the same bisyllabic words used in Experiment 1 were used in the present experiment.

3.1
Method

3.1.1
Participants

Forty right-handed native speakers of English (Midwestern dialect) participated in this experiment. The participants received credit toward an introduction to psychology class for their participation. All participants reported no history of hearing or speech disorders. None of the participants in the present experiment took part in Experiment 1.

3.1.2
Materials

The same stimuli that were used in Experiment 1 were used in the present experiment. Additionally, 56 bisyllabic nonwords were used in this task. Recall that the mean uniqueness point was 4.29 phonemes for the dense words and 4.25 phonemes for the sparse words, and that the mean length for the dense words was 3.9 phonemes, and for the sparse words the mean was 4.0 phonemes. This means that the uniqueness point of the words used in the present experiment fell, on average, after the final phoneme. To avoid giving the participants a cue regarding the lexical status of the stimulus, the nonwords that were used in the present experiment were formed by replacing the final phoneme of a real word with another phoneme to create a nonword. For example, the /s/ from "tennis" /tɛnɪs/ (not a stimulus item) was replaced with /k/ to form the nonword "tennik" /tɛnɪk/. Thus, participants had to listen to the entire stimulus before making a decision regarding the lexical status of the item.

   The nonwords were recorded in the same manner and at the same time as the real word stimuli used in Experiment 1, thereby eliminating possible cues to lexical status based on characteristics of the recordings. The phonological transcriptions of the nonwords are listed in Appendix C. None of the nonwords were created from stimulus items.

3.1.3
Procedure

The same equipment used in Experiment 1 was also used in the present experiment with the exception that a button box (New Micros), with a dedicated timing board to

provide millisecond accuracy, was interfaced with the computer to record responses. Each trial proceeded as follows: the word READY appeared for 500 ms on the screen, followed immediately by a randomly selected word or nonword presented auditorily through the headphones. The participants were instructed to indicate whether the item they heard was a real word in English or a nonword by pushing as quickly and as accurately as possible the appropriately labeled button. "WORD" responses were made with the right (dominant) hand, and "NONWORD" responses were made with the left hand. After the participants made a decision, the next trial began. Prior to the trials in the experiment, the participants received 10 practice trials. These trials were used to familiarize the participants with the task and were not included in the analyses.

3.2

Results and discussion

To maintain the current conventions of psycholinguistic research, separate *t*-tests were used to examine each dependent measure (reaction time and accuracy rates) with participants (t_1) and stimulus items (t_2) treated as random factors. Only accurate WORD responses were included in the analysis of reaction times. Reaction times were measured from the onset of the sound file to the onset of the button press response. Reaction times that were approximately two standard deviations above and below the overall mean reaction time (i.e., below 500 ms and above 1700 ms) were considered outliers and were excluded from the analysis; this accounted for less than 1% of the data.

In the analysis of reaction time, words with sparse neighborhoods (mean = 833 ms, SD = 74.53) were responded to more quickly than words with dense neighborhoods (mean = 846 ms, SD = 77.53; $t_1(39) = 2.10$, $p = .04$; $t_2(54) = 2.38$, $p = .02$). The observed difference in the analysis of the reaction times treating participants as a random factor is considered an effect of medium size ($d = .33$) and has a high probability of being replicated ($p_{rep} = .76$; Killeen, 2005).

A significant difference was also found with regards to accuracy rates ($t_1(39) = 2.41$, $p = .02$; $t_2(54) = 2.15$, $p = .03$), such that words with sparse neighborhoods were correctly responded to 94.3% of the time (SD = 4.6) and words with dense neighborhoods were correctly responded to 92.0% of the time (SD = 5.6). The observed difference in the analysis of the accuracy rates treating participants as a random factor is considered an effect of medium size ($d = .38$) and has a high probability of being replicated ($p_{rep} = .79$; Killeen, 2005). This pattern of results also suggests that participants did not employ a strategy that traded speed for accuracy in their responses.

Luce and Pisoni (1998; Experiment 2) found in an auditory lexical decision task that monosyllabic English words with sparse phonological neighborhoods were responded to more quickly than monosyllabic words with dense phonological neighborhoods. The results of the present experiment further suggest that neighborhood density also influences the processing of bisyllabic words in English in a similar manner. This finding, however, contrasts with the findings of Vitevitch and Rodríguez (2005), who found that bisyllabic Spanish words with dense neighborhoods were recognized more quickly and more accurately than bisyllabic Spanish words with

Language and Speech ———————————————————————

sparse neighborhoods in an auditory lexical decision task. The present result also contrasts with the conclusions of Pitt and Samuel (2006), who suggested that long words are subject to different processing dynamics than short words.

4 General discussion

The present set of experiments examined the influence of neighborhood density on the processing of English bisyllabic words. The results of the perceptual identification task in Experiment 1 showed that listeners identified bisyllabic English words with sparse neighborhoods more accurately than bisyllabic English words with dense neighborhoods. In the auditory lexical decision task used in Experiment 2, listeners responded more quickly and more accurately to bisyllabic English words with sparse neighborhoods than to bisyllabic English words with dense neighborhoods. As predicted by the neighborhood activation model (Luce & Pisoni, 1998), similar sounding bisyllabic English words, like similar sounding monosyllabic English words, compete during spoken word recognition. Words with few competitors are recognized more quickly and accurately than words with many competitors.

In contrast to the findings of the present study, Vitevitch and Rodríguez (2005) found that, in an auditory lexical decision task, bisyllabic Spanish words with dense neighborhoods were recognized more quickly and more accurately than bisyllabic Spanish words with sparse neighborhoods. Given the similar way in which monosyllabic and bisyllabic words (as demonstrated in the present study) in English were responded to, the divergent results of Vitevitch and Rodríguez (2005; see also Vitevitch & Stamer, 2006) may reflect some sort of cross-linguistic difference in spoken word recognition processes—like that seen in the process of segmenting a spoken word from fluent speech (cf., Cutler & Norris, 1988; Mehler et al., 1981)—rather than a difference in how longer and shorter words are processed.

Although there are a number of differences between English and Spanish that could lead to a difference in processing dynamics, one possibility relates to the *morphological*—in addition to the phonological—similarity that exists among some Spanish word-forms. Consider, for example, the Spanish word *niña* (i.e., the singular, feminine form for child, or little girl, and a word used in the study by Vitevitch & Rodríguez, 2005). Words that are phonologically related (based on the one-phoneme metric) *as well as* morphologically and semantically related—such as *niñas* (girls) and *niño* (boy)—would also be activated in the lexicon. In addition, other words such as *niños* (boys), *niñada* (childishness), *niñera* (nursemaid), *niñería* (trifle), and *niñez* (childhood) might also be partially activated in the lexicon, even though they differ from *niña* by more than one phoneme. Other examples of stimulus items from the study by Vitevitch and Rodríguez [and morphologically related items] include *moho* moss [*mohoso* mossy], *heno* hay [*henal* hayloft, *henar* hayfield], and *mudo* mute [*mudez* silence]. The morphological (and semantic) similarity among these partially activated word forms may lead to different processing dynamics in the spoken word recognition system than partially activated word-forms that are *only* phonologically similar (e.g., *cat* and *hat*). We recognize the speculative nature of this hypothesis (however, see Bybee, 2001, among others for similar ideas), but a complete explanation of the influence of neighborhood density on spoken word

recognition in Spanish (e.g., Vitevitch & Rodríguez, 2005) is beyond the scope of the present study and discussion.

The results of the present study suggest that, in English, competition among similar word forms affects the recognition of shorter and longer words and affects them in the same way. Specifically, long and short words with few competitors are recognized more quickly and accurately than long and short words with many competitors. However, the results of Pitt and Samuel (2006) suggest that different processing dynamics are used to recognize shorter and longer words. Pitt and Samuel hypothesized that because longer words have more phonemes than shorter words, they accumulate more bottom-up activation than shorter words. They further hypothesized that shorter words have more competitors than longer words, and that trisyllabic words with a late uniqueness point have more competitors than trisyllabic words with an early uniqueness point. Therefore, short words are subject to more inhibition than longer words, and trisyllabic words with a late uniqueness point are subject to more inhibition than trisyllabic words with an early uniqueness point.

In several phoneme identification experiments using monosyllabic and trisyllabic words, Pitt and Samuel found a larger lexical shift—a measure of lexical influence on phoneme perception—in the labeling of ambiguous phonemes in an /s/–/ʃ/ continuum (Ganong, 1980) for longer words than for shorter words. This result suggested that longer words—which have more phonemes than shorter words, thereby contributing more bottom-up activation to the lexical representation—are activated to a greater degree than shorter words.

Although longer words do indeed have more phonemes than shorter words, it is unclear how *uniquely* informative the additional phonemes in the latter part of a longer word are. Connine, Titone, Deelman, and Blasko (1997) describe what they referred to as "lexical extant," or the information that occurs at the ends of words that can be used as a source of *confirmatory* evidence for the information at the beginning of the word. If the information in the later part of a long word is essentially redundant information (i.e., confirmatory evidence), it is not clear that this information will produce the same amount of lexical activation as the unique information found in the initial part of the word. Therefore, it's not clear that a longer word would indeed produce greater lexical activation than a shorter word, as Pitt and Samuel hypothesized, simply because it has more phonemes than a longer word.

In addition, the increased number of phonemes found in longer words might tax verbal short-term memory more than the smaller number of phonemes found in short words. As a result, *sublexical* representations might be recruited to maintain the longer input in phonological memory, so that word recognition can occur (see Experiments 4–6 of Vitevitch & Luce, 1999). Given that sublexical representations might also be activated by a long word, it is not clear that the effects in the phoneme identification task observed by Pitt and Samuel (2006) can be attributed solely to greater lexical activation for longer words than for short words.

Pitt and Samuel (2006) also observed a larger lexical shift for trisyllabic words with an early uniqueness point (which they took as an indication of few competitors) than for trisyllabic words with a late uniqueness point (which they took as an indication of many competitors), suggesting that trisyllabic words with a late

uniqueness point are subject to more inhibition than trisyllabic words with an early uniqueness point. We have already noted the difficulty in using uniqueness point as a measure of number of competitors (see Figure 1). It is also important to note that words that overlap in the initial portion of the word (i.e., the "cohort") are not the only items that contribute to lexical competition. Newman, Sawusch, and Luce (2005) found in lexical decision and phoneme identification tasks that words that differed from a target word only in their first phoneme still influenced lexical access. The results of Vitevitch (2002a) further suggest that word onsets may only take on a special status in the recognition of spoken (monosyllabic, English) words when overall neighborhood density is equated. Therefore, the claim that trisyllabic words with a late uniqueness point are subject to more inhibition than trisyllabic words with an early uniqueness point should be viewed with caution, as it is not clear that the trisyllabic stimuli used in Pitt and Samuel (2006) actually activated different numbers of lexical competitors.

The results of the present experiment, however, do clearly show that bisyllabic words with few competitors (i.e., a sparse neighborhood) are recognized more quickly and accurately than bisyllabic words with many competitors (i.e., a dense neighborhood). Although these results contrast with the claim by Pitt and Samuel (2006) that the processing dynamics of spoken word recognition differ for long and short words, these results are consistent with the predictions of the neighborhood activation model (Luce & Pisoni, 1998), which predicted that longer (i.e., bisyllabic) words should be influenced by the same properties—stimulus word intelligibility, stimulus word frequency, neighborhood confusability, and neighborhood frequency—as shorter (i.e., monosyllabic) words.

Like Pitt and Samuel (2006), however, we believe that a better understanding of the process of spoken word recognition can only be achieved by examining other types of words, not just monosyllabic words. The non-trivial task before spoken word recognition researchers is to examine how current models of spoken word recognition (e.g., Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994)—which have been primarily tested in studies employing monosyllabic and *monomorphemic* words as stimuli—relate to processing models that account for the recognition of morphologically complex words (e.g., Frost, Grainger, & Rastle, 2005; Greber & Frauenfelder, 1999; Hay, 2003; Schriefers, Zwitserlood, & Roelofs, 1991; Zwitserlood, 2004). Longer words in English tend to be of Latinate or Greek origins rather than Germanic origins (e.g., Reilly, Ramey, & Milsark, 2004), and, therefore, tend to be more morphologically complex than shorter words. In order to have a complete model of spoken word recognition, we must better understand how words of all length and all degrees of morphological complexity are processed.

References

AMANO, S. (1996). Some observations on neighborhood statistics of spoken English words. In D. B. Pisoni (Ed.), *Research on spoken language processing: Progress report no. 21 (1996–1997)* (pp.440–453). University of Indiana.

BAAYEN, R. H. (2004). Statistics in psycholinguistics: A critique of some current gold standards. *Mental Lexicon Working Papers in Psycholinguistics*, **1**, 1–45.

BAAYEN, R. H., TWEEDIE, F. J., & SCHREUDER, R. (2002). The subjects as a simple random effect fallacy: Subject variability and morphological family effects in the mental lexicon. *Brain and Language*, **81**, 55–65.

BOND, Z. S. (1999). *Slips of the ear: Errors in the perception of casual conversation*. New York: Academic Press.

BYBEE, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

CHARLES-LUCE, J., LUCE, P. A., & CLUFF, M. S. (1991). Retroactive influence of syllable neighborhoods. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp.173–184). Cambridge, MA: The MIT Press.

CLUFF, M. S., & LUCE, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, **16**(3), 551–563.

COHEN, J. (1976). Random means random. *Journal of Verbal Learning and Verbal Behavior*, **15**, 261–262.

COHEN, J., MacWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: An interactive graphic system for defining and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research, Methods, Instruments, and Computers*, **25**, 257–271

COLTHEART, M., RASTLE, K., PERRY, C., LANGDON, R., & ZIEGLER, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204–256.

CONNINE, C. M., TITONE, D., DEELMAN, T., & BLASKO, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, **37**, 463–480.

CUTLER, A., & NORRIS, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 113–121.

FRAUENFELDER, U. H., BAAYEN, R. H., & HELLWIG, F. M. (1993). Neighborhood density and frequency across languages and modalities. *Journal of Memory and Language*, **32**, 781–804.

FROST, R., GRAINGER, J., & RASTLE, K. (2005). Current issues in morphological processing: An introduction. *Language and Cognitive Processes*, **20**, 1–5.

GANONG, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 110–125.

GREBER, C., & FRAUENFELDER, U. H. (1999). On the locus of morphological effects in spoken-word recognition: Before or after lexical identification? *Brain and Language*, **68**, 46–53.

GREENBERG, J. H., & JENKINS, J. J. (1967). Studies in the psychological correlates of the sound system of American English. In L. A. Jacobovits & M. S. Miron (Eds.), *Readings in the Psychology of Language* (pp.186–200). New Jersey: Prentice Hall.

GROSJEAN, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, **28**, 267–283.

HAY, J. (2003). *Causes and consequences of word structure*. Outstanding dissertations in linguistics. New York: Routledge.

KEPPEL, G. (1976). Words as random variables. *Journal of Verbal Learning and Verbal Behavior*, **15**, 263–265.

KILLEEN, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, **16**, 345–353.

KUCERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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

LUCE, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, **39**, 155–158.

LUCE, P. A., & CLUFF, M. S. (1998). Delayed commitment in spoken word recognition: Evidence from cross-modal priming. *Perception & Psychophysics*, **60**, 484–490.

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

MARSLEN-WILSON, W. D., & TYLER, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, **8**, 1–71.

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. D., & WELSH, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, **10**, 29–63.

MATTYS, S. L., BERNSTEIN, L. E., & AUER, E. T. (2002). Stimulus-based lexical distinctiveness as a general word-recognition mechanism. *Perception & Psychophysics*, **64**, 667–679.

McCLELLAND, J. L., & ELMAN, J. L. (1986). The TRACE Model of Speech Perception. *Cognitive Psychology*, **18**, 1–86.

MEHLER, J., DOMMERGUES, J. Y., FRAUENFELDER, U., & SEGUI, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, **20**, 298–305.

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.

NEWMAN, R. S., SAWUSCH, J. R., & LUCE, P. A. (2005). Do postonset segments define a lexical neighborhood? *Memory and Cognition*, **33**, 941–960.

NORRIS, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, **52**, 189–234.

NUSBAUM, H. C., PISONI, D. B., & DAVIS, C. K. (1984). *Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words*. Research on Speech Perception, Progress Report No. 10. Bloomington, IN: Indiana University, Psychology Department, Speech Research Laboratory.

PITT, M. A., & SAMUEL, A. G. (2006). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance*, **32**, 1120–1135.

PREACHER, K. J. (2001, April). *Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence* [Computer software]. Retrieved 17 July 2007, from http://www.quantpsy.org

RAAIJMAKERS, J. G. W. (2003). A further look at the "language-as-fixed-effect-fallacy." *Canadian Journal of Experimental Psychology*, **57**, 141–151.

RAAIJMAKERS, J. G. W., SCHRIJNEMAKERS, J. M. C., & GREMMEN, F. (1999). How to deal with "the language-as-fixed-effect fallacy"; Common misconceptions and alternative solutions. *Journal of Memory and Language*, **41**, 416–426.

REILLY, J., RAMEY, C. H., & MILSARK, G. (2004). Confounds in the distinction between high- and low-imageability words: Phonological, etymological, and morphological differences. *Brain & Language*, **91**, 147–149.

SCHRIEFERS, H., ZWITSERLOOD, P., & ROELOFS, A. (1991). The identification of morphologically complex spoken words: Continuous processing or decomposition? *Journal of Memory and Language*, **30**, 26–47.

SERENO, J. A., & JONGMAN, A. (1990). Phonological and form class relations in the lexicon. *Journal of Psycholinguistic Research*, **19**, 387–404.

SMITH, J. E. K. (1976). The assuming-will-make-it-so fallacy. *Journal of Verbal Learning and Verbal Behavior*, **15**, 262–263.

STORKEL, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, **44**, 1321–1337.

VITEVITCH, M. S. (1997). The neighborhood characteristics of malapropisms. *Language & Speech*, **40**, 211–228.

VITEVITCH, M. S. (2002a). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **28**, 270–278.

VITEVITCH, M. S. (2002b). Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Language & Speech*, **45**, 407–434.

VITEVITCH, M. S., & LUCE, P. A. (1998). When words compete: Levels of processing in spoken word perception. *Psychological Science*, **9**, 325–329.

VITEVITCH, M. S., & LUCE, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, **40**, 374–408.

VITEVITCH, M. S., & LUCE, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers*, **36**, 481–487.

VITEVITCH, M. S., & RODRÍGUEZ, E. (2005). Neighborhood density effects in spoken word recognition in Spanish. *Journal of Multilingual Communication Disorders*, **3**, 64–73.

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**, 491–504.

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

WIENER, F. M., & MILLER, G. A. (1946). Some characteristics of human speech. In Office of Scientific Research and Development Washington DC (Ed.), *Transmission and reception of sounds under combat conditions*. Summary Technical Report of Division 17, NDRC, pp. 58–68. Washington, DC: NDRC.

WIKE, E. L., & CHURCH, J. D. (1976). Comments on Clark's "the language-as-fixed-effect-fallacy." *Journal of Verbal Learning and Verbal Behavior*, **15**, 249–255.

ZWITSERLOOD, P. (2004). Sublexical and morphological information in speech processing. *Brain and Language*, **90**, 368–377.

Appendix A

A.1 Dense Words

Stimulus word	Phonological transcription	Length	Freq.	ND	1-Syll Neigh	2-Syll Neigh	3-Syll Neigh	NF	Iso. Pt.
babble	bæ.bl̩	4	1	12	0	100	0	16.08	4
banner	bæ.nɚ	4	8	10	30	70	0	20.20	4
battle	bæ.tl̩	4	87	14	7	93	0	16.57	4
berry	bɛ.ɹi	4	15	13	0	100	0	77.69	5
body	ba.di	4	276	11	15	85	0	4.64	5
bubble	bʌ.bl̩	4	12	10	0	100	0	13.00	3
candle	kæn.dl̩	5	18	11	9	91	0	8.82	6
candy	kæn.di	5	16	10	56	38	6	7.50	6
cattle	kæt.tl̩	4	97	14	62	38	0	9.64	5
dairy	dɛ.ɹi	4	19	11	10	80	10	82.55	5
furry	fɝ.i	3	1	16	9	91	0	31.94	4
hurry	hɝ.i	3	36	13	7	93	0	997.92	3
ladder	læ.dɚ	4	19	10	10	90	0	13.20	4
leather	lɛ.θɚ	4	24	10	0	100	0	53.50	3
letter	lɛ.tɚ	4	145	13	8	92	0	75.23	5
lever	lɛ.vɚ	4	14	13	0	100	0	113.85	4
lighter	lɑɪ.tɚ	4	12	14	14	86	0	52.00	4
mayor	me.ɚ	3	38	12	75	25	0	320.17	3
meter	mi.tɚ	4	7	11	9	73	18	54.73	4
money	mʌ.ni	4	265	10	20	80	0	123.70	5
muscle	mʌ.sl̩	4	42	11	27	73	0	98.82	5
paddle	pæ.dl̩	4	1	10	10	90	0	7.60	4
polar	po.lɚ	4	7	12	17	83	0	5.42	5
puddle	pʌ.dl̩	4	1	10	0	100	0	2.70	4
shallow	ʃæ.lo	4	14	12	0	100	0	26.25	4
tackle	tæ.kl̩	4	10	12	25	75	0	18.08	4
valley	væ.li	4	73	11	9	91	0	4.73	4
worry	wɝ.i	3	55	12	58	42	0	325.75	4

Note: **Length** is based on the number of phonemes, **Freq** is frequency of occurrence as measured by Kucera and Frances (1967), **ND** is neighborhood density, **1-Syll Neigh** is the percentage of neighbors that have one syllable, **2-Syll Neigh** is the percentage of neighbors that have two syllables, **3-Syll Neigh** is the percentage of neighbors that have three syllables, **NF** is neighborhood frequency, **Iso. Pt.** is isolation point.

A.2 Sparse words

Stimulus word	Phonological transcription	Length	Freq.	ND	1-Syll neigh	2-Syll neigh	3-Syll neigh	NF	Iso. pt.
badger	bæ.ʤɚ	4	1	3	33	67	0	5.00	4
barrel	bæ.ɹ l̩	4	24	7	0	100	0	13.57	5
beggar	bɛ.ɡɚ	4	2	3	50	50	0	142.00	4
bottom	ba.tm	4	88	2	33	67	0	38.50	4
burden	bɚ.d n̩	4	44	4	25	75	0	9.50	4
button	bʌ.t n̩	4	10	6	17	83	0	741.50	4
camel	kæ.m l̩	4	1	8	25	75	0	23.13	5
cancer	kæn.sɚ	5	25	6	0	100	0	27.67	6
cashew	kæ.ʃu	4	1	3	33	67	0	13.00	4
devil	dɛ.v l̩	4	25	3	0	100	0	72.67	4
formal	foɹ.m l̩	5	48	3	33	67	0	212.33	5
hero	hi.ɹo	4	52	3	33	67	0	309.67	5
lady	le.di	4	80	6	17	83	0	16.33	4
lawyer	lɔɪ.ɚ	3	43	7	14	86	0	23.00	3
lotion	lo.ʃɪn	5	8	5	0	100	0	26.80	5
lucky	lʌ.ki	4	21	3	33	67	0	16.33	3
lumber	lʌm.bɚ	5	35	6	0	100	0	80.50	5
mighty	maɪ.ti	4	29	2	50	50	0	337.00	4
mitten	mɪ.t n̩	4	1	8	12	88	0	21.75	4
mother	mʌ.ðɚ	4	216	4	0	100	0	426.25	5
movie	mu.vi	4	29	3	33	67	0	59.00	4
person	pɚ.s n̩	4	175	4	25	50	25	53.00	5
powder	paʊ.dɚ	4	28	2	0	100	0	171.50	3
purple	pɚ.p l̩	4	13	2	0	100	0	427.00	5
shower	ʃaʊ.ɚ	3	15	6	17	83	0	61.17	3
tower	taʊ.ɚ	3	13	8	25	75	0	72.38	3
vapor	ve.pɚ	4	12	4	0	100	0	41.75	5
water	wɔ.tɚ	4	442	3	0	100	0	25.67	5

Note: **Length** is based on the number of phonemes, **Freq** is frequency of occurrence as measured by Kucera and Frances (1967), **ND** is neighborhood density, **1-Syll Neigh** is the percentage of neighbors that have one syllable, **2-Syll Neigh** is the percentage of neighbors that have two syllables, **3-Syll Neigh** is the percentage of neighbors that have three syllables, **NF** is neighborhood frequency, **Iso. Pt.** is isolation point.
There was no difference in the percentage of 1-syllable neighbors in the dense (mean = 17.43%) compared to the sparse (mean = 18.22%) condition, $t(54)$ = .158, p = .88. There was no difference in the percentage of 2-syllable neighbors in the dense (mean = 81.34%) compared to the sparse (mean = 80.89%) condition, $t(54)$ = .087, p = .93. There was no difference in the percentage of 3-syllable neighbors in the dense (mean = 1.23%) compared to the sparse (mean = .893%) condition, $t(54)$ = .289, p = .77.

Appendix B

Number of item responses made in Experiment 1 as a function of accuracy rate

Accuracy rate	Dense	Sparse
0–10%	1	1
11–20%	0	0
21–30%	1	0
31–40%	1	0
41–50%	1	0
51–60%	1	4
61–70%	0	1
71–80%	6	6
81–90%	8	5
91–100%	9	11
Total	28	28

Appendix C

Nonwords used in Experiment 2

be.k m̩	kæ.ɹɪk	lɛ.pəɹg	pæ.lɪg
bə.lup	kæɹ.te	lɛ.mɪk	pæ.ɹəb
baɹ.bl̩	ka.fu	lɪ.zəɹb	paɹ.to
beɪ.s m̩	deɪ.zo	la.bə	pa.kɪp
beɪ. bə	de.lu	lo.kɚ	po.no
beɪ.go	fl̩əɹ.o	laʊ.zu	pʌ.z m̩
bi.te	fʌm.bɚ	ma.tu	ʃol.d m̩
ble.zl̩	hæ.mɪt	meɪ.p m̩	ʃʌ.f n̩
ba.tɚ	hæ.pu	maɹ.bə	tɛ.nɪk
bol.du	kɛ.tɚ	maɹ.ʃu	tɪ.kɪp
bʌ.ke	leɪ.b n̩	meɪ.ʤm	vɚ.ʧi
bʊ.l̩ɪg	la.vo	mɚ.su	vaɪ.ɹɪt
kæ.fo	leɪ.zo	mɪn.ɪb	wɪn.du
kæp.sɪn	li.gə	maʊn.to	wɪ.zəɹk